**London School of Economics and Political Science**

# Stabilising liberal societies in a world of radical innovation:

## Committed actors, adaptive rules, and the origins of social order

Reuben Finighan

A thesis submitted to the Department of Government of the London School of Economics and Political Science for the degree of Doctor of Philosophy, London, July 2023.

**Declaration**

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).



I declare that my thesis consists of 98,972 words.

## Abstract

Long-standing questions about social order, and about liberal democratic capitalist orders in particular, remain unsettled. They are of renewed importance in our age of crisis and democratic backsliding.

Adam Smith addressed two such questions at the founding of political economy: First, what are the forces that sustain all societies, and liberal societies in particular? Second, what combination of market and state makes such societies prosperous and powerful? A third question, addressed by Hayek, Polanyi, and Keynes in their own period of crisis and backsliding, pertains to interactions between the two: how does the combination of market and state affect the stability of liberal democracy?

If we are to answer these questions, I argue we need a realistic theory of innovation. Real-world innovation is Schumpeterian: it is uncertain and often radical, so the future may unexpectedly break with the past. Real-world innovation is Baumolian: it is socially ambiguous, and may be productive or extractive. Consequently, the innovations of political and economic entrepreneurs bring the rise, but also the fall, of societies. Given the last two decades, we may be more open to the idea that Fukuyama's "End of History" never arrives.

Our task is to stabilise and optimise cooperation in both politics and the market. "Cooperation" is defined as the alignment of private returns with social returns; it is exemplified by Smith's "invisible hand", and is the precondition for growth.

The usual formal methods for identifying cooperative equilibria fail in a world of Schumpeterian and Baumolian innovation. Beyond the short-run, there are no lasting Nash equilibria. Game forms are destroyed and remade. The institutional forces that we hope will restore cooperative equilibria are themselves subject to innovative attack. How, in this unstable world, is it possible to sustain cooperation over long periods of time? And how can we model and predict cooperation?

This thesis adopts an analytic strategy that makes this problem tractable. I borrow concepts and formal models from evolutionary sociobiology, a field that deals with cooperation under radical and ambiguous innovation. As in Acemoglu and Robinson's *Narrow Corridor*, the core concept is the *adversarial innovation race (*the "Red Queen's race"). Most important in this thesis is the race between

innovating cooperators and defectors. Social order becomes the probabilistic outcome of a dynamic process—of whether cooperator or defector innovations are superior in a given period. Under the right circumstances, outcomes are predictable.

All complex social orders, anthropic and biological, combine "commitment" and "rules" (which, in the definitions of this thesis, includes institutions) into a self-sustaining system. Commitments are essential. They are motives that are exogenous to the innovation race; while all else changes, they continue to draw the system towards a cooperative equilibrium. They come in two forms: one is an intrinsic interest in others' payoffs, and one is an extrinsic dependence on others' payoffs.

However, commitments are impotent, and indeed are destroyed, if there are no rules or institutions that can control defectors—or if committed actors fail to invest sufficiently in adapting rules so that they keep up in the race against defectors.

In short, social order depends on (A) commitments (i.e. motives to run the race that are innovation-proof) that (B) are channelled into the adaptation of rules, to run the race against defectors. Accordingly, the outcomes of innovation races are predictable under two circumstances: when (A) there is no source of commitment to group payoffs, or (B) when committed actors perversely disinvest from running the race, so play the "sleeping Hare" of Aesop's fable. In either case, loss of the race and collapse of cooperation is guaranteed.

On the first question raised by Smith, I present an impossibility theorem for any society built from rules—from institutions, incentives, and so on—alone. Both liberal and authoritarian orders rest on commitment. Smith's *Theory of Moral Sentiments* is supported: the "very existence" of liberal orders rests on other-regarding preferences (which, I show, is a product of trust). It is the only innovation-proof force available to them. Authoritarian orders can be explained via the ruler's extrinsic commitments alone, though other-regarding preferences sometimes play an important role.

On the second question, every regime of economic regulation is within the innovation race and vulnerable to unanticipated counter-innovations. I show that every regulatory regime can be described as a particular "division of regulatory labour" between institutional actors *and* market actors. Institutional actors and market actors are essential complements, with distinct comparative advantages. A

principal task for the institutional regulator is to structurally simplify complex markets; otherwise, those defectors that have advantages in the innovation race (of which there are many) will predictably exploit both regulator and market actor.

Central planners and Hayekian liberals (and libertarians) endorse extreme divisions of labour between regulator and market actor. They are mirror images and fail in predictable ways. Central planners refuse to use market actors, so allocate hyper-complex (and impossible) regulatory tasks to the state. This produces broad inefficiencies and blocks productive innovation. Hayekian liberals refuse to adapt institutions, so allocate hyper-complex (and impossible) tasks to market actors. This produces crises specifically in complex markets—finance, healthcare, insurance, education, and so on—and soaring rents. Its end point is anarchy.

Hayekian liberals suppose advance knowledge of the consequences of basic market institutions. But the unforeseeability of innovation, and distributed nature of knowledge, are double-edged swords: markets produce both productive *and* extractive innovations that the theorist cannot foresee. To block institutional adaptation is to play the sleeping Hare, and guarantees loss of the innovation race.

On the third question, central planning and Hayek's classical liberalism ultimately lead to authoritarianism. In the case of central planning, Hayek's argument is supported: to attempt the impossible tasks allocated to it, the state must concentrate power, and voters cannot win the political innovation race to control such a state. In the case of Hayekian liberalism, the state cannot run the market innovation race. Market anarchy and crisis erode the commitments on which liberal orders depend, fuelling distrust and parochiality. As Smith observes, "faction" and "fanaticism" are the greatest threats to the liberal order.

To use Hayek's terms, central planning and his own classical liberalism are "fatal conceits": they suppose access to distributed and future knowledge that no one possesses. They are both "roads to serfdom": one via excessive control, the other via anarchy.

I describe the "middle of the road", where commitments are channelled into the adaptive, mixed economic strategy advocated by Keynes. As after the Great Depression, this in turn can create economic outcomes that sustain other-regarding commitments. There, the liberal order can make its home.

# Contents

# Acknowledgements

The proverbial lightning struck twice in 2018, once while on an idle walk in the forests of Forges-les-Bains, and later while dozing by a lake near Castelo Branco. So arrived the two main intuitions behind this thesis, which diverted me from my original plan. After that came the hard work. Some sketches were worked out in a tiny London apartment, and the rest in Melbourne and the nearby Dandenong Ranges during the global pandemic. Places seem important, perhaps because, after weeks or months of fruitless contemplation, a breakthrough simply appears without warning on a bend in a forest track. The memory of those moments remains vivid. But, of course, these ideas are born by pressing together things that I have learned from other people, under the conditions created by a network of institutional, financial, and emotional supports. So I will offer a few words of thanks.

This work would not have been possible without the support of David Soskice, my principal supervisor at the London School of Economics. Many obstacles are thrown in the way of interdisciplinary research projects, and especially works of some ambition. David encouraged me to explore my intuitions, to see whether they might survive interrogation. They did, and he immediately understood the argument I was making. From there, he threw his weight behind me, and urged me to develop the connections with Hayek and make the exposition deeper and more accessible. I am especially grateful for his support on those occasions when parenthood and doctorate came into irreconcilable conflict.

This work would not have been possible without my co-supervisor, Ross Garnaut, who has contributed more than any other person—outside of my parents—to my intellectual development. Our conversations and work together date back a few years before the ideas in this thesis coalesced. Although I learned a few things at Harvard, it was Ross that gave me an education in the great movements of politico-economic history. Separate threads I already possessed came together, finally, into a coherent tapestry, and made further leaps possible. He has also been a tireless and penetrating critic of the work.

It would not have been possible without funding from the John Monash Foundation and the Leverhulme Trust. My gratitude goes to the benefactors and their love of knowledge, and the panels who thought I was worth a crack.

And, of course, I could not have done this without my family. My thanks to Yulia for her love, endless patience, and the lending of her razor-sharp mind to the improvement of the work. The same to little Osh, for leavening my working day with his joyful play. My parents' values and curiosity suffuse everything I do, and their love is part of the ground on which I walk. I look forward to spending more time with you all.

I am grateful for conversations with, and comments from, many people along the way, including Fergus Green, Mark Fabian, Jonathan Birch, Andrew Walter, John Sidel, John Daly, James Button, Julian Brown, and Braden Kier. A special thanks to Alan Schwartz for years of stimulations, provocations, and generosity, and Nicholas Gruen for sharing his view of the same mountain from a very different angle. Thanks are also due to my examiners, who helped me to dot i's and cross t's, and have been so generous with advice about next steps.

Beyond that, it is fitting, given the contents of this work, to acknowledge my debts to those countless individuals who contributed to building the culture and institutions that were my inheritance in Melbourne, Australia. Without the moral sentiments of my forebears, and their efforts to convert them into just institutions, I would not have survived much beyond the age of two, nor had the opportunities for so much education. Civilisation has been hard won, and will remain so. This is the beginning of my own little contribution to passing on the torch.

*Adapt or perish, now as ever, is nature's inexorable imperative.*
H. G. Wells (1945, p. 19)

*…he that will not apply new remedies, must expect new evils; for time is the greatest innovator; and if time of course alter things to the worse, and wisdom and counsel shall not alter them to the better, what shall be the end?*
Francis Bacon (1597, p. 171)

*We remain at the end of history because there is only one system that will continue to dominate world politics, that of the liberal-democratic west.*
Francis Fukuyama (2001, para. 14)

# 1

# Introduction

What sustains order in liberal democratic capitalist societies? The first two decades of the century provide us with an unprecedented, if unwelcome, opportunity to study disorder. In 1989, Francis Fukuyama declared the "unabashed victory of economic and political liberalism" (p. 10), a thesis developed in his 1992 book *The End of History*. By 2018, the year this project of research began, Fukuyama had been superseded by Levitsky and Zeblatt's *How Democracies Die,* Albright's *Fascism: A Warning,* and Sunstein's *Can it Happen Here?* Just three years later, "it" almost did happen: the United States Capitol was stormed by a mob that hoped to block the peaceful transition of power. In the following months, many American states altered electoral procedures in ways that will make it easier to overturn results, leading more than 100 of the foremost political theorists (including Fukuyama) to sign a "Statement of Concern" about the deterioration of US democracy.[1] This deterioration is not isolated; while populism was once associated mostly with Latin America (Dornbusch & Edwards, 1991), the current success of populism is historically unprecedented, with 26 percent of major economies run by populists in 2018 (Funke

[1] Those changes to procedures "transforming several states into political systems that no longer meet the minimum conditions for free and fair elections". See New America Foundation (2021, para. 1).

et al., 2020). In the 21st century, some 80 percent of cases of democratic decline have been voter-driven (Mounk, 2018).

Capitalism, meanwhile, has its own maladies. The stylised macroeconomic facts of the post-1980 era are slowing growth (in Summers, 2015, "secular stagnation"), weak investment, heightened systemic risk, and inequality. In the U.S., median wages have halved relative to GDP since 1980, so that average standards of living have stagnated or reversed. This marks a return to the pre-war Gilded Age pattern of capitalism. Each of these trends appears linked to a roughly fourfold expansion of economic rents[2] as a share of GDP (De Loecker & Eeckhout, 2020; Barkai, 2016; Autor et al., 2020; Jones & Phillipon, 2016; Crouzet & Ebery, 2021),[3] and the sharp decline in various indicators of market competitiveness over the period is consistent with this story (Furman & Orszag, 2015).[4]

Scholars' explicit and intuitive models of the world predicted more or less the opposite—continued democratisation and strong growth in a globalising world. To have forecasted these disturbances in democracy and markets in the 1990s would have marked one out as the most incurable of pessimists. One possibility is that these developments are, in technical terms, a "shock"—some sort of unpredictable, exogenous event, just as a healthy person may blamelessly catch a serious flu, but has reason to hope for a natural return to good health. I want to suggest, however, that it is rather like the cirrhosis of someone who repeatedly ingests a poison. We caused this, with the right model it was predictable some decades ago, and it will continue to worsen until something changes. In Hayek's (1945) terms, we have wandered down a "road to serfdom". It is not, however, the road he foresaw. He warned of the (real) dangers of central planning, but that particular threat receded long ago, and more or less decisively with the fall of the USSR and communist China. The spectre of serfdom has appeared to us in a Hayekian age, and we need to understand why.

[2] Rents in Ricardo's (1814) sense, of persistent excess profits, above those required to bring a factor (e.g. labour, capital, or land) to market. They indicate the absence of competition.
[3] Barkai (2016) finds that the growth in rents implies a reduction in output of at least 10 percent, Jones and Philippon (2016) find that it explains a large share of the reduction in investment, and reduces interest rates by almost two percentage points, and Crouzet and Eberly (2021) find that it may explain half of the decline in total factor productivity growth (TFP, the driver of long-run growth).
[4] Including reduced rates of market entry and exit, increases in the number of firms sustaining ultra-high profits over decades, and divergence of profitability between low-rent and high-rent firms.

I argue that the cause is our failure to grapple with the consequences of *innovation*. Later I will discuss historical shifts in thinking about innovation—from being "bad change" in the time of Bentham (1824, p. 144), to "unforecastable improvement" by the twentieth century.[5] Both views are incomplete. Both are half a theory of innovation, and—characteristically for a species inclined to thinking in blacks and whites—we switched from one to the other, rather than to the middle position. I will later argue that this gives us half a theory of economics, and half a theory of the liberal society. Scholarly understandings of liberal societies are largely models of health and stability, rather than of sickness. But, as in medicine, we can only properly understand the former if we understand the latter.

In this thesis, I modify the treatment of innovation in standard models of capitalism and democracy. The implications are far-reaching and profound. If my analysis is correct, strong claims can be made that carry the force of logical necessity. First, the resulting model tells us about the ground of human societies. In the case of liberal societies, it will return us to Adam Smith's *Theory of Moral Sentiments*, and give concrete form to the musings of many of the most significant economists and political economists of the last century. Specifically, the fact of innovation means that liberal order logically rests on other-regarding preferences. Second, a more complete model of innovation tells us what is involved in efficient long-run management of the market economy, and answers some long-standing questions about the balance between market and state.

It shows us, I suggest, the nature of our contemporary road to serfdom, and why our economies have become stagnant and concentrated—why liberal societies today appear like cautionary tales rather than beacons of prosperity and liberty. Most importantly, it points us to a better course.

### 1.1 Objectives

In broadest terms, this thesis seeks to develop a general theory of how social orders adapt, persist, and prosper—or fail to do so—in a world where they are perpetually challenged by technological and strategic[6] change. In brief, it seeks to explain how

[5] The latter term intended critically by Awrey (2012, p. 258), discussed further in Chapter 5.
[6] If "hard" technological change ceased, history would continue. Leaders would rise and fall, political systems would decline or be remade, groups would shrink, grow, or change, firms would rise and decline, collude and collaborate, and interact with political systems in various ways. We could call this

*cooperation* can be maintained amid *innovation*. I will define cooperation and innovation, describe their interactions, and comment on the origin of this general theory in a moment.

I call it a "general" theory for two reasons. First, because it allows for uncertain change in game parameters, and so relaxes assumptions that are in standard economic and game theoretic approaches. These standard approaches are, in this general model, only special, short-run cases. Second, it is general in that it can be used to describe all manner of social orders, from the family to the state, and even sociobiological orders. But that is a matter to be explained shortly.

The main case of interest will be the liberal, democratic capitalist society. In my view, where they have been established for some time, such societies offer standards of human welfare far beyond those reached in authoritarian political systems. This is a view with some empirical support.[7] To understand the conditions in which liberal societies can survive and flourish is worth considerable analytic effort. The analysis of this case offers some novel answers to questions about the forces that hold democracy together, the appropriate roles of market and state, and how capitalism and democracy are in some ways in conflict and how they may be made compatible. A sub-objective is to offer a diagnosis of the present malaise of liberal societies, and to gesture towards some promising remedies.

Note that the international order is beyond scope, and a matter for future work; this thesis is focused on innovation and cooperation within nations and their economies.

### 1.2 Core concepts: Cooperation, innovation, and their interactions

By "cooperation", I mean the coordination of individual behaviours such that they optimise group payoffs, by ensuring that individual and group payoffs are *aligned*. As North and Thomas (1973, p. 1) put it, efficiency and economic growth are the consequence of various forces that "bring the private rate of return close to the social rate of return." This alignment is what contemporary economists mean when they

"social innovation", but this has a commonplace meaning I do not intend here. To keep it as broad as possible, I will simply refer to "strategic change".

[7] E.g. see Acemoglu et al. (2019) on growth, Dorn et al. (2007) on happiness, and Ray (1993) on lower propensities for war.

refer to Smith's (1776, p. 456) "invisible hand".[8] Thus, intense competition between self-interested butchers, bakers, and aircraft engine-makers is still *cooperation* with society at large, so long as that competition is well-governed. Some formalisms will help make this clear in the early chapters.

By "innovation", I intend a broad meaning that could be summarised as the search for, and adoption of, new technologies and strategies of all kinds. It includes "hard" technologies, from the loom to the microchip, and "soft" technologies, such as methods of management or decision-making. It includes shifts in political strategies, such as the formation of new coalitions or new rhetoric. It includes imitations, which may not be new under the sun, but are always new in the sense that they bring a change. It can be summarised as "strategic change", and it is a source of dynamism in all social systems.

Crucially, this innovation is of the Schumpeterian (1942) variety, in that it is subject to profound uncertainty and periodic revolution. As Schumpeter puts it, in capitalism

> *we are dealing with a process whose every element takes considerable time in revealing its true features and ultimate effects.* (p. 83)

Understanding the full "effects and counter-effects" of any enterprise is an "impossibility" (Schumpeter, 1934, p. 85). In a Schumpeterian approach to innovation, departures from perfect rationality are essential; it necessarily implies uncertainty of the Knightian/Keynesian variety, as opposed to calculable risk (see Kay and King, 2020, for a recent overview). I explore the necessity of this approach to innovation, and so to understanding social order, in Chapter 4. In any case, throughout this thesis I will take individual behaviour as rational in the more limited sense that it is *reasoned* and *aimed* at optimisation, even where actual optimisation is as unlikely as Schumpeter suggests.

Views on the implications of innovation for cooperation have differed markedly throughout history. As we will see in later chapters, "innovation" was a pejorative for most of history, while today the dominant view is almost wholly meliorative. Innovation was once for worse, now it is for better. Baumol (1990), like Bentham

[8] This not Smith's originally intended meaning of "invisible hand", but an erroneous and now conventional interpretation by Samuelson. See Kennedy (2010).

(1824) in the nineteenth century, offers a balanced perspective on innovation. He observes that innovators and entrepreneurs may invest their efforts in either productive or extractive activities: they may produce steam engines and automatic teller machines, or they may produce mortgage-backed securities that raise systemic risk, marketing strategies for addictive opioids, and "deep fake" videos that make the collective determination of truth impossible. Innovation is, in itself, always *ambiguous* for the group; it may build up or destroy economies and political systems.

### 1.3 Motivating the thesis: Grand lacunae and a new way forward

With that, we can turn to a few early comments on the motivation of this research. In standard economic and politico-economic models, the unruly force of innovation is domesticated by institutions. Effective institutions, such as those establishing healthy democracies or competitive markets, favour productive innovations from economic and political entrepreneurs, while suppressing the extractive. Individuals are led to make socially efficient choices. Prosperity results.

Yet there remains great uncertainty and long-standing disagreement over:

1. the sufficiency of institutions alone as a driver of adaptation, and the role, if any, of Adam Smith's "moral sentiments"; and
2. the nature and extent of institutional adaptation required as innovation proceeds.

On the latter, take the contrast between Hayekian and Keynesian economic management. Hayek's approach is minimally interventionist, and the core institutions of the free market—property rights, free entry—more or less solve the adaptive problem. There is a "big bang" moment where efficient market institutions are set in place, and thereafter the adaptive process spontaneously unfolds as a result of market actor optimisation. Keynes' approach is one of observation, institutional experimentation, and intervention where the benefits are thought to outweigh the costs. Rather than a big bang, we have continual institutional adjustment and reform.

On the former, there is a broad mix of views on how economic and political institutions interact with moral sentiments (e.g. actors' other-regarding preferences,

or preferences over others' welfare). In the dominant self-interested model, such sentiments are fictitious or at least impotent. Others—such as Adam Smith, Arrow, Hirsch, North, Polanyi, and Schumpeter—have speculated about their significance as a social lubricant, or asserted that they are the very ground of social order. But the latter work is overwhelmingly informal, discursive, and loose on the mechanisms involved; it does not lay out a systematic model or "blueprint" for social order.

Thus, there are yawning gaps between different theorists' overall understandings of social order and prosperity, and how they are maintained in an innovative world. One area of progress, on the first of the two problems above, is the broad rejection of central planning. If we can rule out other flawed conceptions of social order—and particularly conceptions that underlie the present degeneration of liberal societies—this would be worth doing.

The claim of this thesis is that a simple formal approach to studying social order, built on Schumpeterian and Baumolian innovation, can provide stringent logical constraints on these two problems, and a couple of impossibility theorems to boot. The method, discussed in more detail in Chapter 2, departs from conventional formal approaches that cannot handle Schumpeterian innovation—e.g. from standard game theory, which necessarily specifies the nature of the technical and strategic changes that are allowed to emerge. An unusual step is to import theory and methods for dealing with genuinely uncertain, open-ended innovation from the cognate field of sociobiology. There too, short-run game theoretic analyses are special cases, and no longer apply when the theorist allows genuine innovation. As I explain in Chapter 2, the result is a highly general theory of social order that applies to any system of cooperation with open-ended innovation—whether human, sociobiological, or involving artificial agents. I point to similar raids on biology, exploiting similarly general theories, conducted by economic luminaries such as Milton Friedman.

The next two sections of the introduction broadly summarise some key literature relating to the two questions raised above—about institutional adaptation and institutional sufficiency—and summarise the main results of this thesis. Finally, the last section of this introduction offers a more technical, step-by-step summary of the argument.

## 2.1 The first result: On the foundations of social order, and of liberal orders in particular

The first result addresses the first dot point in the previous section. To sketch it out, let us start with Kenneth Arrow. In 1972, he wrote that

> *Many of us consider it possible that the process of exchange requires or at least is greatly facilitated by the presence of several of these virtues (not only truth, but also trust, loyalty, and justice in future dealings).* (pp. 345-6)

It is not an unreasonable hypothesis, he suggested, that people may contribute to public goods so they may "have the satisfaction of personal participation in social welfare". In the production of the market society,

> *one might loosely say that the categorical imperative and the price system are essential complements.* (pp. 357)

That is, moral sentiments and markets must go together. On the few occasions Arrow considers this topic—one that interested him more than his published body of work might suggest—he is openly tentative. It is "possible", speaking "loosely", that moral preferences of some kind are necessary or at least useful. Decades later, he explained that "it's not so easy to construct a broad theory" (Arrow 2006), and

> *I generally try to write things I feel sure about. As I get older, I'm a little more speculative and start to stimulate other people to think more about it.* (p. 12)

Let us be so stimulated, and see if we can find a way to be sure about it.

## 2.2 An entry point: Social order in Smith

Not all have been as coy as Arrow. Adam Smith (1759/1976, p. 13) regarded the existence of other-regarding preferences as "a matter of fact too obvious to require instances to prove it", and he regarded them as the ground of social order. The first of our two results in this thesis could be construed as a logical proof of Smith's complete system of social order presented across *The Theory of Moral Sentiments* (*TMS*) and *Wealth of Nations* (*WoN*), or at least some close neighbour of that system. This was not the goal of this thesis, but is, rather, a happy outcome of the

work. The core parts of the analysis were completed before I learned that Smith had arrived at a similar view of social order. If the argument sustained by this thesis is correct, Smith cuts an even greater figure in the history of liberal thought than is generally recognised—he showed us how to understand the market order, and how to understand the broader social order in which markets are necessarily embedded.

At the foundation of Smith's system is a force that Smith calls "sympathy" or "fellow-feeling".[9] It is a second force, alongside institutions, drawing private payoffs closer to group payoffs and so generating cooperation.[10] Smith holds little doubt that sympathy is, against self-interest, the weaker force.[11] And yet he argues that fellow-feeling is the ground of every social order. "The very existence of society" depends on it (p. 89, 106).[12] The *TMS* offers Smith's *general* explanation of social order, while *WoN* offers a *subordinate* explanation of the market order.

The *TMS* attempts a positive, scientific description of human psychology and the resulting "system of sympathy" (Smith, 1759/2009, p. 374), where social preferences produce good governance (and, following *WoN,* wise governors will frequently employ markets). Today, research is converging towards this system, with a spate of formal models where social preferences determine equilibria. These models, like this thesis, have been a response to the shock of growing democratic dysfunction. They follow a few decades of suggestive empirical research, such as Robert Putnam's (2000) influential *Bowling Alone*, charting the decline in American trust and predicting dire consequences for the quality of government; Elinor Ostrom's (1990) studies of trust-based public good provision in small-scale communities; and a large empirical literature finding unusually strong correlations, and causal linkages, between trust and the key economic and political measures we care about: economic growth, institutional quality, and political stability (surveyed in Chapter 15). These

[9] Today, it is common to translate Smith's "sympathy" as "empathy", but his model is much richer. It is discussed in Chapter 15.
[10] The mechanism is obvious enough: others' payoffs enter into our own. If personal payoffs and others' payoffs are given the same weight, the individual tries to optimise for the group.
[11] Especially when people are dealing with strangers, this may seem another matter "too obvious to require instances to prove it." In the economic game theoretic studies discussed in Chapter 15, norms are more potent than expected—especially in the field. But people do on average tend to value themselves a little more than strangers.
[12] See also Smith (1759/1976, p. 86): "Nature has implanted in the human breast that consciousness of ill-desert, those terrors of merited punishment which attend upon its violation, as the great safe-guards of the association of mankind, to protect the weak, to curb the violent, and to chastise the guilty."

**Box 1.1 Resolving the "Adam Smith problem"**

Given the precedence of sympathy in Smith's social order, it is unsurprising that he regarded the *TMS* as his most important work. Curiously, many scholars have expressed surprise that Smith should be the author of both, viewing the *TMS* and *WoN* as somehow dichotomous. This is the so-called "Adam Smith problem",[13] the puzzle that he could place morality at the ground of things in the *TMS*, and offer a theory of order emerging from self-interest in *WoN*. Given their original publication dates (1759 and 1776 respectively), one could be forgiven for thinking the *TMS* is superseded by *WoN*. Smith, however, developed the substance for both in his four-part lecture series at the University of Glasgow. His main effort towards the end of his life was to substantially extend the *TMS* with a sixth edition.[14] As Sen (2009) observes, it was Smith's first and last book, with *WoN* wedged in between.

The confusion arises, perhaps, because social scientists' knowledge of Smith's work is dominated by a small fragment of it: a few lines about the self-interest of butchers and bakers. It is a delightful and important fragment. Yet when ripped from his larger system, and applied well beyond its intended domain of market activity, it suggests Mandeville's (1714) view that broader social order can emerge from pure self-interest. This view, which Mandeville shared with Hobbes and which was much admired by Hayek (2005, p.80), was in important respects the converse of Smith's view. Smith (1759/1976) commented at length on Mandeville's thesis, arguing that it was "wholly pernicious" and "in almost every respect erroneous" (p. 159), although he credited Mandeville for recognising that self-interest had its uses.

Smith's argument is straightforward. Sympathy is essential for generating the institutions of society. In turn, those institutions, and especially in the market, may fruitfully harness self-interest. But it leaves many questions unanswered. Why can we not get by with institutions and constitutions alone? How can sympathy achieve something so profound as the grounding of liberal social orders when theorists such as Smith, and the empirical literature, suggest that it is in everyday life a weaker force than self-interest? How do these forces—sympathy, self-interest, and institutions—come together to form a coherent, self-perpetuating social order?

This thesis aims to complete Smith's project by providing clear answers to these questions. Some practical advice for policymaking will emerge along the way.

[13] See Montes (2003) for an overview of the "Adam Smith problem" and surrounding debate.
[14] As recounted by John Millar, one of Adam Smith's students. See Sen (2009).

kinds of claims have a longer history, including in Karl Polanyi's (1944/2001) warning that loss of trust in liberal societies led to fearsome backlashes—to communism, fascism, and war.

Nonetheless, these do not yet amount to a revolution in thinking, and the causes of Arrow's hesitancy remain in place. The various findings are impressive and strongly suggestive, yet they remain too soft to scratch the diamond-hard surface of the self-interested model.

This leads to the following proposition:

> **We still disagree on the matter of whether other-regarding preferences are *critical necessities*, *nice-to-haves*, or mere *illusions*.**

A stimulating contrast is between Hayek, as a chief influence on the post-1980 liberal order we live in today, and Smith.

- Hayek saw sympathy as an archaic force suited to tribal life, small-scale community, and family interactions, which is destructive when applied to reform the market society (Hayek, 1985).[15]
- Smith saw sympathy as the key to the past, present, and future prosperity of human society, and the source of publicly favourable market institutions. Without it we cannot sustain justice or efficiency.

### 2.3 An impossibility theorem for the purely institutional society

Since we began with Kenneth Arrow, it is fitting that the first result can be characterised as an "impossibility theorem".[16] What is impossible is the purely institutional society. Beginning from conventional premises, I tighten the screws on the standard model until it is forced to confess: it cannot explain the data. Schumpeterian innovation makes it logically impossible to build a liberal society

[15] For Hayek, other-regarding preferences are in "constant conflict" with the "traditional" moral rules, which are "not altruistic" but that underpin the market order (Hayek, 1985, 14:44). Altruism must be "restrained". Other-regarding preferences are suited to the family and small community: "our intuitive moral feelings … are adapted to the small person-to-person society".
[16] Arrow (1951/2012) offering his own famous "impossibility theorem" in social choice theory.

from self-interested actors and institutions alone. No Coase theorem, nor any constellation of private interests, such as balances of power, can provide a way out.

The explanatory gap has a very specific shape. For the liberal society, at least, individual utility functions must include some positive weighting of others' or group payoffs. This social preference must not be the "artificial" product of institutional incentives. People must be *committed* to group payoffs in the same manner, if lesser in degree, that they are committed to their own private payoffs. Smithian "sympathy" is a force of this kind. The framework allows us to identify some necessary, and some probable, properties of social preferences, and I canvas broad and deep empirical evidence that such properties exist in practice. Smith's powers of observation are impressive: the empirical evidence, and even the latest neuroscience,[17] is in broad agreement with his original approach.

This impossibility theorem also applies to authoritarian societies, although they may draw on a source of commitment that is unavailable in democracies. A more general way of stating the theorem is as follows:

> **Institutional structures are fatally incomplete without an exogenous source of "commitment". In democratic societies, commitment can only come from other-regarding preferences. Authoritarian commitment may arise under pure self-interest, though other-regarding preferences are stabilising.**

Hobbes (1651) was correct that the democracy of self-interested actors would be mere anarchy, before the inevitable return of the Leviathan.

The analysis suggests a *universal principle* for complex cooperative systems with open-ended innovation: no such systems can exist without a source of commitment. It applies in every complex cooperative order in sociobiology (Box 1.2), and to artificial agent systems.

[17] The function of "mirror neurons" is to represent others' bodily states, e.g. pain, as if we were feeling it ourselves. Smith remarks on this capacity and its significance for morality. See Chapter 15.

**Box 1.2 Borrowing from sociobiology**

Sociobiology plays a useful, if inessential, role in this thesis because (i) these conditions effectively always apply, and therefore every sociobiological system that sustains deep and complex forms of cooperation—whether a bacterial community, cross-species symbiosis, or an organism—involves mechanisms that lead actors to weight the payoffs of others, and (ii) sociobiologists already have well-developed formal approaches to talking about this problem. I am, however, running ahead into complex parts of the argument. Again, the ways in which the argument borrows from sociobiology are explained in Chapter 2.

This complementarity between commitment and institutions goes both ways:

> **No source of commitment can survive without a supporting system of institutions.**

Or in Smithian terms, one might say that justice depends on sympathy, but sympathy cannot persist in a world without justice. This too is a universal principle in complex cooperative systems, again found in sociobiology. Together,

> **Institutions and commitment—and institutions and other-regarding preferences in the case of liberal democratic societies—are *essential complements* in the production of social order.**

I explore the mechanisms by which other-regarding preferences are channelled into the institutional structure. Downs' paradox—the absence of private incentives to vote or provide other key democratic public goods (Downs, 1957)—is thought to represent a threat to democracy. We will see this is not a bug but a feature. Democratic societies depend on exploiting choice architectures that weaken or eliminate private incentives, allowing other-regarding preferences to determine individuals' choices even where those preferences are weak.

## 2.4 Parochial other-regarding preferences: The principal threat to democracy

Reliance on other-regarding preferences has its problems. I echo Smith in another aspect of the argument: social preferences have a dark side. Smith's disdain for purely self-interested individuals is palpable, but it is "faction" and "fanaticism" that he regards as the real threats to social order (Smith, 1759, p. 71). Note that, with Smith, I view mild parochiality—e.g. modestly greater interest in those who live in the same community, province, and country—as an innate tendency of little harm. It may even play an essential role in gluing human groups together. By parochiality in the thesis, I am referring to *deeper* social cleavages, and cleavages *within countries*, of the kind that are familiar in our age of populism (and that befit the language of "faction" and "fanaticism").

*Parochial social preferences* lead individuals to supply *narrow public goods*, and the latter are often broader public "bads". Every war, system of slavery, and genocide has depended substantially on narrow public goods, and so on parochial social preferences. It is those social scientists who regard other-regarding preferences as mere "sentimentalism" that are the sentimentalists; they hold a dewy-eyed view of what other-regarding preferences entail.

> **Liberal democracies rest on institutions that reveal and channel underlying other-regarding preferences, leveraging them to provide fundamental democratic public goods. If other-regarding preferences are parochial, then those institutions reveal that parochiality. Strong parochiality corrupts all institutions and leads to civil conflict and democratic decline.**

Maintaining a favourable pattern of social preferences, compatible with the liberal society, requires attention to their causes—what an economist might term their "production function". A survey of the political science literature provides a rich, if preliminary, view on the most important factors. One clear message is that societies must generate broadly-based prosperity if they are to sustain democratic capitalism. Equity and efficiency, values that are often considered opposed, are, in the long run, inseparable.

It is hoped that this research will give greater clarity and definition to our picture of how human societies work—or indeed how they fall apart. Unlike previous qualitative works, like those of Smith or the Polanyi brothers, we end up with something akin to a general "blueprint" for the market society. We will see how institutions and social preferences act as essential complements, each neatly filling in for the weaknesses of the other. The analysis will shed light on some of the most challenging problems of institutional design in democratic capitalism, such as maintaining trustworthy political media, and suggest some ways forward.

### 3.1 The second result: On the technical problem of economic management

Now we turn to the second dot point or "gap" introduced in Section 1. If a social order is to survive innovation, some source of "commitment" to that order is necessary but not sufficient. A society may be committed to raising the general welfare and yet run itself into the ground with incompetence. Commitment must be combined with *competent rule-making*. Here too the overall structure of the thesis, or at least its applied sections, echoes Smith's work. While the *TMS* was concerned with the ethical preferences that grounded social order, *WoN* was concerned with matters of "expediency",[18] or the prudent and practical methods—such as appropriately regulated markets—that "public spirited", other-regarding people could use to raise the standard of human life and strengthen the state. Here we enter the territory of expediency.

With Bentham (1824) and Baumol (1990), the framework holds that individual innovation is, *ex ante*, always ambiguous. Individuals innovate for private gain, with unknown consequences for the group. As the long history of barely perceptible improvement in humanity's material conditions suggests, there is no natural bias towards productive contributions. That bias comes from institutions, and institutions are themselves *strategies*.[19] Any particular set of institutional strategies—and any particular approach to using the market—is vulnerable to counterstrategies. This is

[18] As put by John Millar, Smith's disciple and friend, and recorded by Dugald Stewart (1795/1980, p. 10).
[19] I.e. they are social technologies, innovated and implemented by individuals working in groups.

akin to Myron's law: "Asymptotically, any finite tax code collects zero revenue" (Romer, 2010, p. 1, attributed to Myron Scholes).

How do we achieve cooperation—the alignment of individual and group payoffs? A useful starting point is to note the silence of the neoclassical model: it is consistent with anything from *laissez faire* to central planning (e.g. Pareto, 1906/2014, Barone, 1908, and Hayek, 1945). Hayek understood that we can only approach efficiency if we understand the concrete capabilities of different actors in the system—particularly the regulator and market actors (individuals and firms). Put another way, efficiency is always the result of a *total regulatory structure* that emerges from a *division of labour* between regulatory actors and optimising market actors. This division of labour, like any other, ought to be designed in light of the *comparative advantages* of these actors. Hayek's argument, for example, was that market actors are advantaged in the use of distributed knowledge.

Crucially for our argument, when the regulator and market actors play their role in this system, both confront *antagonistic innovation races*. Both the regulator and market actors must be able to run these races. The framework of this thesis suggests that policymakers and economists neglect the all-important practical matter of whether market actors (especially consumers) and regulatory actors can win these races. If they cannot, the total regulatory structure fails and produces crisis. The argument is summarised in Section 4, but I will continue summarising conclusions here:

> **Prosperity requires institutional design that is sensitive to (A) asymmetries in firms', consumers', and regulators' capacities to innovate, and (B) the complexity of economic activity. Asymmetric capacities, in combination with complexity, allow rent extraction and provoke crisis.**

Each of the major crises of twenty-first century democratic capitalism is a consequence of this combination, including the Global Financial Crisis, opioid crisis, and crisis in information markets. Instances of this argument can be found especially in the literature on financial regulation, which reflects the great complexity of financial markets. It is implied in the work of Hyman Minsky, who argued that regulators who focus on allocative efficiency will pay costs in terms of system

stability (e.g. Ferri & Minsky, 1992). Some similar arguments can be found through the commentary that followed the Global Financial Crisis (e.g. Haldane, 2012, and many others cited in Chapter 12). I give this argument some additional formal grounding, tighter definition, and generalises it to all complex markets. It results in the following prescription:

> **A core task of regulation is to structure and simplify the tasks that are allocated to regulatory and market actors, so they may feasibly accomplish them.**

Structural regulation and market actor optimisation are essential complements.

### 3.2 Central planning and libertarianism: Two fatal conceits

The arguments in this thesis have clear implications for the contest between grand theories of economic management—for the different proposed divisions of labour between market and state.

If Smith is the hero of this thesis, Hayek is the anti-hero. One of Hayek's great contributions was to articulate what Gray (1998/2013, p. 145) calls an "impossibility theorem" for the (efficient) centrally planned society.[20] This research project offers fresh support for Hayek's result, but also produces its mirror image: an impossibility theorem for the self-regulating market, libertarianism, and the classical liberalism that Hayek advocated. The analysis suggests that:

> **The impossibility theorem is symmetric, eliminating central planning but also libertarianism and Hayek's own classical liberalism.**

Hayek's "distributed knowledge" and epistemological arguments undermine the case for central planning, but they cut just as sharply against the unregulated market. The fatal conceit of libertarianism and its bedfellows is that it purports to have identified an "unbeatable" institutional strategy in a world of uncertain, radical innovation where institutions are vulnerable to unknown counter-strategies. Knowledge about

[20] Gray (1998, p. 150) writes that "[i]t works only as an impossibility theorem against the most hubristic types of economic planning… It demonstrates that a powerful twentieth-century project, the Marxian project of replacing market processes by central planning—is unachievable. It tells us little else."

present-day extractive strategies is distributed, and unavailable to the armchair theorist. The libertarian and Hayekian must further claim knowledge of future strategies, if they are to be confident that simple market institutions will be adapted to them. They sleep in a fierce innovation race, believing they have already won it. The inevitable and predictable result is catastrophic failure, as creative firms outsmart the libertarian's simple designs. The very point of markets is that they enable innovations and counter-innovations that no theorist or bureaucrat could imagine. To our benefit and peril, individuals and firms are, *en masse*, smarter than the central planner and smarter than the libertarian.

One may say that both central planners and libertarians make *a priori* commitments with the division of labour in the total regulatory structure—both making strong presumptions about the relative powers of market actors and the state. Both are uninterested in *a posteriori* knowledge, won from hard experience and study, of the actual effects of markets in practice. Both involve the same kind of "fatal conceit".

I point to a "Libertarian's Trilemma": in complex markets, it is impossible to combine (A) efficiency, (B) freedom of contract, and (C) a small regulatory apparatus. Libertarians choose B and C, and so sacrifice efficiency. This sacrifice is often catastrophic in scale. In turn, mismanagement of these innovation races undermines political stability and the trust that underpins liberal societies. It raises the relative competitiveness of authoritarian states. This is the tale of the last few decades, and the harvest sown by the economic philosophy of Hayek and Friedman.

Adaptation, and the prevention of emerging crises, requires running the innovation race: collecting evidence about how the system is functioning, how relative capabilities are changing, and how the division of labour between structural regulations and market actor optimisation must be altered.

### 3.3 Two roads to serfdom

Our story of complementarities—between institutions and other-regarding preferences, and between market and state—has one final movement. Democracy and the mixed economy are essential complements, such that both central planning and libertarianism end in authoritarianism. Central planning because, in the absence of markets, institutional adaptation requires extreme concentrations of knowledge and control. Libertarianism because it refuses to adapt the institutional structure at

all. Amid a heated innovation race, it sleeps idly at the side of the track, allowing special interests to engage in an anarchy of extraction. This undermines the apex public good of the liberal society: the social trust that underpins all adaptation. It invites the return of that promisor of order, the Leviathan.

> **There are two "roads to serfdom", each heading in opposite directions. Extreme information and control requirements mean central planning must concentrate power. The anarchy of rent extraction under libertarianism sows distrust in institutions, provoking a retreat to parochial populism and the strongman. The road to freedom wends a careful path between them.**

We are fortunate that genuine libertarianism has never been tried. Central planning leads to inefficient autocracy, but, to its minor credit, it maintains some means for adapting to change. The libertarian society—without recourse to central banking, to licensing for doctors, to product standards, and so much besides—is an endless parade of catastrophes. In the mixed economies of the real world, even relatively minor and partial movements towards libertarian-style deregulation have been enough to substantially tear the fabric of economy and society, and create opportunities for parochialists with authoritarian ambitions.

Adam Smith would regard the passions for central planning and the minimal state as manifest perversities. Central planners and libertarians deny the complementarity between state regulation and market actor choice. Both pre-emptively close off one of these two tools for adaptation, in an evolving universe that will set us challenges we cannot foresee. The market is, for Smith (1776), a tool—a matter of expediency and practical judgement—that serves us well wherever it tends to raise human happiness and sustain social order. Wherever markets were well-suited to serving society, he favoured them. Wherever he saw significant limits to markets, he proposed public remedies. Wherever he was uncertain, he suggested experiments with both markets and public provision. He put the maintenance of social order above all else. He was a pragmatic adaptationist, and so a fine man to be the first economist.

# Technical argument summary

Now we move on to the technical argument summary. While it is conventional to present chapter-by-chapter summaries, complexities in the nature of the argument mean that it is most lucidly presented in its four parts. I hope that the reader may find this a useful entry to the argument, and a reference to return to as needed. Some readers may find conventional chapter summaries a useful complement, and can find them in A.1.1.

## PART I: Cooperation and innovation

### Cooperation: Individual-group payoff alignment

Growth and prosperity depend on the alignment of individual payoffs with group payoffs, so that the optimising individual optimises for the group (e.g. North and Thomas, 1973). This is the meaning of Smith's (1776, p. 456) "invisible hand", and it is a principle common to all systems of cooperation—anthropic, biological, and digital—that fall under the abstract framework developed in the first two parts of this thesis.

In human societies, this general alignment does not occur as a matter of course. Historically and today, absent or poorly designed institutions produce stagnation or only slight improvement in economic conditions, and growth flowers only when institutions improve enough (Jones, 1988).

There are only two generic forces that align payoffs. They are:

1. **Rules**. These are punishment and reward strategies used to *modify individuals' payoffs*, so that they choose strategies that benefit rather than harm others. Examples include individual-level Tit-for-Tat, and institutions such as property rights, Pigouvian taxation/subsidy, and the laws that govern politics. This novel definition of the term "rules" may be unfamiliar, but the logic behind it will, I hope, become clear in the body of the thesis.
2. **Commitment**. Individuals may have a stake in others' payoffs that persists independently of punishments or rewards. This stake is a consequence of

*structural or motivational dependence* on the payoffs of others or the group. These comes in two general types.

a. *Self-interested* individuals' payoffs may *structurally depend* on others' payoffs. For example, if the dictator's income were tied to citizens' incomes via taxation, and he/she was motivated only to maximise personal income, this would produce a motive to increase economic prosperity via interventions such as property rights.[21]
b. *Non-selfish* individuals may have *preferences*—benign or malign—over others' payoffs or group performance. For example, parents care for a child's welfare, nationalists have regard for national glory, Smithian optimisers are sympathetic to their fellows, and fanatics may delight in non-believers being "dashed against the rock".[22]

Actors who cooperate because of rules (punishments and rewards) are *mercenaries*. They have no underlying commitment to the group, and are unreliable opportunists where rules fail.

Actors shaped by commitment (structural payoff dependence and motivational payoff dependence) are *committed actors*. They retain an interest in strategies that benefit the group, even where rules are exploitable or absent altogether.

I represent these forces with a utility function that includes private payoffs plus some weighted function of others' payoffs. There is a long tradition of such functions. Other-regarding commitments are informal in Smith (1759/1976) and Hume (1739), and formalised in Edgeworth (1881), Fehr and Schmidt (1999), and Bolton and Ockenfels (2000). Structural payoff-dependence is informal in Hobbes (1651) and formalised widely, with the most important instance for the argument being Olson (1993).

**Rules**, as just noted, include *institutions* such as property rights, and also include *individual-level* strategies, such as Tit-for-Tat.[23] They take the form "if X, then do Y". Rules are:

[21] Note that a dictator may also hold the non-selfish preferences described in point b.
[22] Psalms 137:9.
[23] Tit-for-Tat is a strategy for inducing cooperation in repeated games (e.g. the repeated Prisoner's Dilemma), where the player adopts a simple "rule": if the other player defected last round, then defect; if they cooperated, then cooperate. See Alexrod & Hamilton (1981).

- *strategy-dependent,* in that their effect depends on the successful execution of strategies for detecting that X has taken place, and for enforcing Y (North, 1986). This naturally leads to the possibility of concealing X or evading Y.
- *complex* and *narrow* in scope, in that myriad detection and enforcement strategies are required to align the myriad strategies pursued by political and market actors (i.e. harmful extractive strategies and underprovided public goods).[24] Achieving completeness is costly and frequently impossible.
- *adversarial,* in that the targets of rules have incentives to innovate counterstrategies for evading, exploiting, and capturing the rules—to escape due punishments or receive undue rewards.

Rules may be private goods or public goods. Most importantly for our purposes, *institutions are public goods*.[25]

**Commitment,** whether in its aforementioned structural or motivational forms, is, in the main cases of interest, the product of features of the world or preferences that are outside the strategic game.[26] Commitment is:

- *strategy-independent*, in that its incentive effect usually does not require any actor to successfully execute a strategy;
- *simple,* and ranging from *broad to complete* in scope, in that just one dependence mechanism can align most or even all individual payoffs; and
- *non-adversarial*, because the affected individual cannot, or will not, resist that payoff alignment.

### Some illustrations of rules versus commitment

To make these differences between rules and commitment more tangible, first take the example of two parents who are exceptionally "efficient" in nurturing the physical, intellectual, and emotional development of their child.

[24] On the specificity of rules, consider that every novel means of violating property rights, for example, requires novel adjustments to the detection and enforcement apparatus. On completeness, the general impossibility of complete contracts is well-known (e.g. Hart, 1995).
[25] Some individual-level rules, like Tit-for-Tat, are private goods (i.e. they are privately rewarding in the repeated Prisoner's Dilemma). Institutions are public goods and it is privately costly to contribute to them; their properties are described further in Chapter 9.
[26] That is, they tend to be neither strategy-dependent nor vulnerable to counterstrategies.

- If the parents are mercenaries, we would require an extraordinarily complex regulatory apparatus to administer rewards and punishments for each incremental word, gesture, and activity in the process of nurturing care. Achieving completeness would be impossible: good parenting is too complex to codify. These mercenary parents are also innovative, and over time learn ways to game, or perhaps even capture, the rules so they may reap rewards while neglecting their charge.
- If the parents have regard for, and are committed to, their child's welfare, rewards and punishments are superfluous. The mechanism motivating cooperation is complete and innovation-proof: every benefit for the child is a reward for the parents. This will remain true regardless of technical and strategic change. There is no sense in which the parents can escape, evade, or exploit their other-regarding preferences.[27]

We will not be concerned with familial preferences, but the example is useful for building intuitions about commitment born of other-regarding preferences.

For a second example, this time of structural payoff-dependence, take Olson's (1993) stylised self-interested dictator (noting that authoritarians may have more complex moral and cultural preferences in practice). All dictators depend on domestic economic output for their tax income; if rational, the self-interested, Olsonian dictator provides a range of public goods such as rule of law, roads, and so on, that raise the general level of prosperity. The key point is that we do not need any system of reward and punishment to motivate the dictator to provide public goods or to invest in reforming and repairing the institutional structure, where this increases expected tax income.

## PART II: Rules and commitment in the innovation race

### Innovation and the Red Queen's race

To this foundation I add uncertain and sometimes revolutionary Schumpeterian innovation. Like Baumol (2004), Azar Gat (2006), Acemoglu and Robinson (2019), and several others, I borrow from biology the concept of the *Red Queen's race*. Much innovation is adversarial. Competitors perpetually search for strategies and

[27] It is incoherent to "prefer different preferences".

counterstrategies that will give them an advantage—in markets, in politics, and in warfare. Van Valen (1973) coined the term "Red Queen's race" with reference to Lewis Carroll's *Through the Looking Glass* (1897), where the Red Queen and Alice compete in a race where they run but remain in the same place:

> *"Well, in* our *country," said Alice, still panting a little, "you'd generally get to somewhere else—if you run very fast for a long time, as we've been doing." "A slow sort of country!" said the Queen. "Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!"* (p. 50)

In the adversarial innovation race, competitors must run as fast as they can in order to stay in the same position *relative* to one another. However, because innovation is unpredictable, in practice the race results in *stochastic gaps in the relative performance* of competitors over time. In a period of innovation that is sufficiently lucky for one racer and/or unlucky for the other, the gap grows large enough that one racer is eliminated. We can see why Gat (2006) would use this to describe arms races, Barnett and Hanson (1996), Derfus et al. (2008), and Haldane (2012) the innovation races between firms, and Acemoglu and Robinson (2019) the contest between government and civil society. Baumol (2004) applied it to all these cases. Others have framed problems of pharmaceutical firm regulation and misinformation as innovation races, matters discussed in later chapters.

A key aspect of the Red Queen's race, and an implication of Schumpeterian innovation, is that there are no "unbeatable" strategies. One must therefore keep running and responding to the latest moves. If a racer stops innovating in an antagonistic contest, strategic change will be unfavourable and will tend to favour monotonic decline. I call this the *Tortoise and Hare principle*, after the fable with the same lesson: falling asleep in the middle of a race guarantees loss.

Finally and relatedly, the probability of victory/loss in the Red Queen's race can be predicted based on the racers' *relative innovation capacities*. This is the capacity for

actors to search for and identify useful strategies.[28] It interacts with complexity in ways that become important a little later.

### Rules as strategies: The institutional Red Queen's race

Actors who contribute to aligning payoffs via rules are given the label "regulatory actors". As will become clear later, they are not only state actors working via institutions; the category will include market actors when contributing to the total regulatory structure.

All rules, including institutions, are within the strategic game. They are strategies vulnerable to counterstrategies. A Red Queen's race takes place between actors that work to align payoffs via rules—regulatory actors—and those who seek to evade, exploit, or capture the rules. Where rules are important for driving cooperation, failure to run this race—playing the sleeping Hare—guarantees the collapse of cooperation. This is akin to Myron's law: "Asymptotically, any finite tax code collects zero revenue" (Romer 2010, attributed to Myron Scholes).

### Innovation makes purely rules-based orders technically impossible.

This leads to an *impossibility theorem for purely rule-based orders*. All such orders require either structural or motivational dependence, i.e. an *exogenous commitment* to restoring that order. Smith, Arrow, Hirsch, North, Polanyi, and Schumpeter were correct in supposing that the liberal order has foundations in other-regarding preferences.

Here, I show that the complementary relationship between "commitment" and "rules" is a fundamental logical necessity for any system of large-scale cooperation. It is equally necessary in human, sociobiological, and artificial agent systems.[29] I propose it as a natural law of social order.

Why? The rules that drive large-scale cooperation are public goods.[30] The only way mercenaries will supply them is if they are incentivised by that same rules-based

[28] Innovation capacity increases with "trial number", i.e. the number of strategies one can experiment with, and "trial quality", i.e. the extent to which experiments are directed randomly versus rationally. A "sleeping Hare" has zero innovation capacity. This is explained further in Chapter 5.
[29] A conclusion explained at the end of Part 2, in Chapter 8.
[30] In the cooperation literature, a "large" group has more than twenty actors. We are concerned with groups of many thousands or millions.

order. The problem is that we always end up needing the mercenaries to step beyond the rules, to run the race on behalf of the institutional order even when they have no incentive to do so. There are two reasons:

1. The first is a completeness problem. Within the Red Queen's race, the system of rules will be compromised in unanticipated ways. *Novel* public goods are required, over time, that can restore the original equilibrium from any possible corrupted equilibrium. It is impossible to codify the supply of all these possible public goods in advance, and mercenaries will not supply uncodified goods.
2. The second is a circularity problem. Any rules-based repair mechanism would itself be part of the strategic game. When the repair mechanism is compromised, it cannot repair itself, and any additional layer of rules we add to solve the problem suffers from the same weakness.

I show that alternative approaches, such as balances of power or Coasian bargaining (Coase, 1960/2001), provide no solution.

Social order requires a *universal repair mechanism* that can motivate individuals to repair the system (and so provide public goods) in ways that cannot be specified in advance. That mechanism must be *innovation proof* and so external to the Red Queen's race. Commitment (i.e. structural/motivational payoff alignment) is such a mechanism. It is complete, in that committed actors will value any strategy that improves group outcomes, and there is no sense in which those strategies need to be specified in advance. Again, it is also strategy-independent and therefore innovation-proof.

In the long-run, the strength of this commitment *anchors* the efficiency of the rules: with zero commitment, no institutions survive and we get anarchy;[31] with perfect commitment, actors make socially optimal investments in the system of rules.

"Closed" games are rare in practice; almost every real-world game is an "open", innovative game. A general property of these games is that there are no long-run Nash equilibria. There are still equilibria in such a world, or at least points of

[31] Only individual-level rules that are private goods, like Tit-for-Tat, can be supplied in cases of repeated action in small groups (two to twenty players).

attraction, but they must arise from a source of payoff-alignment that is external to the Red Queen's race. That is, the only equilibria are those set by one of the two varieties of commitment "anchor".

### Commitment without rules: Instability and inefficiency

Finally, commitment by itself cannot drive stable and efficient cooperation. First, committed actors who do not employ rules are effectively unconditional cooperators; a standard result in the literature is that such actors are vulnerable to invasion by defectors and free-riders, who benefit from public goods but do not contribute to them, or who parasitise others without punishment. Second, while committed actors may intend to contribute to group optimisation, commitment itself does not assist with solving complex allocation problems—with identifying the best ways to cooperate.

Rules address both of these problems. First, they control defectors and ensure that committed actors are not disadvantaged. Second, rules can be used to create systems of *domesticated competition* (termed "Darwin machines" after Calvin, 1987, Plotkin, 1994, and Wilson et al., 2014, and introduced in Chapter 5), such as the market and its sociobiological analogues, to optimise resource allocation and utilise distributed knowledge.

The result of this abstract framework is that in the production of social order, **rules and commitment are essential complements** (Figure 1.1). Each fills in for the weaknesses of the other. This result is the natural consequence of the Red Queen's race.

## INTERLUDE: Applying the framework to the market society

### Innovation in the market society: The Red Queen's relay

The framework is complete and now we prepare to apply it to the cases of democratic and authoritarian market societies. The dominant explanation of social order in the social sciences is of self-interested actors working within *institutions*. A key goal is to see how far self-interest and institutions can take us.

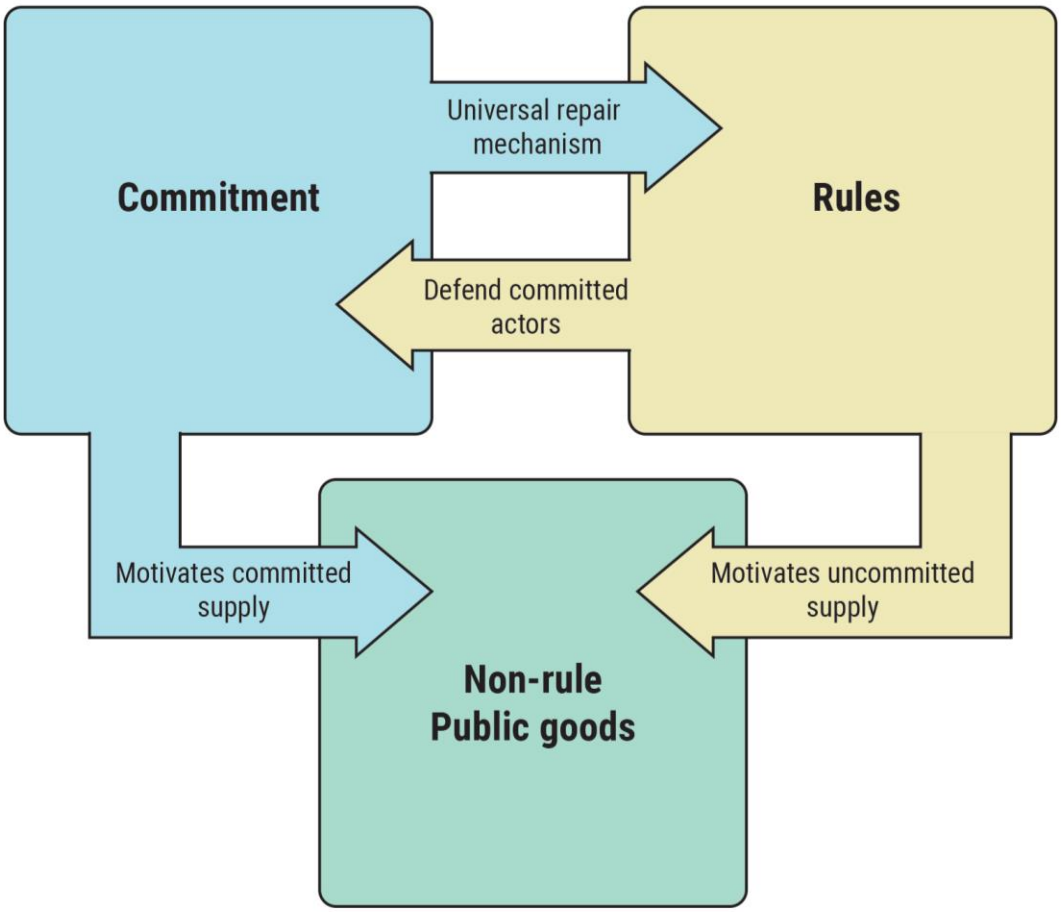



**Figure 1.1** Commitment and rules are essential complements. Commitment is an exogenous motive for providing rules, given the impossibility of a self-repairing rule-based order. In turn, rules support the maintenance of commitment by penalising defectors. Both rules and commitment motivate the provision of various other economic and political public goods.

The institutional structures of market societies can be understood as arranged in a hierarchy (as a multi-layer principal-agent system as in Peltzman, 1976). At the bottom, people exchange goods and services within the rules of the market. In the middle layer, various institutional actors develop and police market rules, and also police one another. At the top layer, the meta-institutional layer, we have the policing of the rule-makers. The payoffs of rule-making actors may themselves be shaped by formal political institutions, but also informal and/or non-institutional factors (financial co-optation, threats of violence, and so on).[32]

[32] "Regulatory capture" may occur with or without changes in formal rules, and indeed the meaning of formal rules changes over time with strategic innovation.

In an orderly regime, whether democratic or authoritarian, motives at the meta-institutional level (e.g. of voters, an oligarchic clique, or a ruler) are channelled from the top of the hierarchy down to its lowest levels. For example, voters or a ruling elite shape the incentives of the regulator, such that it develops and enforces property rights, and then market actors follow these incentives.

In this, we have a first order, rules-based explanation of social order.

Every level of the institutional hierarchy must function well to produce prosperity. Because all rules are strategies with counter-strategies, the use of rules throughout the hierarchy creates three main adversarial contests—three Red Queen's races—in the market society:

1. the race between market actors;
2. the race between market actors and the regulator; and
3. the meta-institutional race for control of the regulator.

These are the three legs of the **Red Queen's relay**. The prosperous market society, whether democratic or authoritarian, must set a good pace in each leg.

## PART III: The technical problem of economic management

Part 3 is concerned with the purely technical problem of managing the Red Queen's race in markets, temporarily setting aside the problem of regulator motivation. Following standard analyses of the purely positive problems of regulation, I assume a benign regulator and take market actors as self-interested "mercenaries".

Hayek pointed out, following Barone (1908) and Pareto (1906/1971), that the neoclassical equilibrium is theoretically consistent with anything from anarcho-capitalism to central planning. We arrive at some equilibrium because of a concrete system of rules, a combination of various detection and enforcement strategies that align private with social payoffs. Neoclassical theory is silent on the concrete processes that produce efficiency; it "systematically leaves out what is our main task to explain" (Hayek, 1945, p. 530).[33]

[33] For Hayek (1945), as for us, the problem to engage with is the "unavoidable imperfection of man's knowledge" (p. 530).

This system of rules—which I term the "total regulatory structure"—is always composed of some *division of detection and enforcement labour*. The total regulatory structure is always a *hybrid* of institutional and individual-level detection and enforcement powers. State regulation and individual-level "rules" may be perfect substitutes, essential complements, or one may dominate the other. Central planning uses institutional detection and enforcement more intensively, and free markets use individual-level detection and enforcement more intensively. *A priori*, either system may be efficient or disastrous.

The optimal division of labour between market actors and state institutions depends on their actual capacities—on their *comparative advantages* in detection and enforcement—and appropriate combinations of their respective powers. As Hayek remarked, the all-important question of these comparative advantages has been left largely to assumptions.

It is dangerous to proceed on the basis of assumptions about actors' capabilities. The ideal system, or at least the range of workable systems, is substantially determined by the observed pattern of actual capabilities. It is pointless to fervently hope that certain capacities exist, to try to will them into being, and to articulate and rearticulate the desirability of the orderly systems we could build from them if they indeed existed. They are constraints that we must discover and work within.

We investigate those capacities in the context of the Red Queen's race. All components of the total regulatory structure—both contributions of market actors and state regulators—involve strategies vulnerable to counterstrategies. The contest over market actor detection and enforcement creates the first leg of the Red Queen's relay. The contest over institutional detection and enforcement creates the second leg. These two legs are the subject of Part 3.

**The first leg, the race between market actors,** is invisible in a purely "rational" model. But we are dealing with Schumpeterian innovation in a Red Queen's race, and all the violations of perfect rationality that entails.[34]

[34] With "perfect" rationality implying either that actors actually optimise, or that they have rational (on average unbiased) expectations. Both are inconsistent with Schumpeterian innovation, as quoted in Section 1.2.

In the world of the Red Queen's race, bounded rationality creates an *adversarial innovation race*. There is a race between market actors who aim to measure private payoffs, and those who aim to conceal or misrepresent those payoffs. As in all innovation races, differences in *innovation capacity* tend to predict race outcomes.[35]

Relative innovation capacities evolve over time. Today's markets pitch consumers with largely static capacities against firms' rational-scientific innovation systems, composed of armies of marketers, psychologists, and analysts increasingly equipped with big data and machine learning techniques. At the limit, firms may acquire an almost complete understanding of their customers' weaknesses. The converse will not be true.

These asymmetries bite in *complex* markets,[36] and the resulting harms are most severe where markets are welfare-critical.[37] The framework suggests an economic law: *in complex and welfare-critical markets, asymmetric innovation capacities guarantee crisis*. I examine the consequences of allocating market actors impossible tasks in two complex markets in the 21st century: financial markets (producing the Global Financial Crisis) and pharmaceutical markets (producing the opioid crisis).

We posit a general principle of regulatory design: in complex markets, we must manage asymmetries in innovation capacity via strategies that reduce complexity. In short, either we allocate market actors payoff-detection tasks[38] that are achievable, or we accept a capitalism with large inefficiencies (in welfare-significant markets) and regular and severe calamities (in welfare-critical markets).

**The second leg, the race between market actors and the regulator,** is another adversarial innovation contest. Innovating firms may create (A) first-leg problems (asymmetric innovation capacities) or (B) pure incentive problems, and both require the regulator to play catch-up. The expected outcome of the race again depends on *innovation capacity* and *complexity*. In a useful general metaphor,

---

[35] This produces differences in the likelihood, in a given period, of identifying successful strategies for outplaying an opponent.
[36] A matter discussed in Chapter 11. Consider that marginal investments in rationality continue to improve performance in complex games such as chess or Go, but rapidly hit a ceiling for tic-tac-toe.
[37] Welfare-critical markets powerfully affect individual outcomes and/or the quality of important public goods. For example, the private consequences of healthcare, or systemic effects in financial markets.
[38] Recall that rules are composed of strategies for detecting that X has taken place and enforcing Y, and that every society is a hybrid of individual and institutional detection and enforcement.

Heimann (2000, p. 332) describes this as the problem of "bloodhounds chasing greyhounds". In complex markets, firms generally have advantages in innovation capacity over the regulator, and race ahead.

The regulator can control the pace of this Red Queen's innovation race by its choice of regulatory strategy. I distinguish two approaches to regulation, which are poles on a spectrum:

- *Marginal regulation*. The marginal regulatory structure maximises theoretical allocative efficiency. It is maximally permissive and structureless, allowing market actors maximum choice. It aligns payoffs via actors—whether market actors or state regulators—identifying the marginal payoffs of strategies, and responding accordingly. It is complexity-increasing, and so increases incentives for investment in gaining asymmetric advantages in innovation capacity. It therefore intensifies the Red Queen's innovation arms race.
- *Structural regulation*. The structural regulatory approach uses blunt tools and broad, bright lines to demarcate acceptable domains of economic activity. Its purpose is to simplify detection tasks so that they are achievable. It sacrifices theoretical allocative efficiency in order to achieve actual efficiency, by making it possible to keep up in the Red Queen's race. To extend Heinmann's (2000) metaphor, the regulatory bloodhounds have the power to structure the topography of their race against the greyhounds. The goal of structural regulation is to limit complexity so that they need only guard narrow passes, rather than chase in an unlimited, multi-dimensional space.

The analysis is of very broad application. I consider applications from basic property rights, to intellectual property, mineral extraction, and tax policy, before paying particular attention to two main cases: the Global Financial Crisis and opioid crisis. A third case, of political information markets, is explored in Part 4.

Market actors and state regulators are *essential complements*. Marginal regulation by market actors is essential for efficiency; distributed knowledge gives them advantages especially in markets for private goods. Structural regulation is used to simplify detection tasks, by a variety of means discussed in the text, so that market actors can play their role. Only then does voluntary exchange work its magic.

**Table 1.1** The complementary roles of marginal and structural regulation

| | **Marginal regulation** | **Structural regulation** |
|---|---|---|
| **Market actors** | Identify marginal private payoffs, especially for private goods/services | N/A |
| **Institutional actors** | Identify marginal group payoffs, especially for public goods | Reduce complexity so that marginal regulation can function |

**I conclude with an "impossibility theorem" for both central planning and libertarianism (and Hayek's classical liberalism).** The argument is Hayekian in nature, but finds that Hayek's own free market system is the mirror image of the flawed system of central planning. Both are non-adaptive systems that commit on an *a priori* basis to using a particular combination of institutional and individual-level detection and enforcement. They commit to doing so in perpetuity, in a world where detection and enforcement tasks change over time, and so the comparative advantages of institutions and individuals change over time.

Hayek brilliantly pointed out the error of central planning: it neglects the limits to institutional actors' actual capacities to detect payoffs. Libertarianism and Hayekian classical liberalism, however, make a symmetrical error: they neglect the limits to market actors' actual capacities to detect payoffs. Both systems are justified by abstract, logical models that simply presume the necessary capacities exist. Empirically, the necessary capacities do not exist. Both systems generate predictable failures: inefficiency and discoordination in central planning; epidemics of rent-extraction and regular crisis in "free markets".

I lay out the "Libertarian's Trilemma" mentioned in Section 3.2. The full explanation is presented in Chapter 13, but the nub is that:

- the goal of (A) prosperity is only possible where we keep up in the Red Queen's race over detection and enforcement; and

- keeping up in the race, in complex markets where market actor detection fails, requires either (B) structural constraints on contracts, and/or (C) a large marginal regulatory apparatus that can act as a substitute for market actors.

Libertarians rule out both means (B and C) of keeping up, and thus commit to losing the race—to playing the sleeping Hare—and sacrificing prosperity (A). The influence of free market theory spurred a "great regulatory complacency" from 1980 onward, and contributed decisively to the weakness of contemporary democratic capitalism.

Hayek's own colourful language is apposite: both central planning and his own classical liberalism are "fatal conceits".

- The conceit of central planning is the idea that it may act as a substitute for market actors. As Hayek observed, its pretence of godlike interventionism requires something approaching omniscience if it is to be efficient.
- The conceit of Hayek's classical liberalism is the pretence of the godlike knowledge that basic market rules are "unbeatable". Firms innovate in myriad ways that the armchair theorist cannot foresee.

## PART IV: Sustaining democracy in the Red Queen's world

**The third leg, the meta-institutional race,** is the contest to control the payoffs of the regulator who forms market rules, and so determine whose interests are served by the rules it makes. Control may be won by modifying institutional constraints on the regulator, and/or wielding resources such as money and violence to alter the meaning of those constraints in practice.

I begin with a purely self-interested model of authoritarianism and democracy, to see how far it can take us. This will not be the preferred model, but a stepping-stone.

Authoritarian regimes based in self-interest are explicable in the framework of this thesis. Following Part 2, social order emerges from commitment and rules acting as essential complements. The commitment "anchor" is the ruler's structural dependence on the state for tax income (their Olsonian encompassing interest). This innovation-proof motive is forever channelled into institution selection. In turn, institutions for co-optation and repression maintain the ruler's commitment to

economic output, by securing regime survival. Self-interest is sufficient to build a working authoritarian social order.

Olson (1993) argues that the dictator's structural payoff dependence creates incentives for efficient institutions. This is true, but the primary role of institutions is securing regime survival (so maintaining commitment). Rulers are engaged in a complex Red Queen's race with usurpers,[39] and use "structural" institutional strategies to simplify that race. These strategies are frequently crude and severely undermine allocative efficiency. Even then, rulers cannot halt the race and, on average, are dethroned after a decade. In a world of self-interest, authoritarianism is possible—but neither the ruler nor their subjects may sleep easy.

**Box 1.3 Authoritarianism beyond material self-interest**

Real-world dictators may have motives that go beyond self-interested accumulation, such as national glory, moral or cultural diffusion, or even the welfare of the people. Insofar as the realisation of these other motives depends on tax income and efficiency, however, the model of Olsonian structural payoff dependence remains illuminating.

Empirical evidence on regime survival contradicts the self-interested model: authoritarian systems based in powerful ideological commitments ($r_c > 0$) are outliers, with greatly increased survivability. The framework provides an obvious explanation, where ideological commitments among elites and the citizenry act as a partial substitute for rules—for fragile mechanisms of co-optation and repression—and so help win the Red Queen's race against usurpers.

Democratic regimes composed purely of mercenaries are, in contrast, anchorless. No citizen has incentives to provide the public good of a *vote* that is *informed* and *in defence of democracy*. As Downs observed, the vote has negative expected value. If we bring in the *deus ex machina* of a zero-cost vote, its value is positive but miniscule.[40] Yet voters are engaged in a Red Queen's race with special interests for

[39] Again I am assuming self-interested actors in this analysis, and as North (1979) observes, this rules out mass collective action. All internal regime change is by "palace coup" (p. 258)—hence usurpers.
[40] E.g. participating in an election in a population of one million, where victory is worth $1000, would be worth less than one cent. Discussed in Chapters 14 and 16.

regulatory control, and we need them to invest in counterstrategies and run the race. They will not do so. Finally, voter-mercenaries will minimally resist, and will in fact participate in, the dismantling of democracy. Disenfranchising others with different interests, for example, increases the value of democracy and the vote.

I add that mercenary judges, and a mercenary military, are unreliable defenders of constitutional constraints and will opportunistically join in rent extraction.

**Mercenary democracy is the anarchy before the Leviathan.** It is a contest between special interests with the incentives of Olson's (1993) roving bandits,[41] with no actors committed to resisting them. Social order will only recrystallise when an innovation-proof source of payoff alignment emerges. One will always emerge: the roving bandit has a natural incentive to build a winning coalition sufficient to become a stationary bandit.[42] Thus, as Thomas Hobbes (1640/2019), observed, the democracy of self-interested actors is an interregnum before the return of the Leviathan.

> *[A] democracy is by institution the beginning both of aristocracy and monarchy*. (p. 22)

So it would be in any world of purely self-interested actors.

### Moral sentiments: Updating Smithian sympathy

The framework of this thesis suggests that Hobbes' analysis of the self-interested society is correct. Yet in practice, Hobbes was wrong: democracies have survived and thrived. The rest of the thesis addresses this puzzle.

Democracy rests on commitment. Lacking structural payoff dependence, there is only one alternative: democracy must rest on motivational payoff dependence, i.e. other-regarding preferences. That other-regarding preferences are a *logical necessity* for democracy is one of the key analytic results of this thesis.

[41] "Roving bandits" move from victim to victim, so unlike the "stationary bandit", they have no incentive for restraint in extraction. They have roughly zero structural payoff dependence. The roving bandit is a metaphor for any actor whose share in total societal output is small, so they have no incentive for restraint in extracting from that society. All actors in democratic capitalist societies have a small share and, as far as *material* incentives are concerned, are roving bandits.
[42] Just as failed states are soon occupied by competing local warlords—small-time stationary bandits.

Adam Smith viewed markets as nested within a broader social order that rested on a form of other-regarding preferences he termed "sympathy" (Smith, 1759, p.1). Much of Smith's *TMS* presages today's empirical evidence, and the argument in this part of the thesis.

I begin by surveying that evidence. Humans are *strong reciprocators*. Strong reciprocators will pay costs to benefit others, on the condition that those others are also cooperative. My commitment is conditional on your commitment. Unlike unconditional altruists, strong reciprocators refuse to be "suckers" and defect in response to cheaters. They will, in fact, pay costs to punish cheaters—and so they are the kind of being that will invest in institutions as public goods. For Smith (1759/1976), strong reciprocity is "the great law which is dictated to us by nature" (p. 4): the attitude that benevolence is due to benevolence, and punishment due to harm.

Strong reciprocation is a common equilibrium strategy in nature. A large game theoretic literature shows that strong reciprocation is the optimal strategy where there is between-group competition, but also the threat of within-group free-riding.

The commitment of the strong reciprocator hinges on *trust* that others are so committed. We may therefore expect trust to predict macro political and economic outcomes. Indeed, a large empirical literature shows that trust predicts economic growth, quality institutions, and political stability. Trust is the strongest correlate of growth in multi-variable cross-national studies, and the relation is causal. Contracts are incomplete, investments carry risk, and contributing to democratic life is costly with little private reward. Trust with strong reciprocation solves these problems.

**Other-regarding preferences have vulnerabilities and a dark side.** First, because cooperation is conditional on trust, a society of strong reciprocators can become caught in a low-trust trap where distrust is self-reinforcing: distrust leads to defection, and defection increases distrust (e.g. see Cohen, 1999, Levi, 1998). Well-designed institutions prevent defection and sustain trust, while poorly designed institutions that allow or reward defection will see trust decline.

Second, other-regarding preferences are often *parochial*. Trust and cooperation are bound up with cultural identity, and vary according to cultural distance. The greatest threats to social order, Smith (1759/1976, p. 71) wrote, are "faction" and "fanaticism",

and narrow group affiliations are indeed among the most significant forces of human history. Liberalism, with its insistence on equal treatment of individuals in the polity, is novel in its (at least within-nation) universalism, and its success and persistence will need to be explained.

In what follows, I first assume the liberal (i.e. universal) pattern of other-regarding preferences and later allow for parochiality. Note that truly universal (i.e. global) other-regarding preferences, and the global public goods they may support, are beyond scope; this thesis is focused on explaining social order at the level of the state and below.

**How does democracy leverage weak other-regarding preferences into broad prosperity?** The analysis so far presents a puzzle: the framework indicates that commitment "anchors" the institutional order within the Red Queen's race, yet like Adam Smith and following the empirical evidence, I expect other-regarding preferences to be relatively weak in everyday conditions (e.g. outside of wartime). How can a weak anchor create an efficient social order?

The answer is that democratic institutions function as a *fulcrum* for leveraging weak other-regarding preferences. There are two fulcra: electoral institutions, and the institutions that create "professions". The design of each works to nearly eliminate the private returns to actors' choices, so that, even if weak, other-regarding preferences dominate the choice function.

**Electoral institutions are the first fulcrum.** Following Downs' paradox, large elections reduce the private benefits of voting to effectively zero. In the analysis here, Downs' paradox is not a bug, but a means of getting work out of other-regarding preferences: the decision to vote is sociotropic. Following Edlin et al. (2007), I show formally how other-regarding preferences resolve Downs' paradox.[43] A problem in models with altruism, however, is that voters frequently behave like martyrs. Adding strong reciprocation solves this problem.

Having discovered the willing voter, the analysis turns to Olson's model of democracy, and his question of how the winning majority will treat the minority. In Olson's self-interested model, the voter monolith values only economic efficiency.

[43] Also see Coate & Conlin (2004) and Fowler & Kam (2007).

Like a dictator, it has zero interest in welfare or liberal values, and the majority may be profoundly extractive of the minority so long as doing so is efficient; Olson's superencompassing interest is fully consistent with a slaveholder democracy.

Weak other-regarding preferences solve the problem; they are enough to prevent egregious forms of extraction. They do not support intensive redistribution, but only to the neediest cases. Higher-trust societies with strong other-regarding preferences are expected to redistribute more, and invest more efficiently in public goods.

**The "profession" is the second fulcrum,** and it is required to provide a range of other democratic public goods (as well as classical public goods) that cannot be provided via elections. The archetypal case is the judiciary. The judiciary is inexplicable in the workhorse economic model: in healthy democracies we trust it to provide some of the most complex and profoundly important public goods, while it is largely insulated from punishments and rewards for doing so.

Mirroring the electoral case, and the upside of Downs' paradox, it is the weakness of private incentives that makes the judiciary possible. It creates the fulcrum I call a *profession*. Professional motivation is developed by structural regulatory interventions, the most important of which is *extrinsic insulation*. This entails actors' isolation from private incentives and competition, such that decisions are then made on the basis of social payoffs.

Why must some public goods be provided via professionalisation? Because they are degraded when governed by externally imposed rules. There are two reasons:

- a codification problem. Some public goods are too complex and context-sensitive to codify. In the case of the judiciary, for example, its function of flexibly realising the "spirit" of the law cannot be codified but requires judgement.
- a political influence problem. Some public goods, especially those that hold regulatory actors accountable, require independence from the rule-makers. For example, the judiciary's independence from lawmakers is central to its function of holding lawmakers accountable.

As an aside, the first of these, the codification problem, is common to many complex public goods including healthcare, education, legal services, and science (including

the practice of political economy). I observe that attempts to increase efficiency via incentive structures and competition can have unintended consequences, eroding commitments and leading to deprofessionalisation and the degradation of the public good supplied.

**Political information markets are the remaining problem, and professionalisation is part of the solution.** Downs' (1957) primary concern was not turnout, but that voters would be uninformed. In the analysis, the key insight is that information *markets* fail to harness either of the fulcra: weak other-regarding preferences cannot be leveraged into informed voters, and the public good will not be provided.

Voters are engaged in a Red Queen's race with special interests over information. Weakly other-regarding voters (A) invest too little in the race and, even setting motivation aside, voters (B) face the "first leg" problem: political information is hyper-complex, and special interests have large advantages in the race.

From Part 3, we have a ready-made solution: *simplification*. Citizens must be allocated detection tasks that they can accomplish. This means intervening in the supply-side of information, so that *cheap* information is *reliable* information. Yet this confronts the two problems discussed in relation to the judiciary: it is not feasible to *codify* acceptable output for political media, and political media must be *independent* if it is to hold regulators to account.

The solution is professionalisation—that is, structural simplification of the detection problem by creating the conditions for trust. I observe that the professionalisation of journalism in the post-war period meant that public trust in media was rational; accordingly, rates of trust were extremely high compared to today. I examine some means for re-professionalisation, though a full solution is complex and beyond scope.

**Parochiality[44] corrupts these two democratic fulcra and is the greatest threat to the liberal order.** While benign other-regarding preferences ease all legs of the Red Queen's race, Smith's "faction" and "fanaticism" intensifies them. I begin by examining the effects of parochiality on democratic choice. Parochialists

[44] Again, as described in Section 2.4, by "parochiality" I mean intensive, within-country social cleavages, not the nearly ubiquitous fondness that people hold for their own hometowns, countries, and cultures.

will pay private costs to supply *narrow* public goods (i.e. in-group-beneficial public goods that may be net socially harmful, sometimes grossly so).

With respect to voter choice, parochialists value transfers from the out-group to the in-group. If the in-group is valued positively and the out-group is weighted zero (i.e. mere indifference), this is enough to motivate the grossest forms of collective extraction. The great horrors peppered throughout human history—slavery, war, genocide, and so on—depended on narrow public goods that self-interested actors would not supply.

A feature of parochial politics is the *primacy of victory*. Victory produces a *double-dividend*: the avoided costs of the out-group's extractive platform, plus the benefits of the in-group extractive platform. Its magnitude may be very large. The greatest fear of the parochialist is that the loss of the double-dividend may be *permanent*, given the group that holds power may modify meta-institutions.

Parochiality corrupts professions. Parochial journalists prefer to produce strategically favourable (mis)information that benefits in-group candidates and special interests, even in the total absence of private inducements to do so. Judges, police, teachers, doctors, scientists, and others may depart from the values of the profession wherever their activity relates to the between-group conflict.

The primacy of victory also corrupts the role of information in democracy. The *strategic* effect of information, its contribution to in-group power, matters far more than its truth. Parochialism leads to information markets awash with misinformation that has purely strategic value.

In turn, the double-dividend becomes the *strongman's dividend*. Parochial voters will accept corrupt leadership, so long as its costs are lower than the double-dividend. The enormous size of the latter provides enormous cover. Indeed several empirical studies have found that partisanship provides cover for corruption and reduces sensitivity to economic performance (e.g. Eggers, 2014).

All this, I suggest, is why empirical studies find such a strong effect of generalised trust, and between-group trust, on economic and political performance.[45]

## CONCLUDING: Democratic capitalism as a trust-building machine

This thesis argues that under Schumpeterian innovation, purely rule-based social orders are logical impossibilities. All are anchored by at least one of the two forms of commitment. In turn, committed actors devise rules to protect their communities from defectors, and so sustain the conditions for commitment.

Liberal democracy and efficient markets depend on Smithian moral sentiments—on other-regarding preferences. For the strong reciprocator, commitment to others depends on trust that others are so committed. The primary role of rules is to sustain trust, as the foundation of the social order; this requires ensuring that cheaters do not get ahead, and that prosperity is broadly shared.

The analyses in Parts 3 and 4 are therefore connected. Complacency in running the Red Queen's race in markets since the 1980s has had two predictable outcomes. First, it has permitted rampant innovations in rent-seeking, produced severe crises, and led to a secular decline in the capacity for capitalism to deliver for ordinary citizens. Second, worsening outcomes produce distrust; they indicate to voters that the compact between social groups is no more. Even if this failure were purely a matter of technocratic incompetence, poor outcomes lend support to conspiratorial interpretations and create tinder for between-group conflicts.

I briefly point to a large body of evidence showing that inequality and economic crises have, over the past century, repeatedly fuelled populism and parochialism. Interventions that reduce inequality have the opposite effect.

This is a variation on the theme of Karl Polanyi (1944), who drew lessons from the fall of Vienna—and its turns towards fascism and communism—that were quite different to those drawn by his fellow Austro-Hungarian, Hayek. The political movements of fascism and communism sprouted in the extreme hardships of nineteenth century capitalism, and flowered during the Great Depression. As McRobbie and Polanyi-Levitt (2008, p. 398) summarise, Polanyi believed that "the

[45] Noting that the various forms of trust—social, political, and institutional trust—are quite well correlated and tend to move together.

greatest threat to freedom was a poorly administered economy". And in Polanyi's terms,

> *[t]he fascist solution of the impasse reached by liberal capitalism can be described as a reform of market economy achieved at the price of the extirpation of all democratic institutions…* (Polanyi, 1944, p. 237)

It was the promise of a fair deal that rebuilt trust, allowed democracy to survive, and made the post-war Golden Age of Capitalism possible.

**To conclude, there will be no "end of history", no self-repairing constitutional clockwork, and no self-organising market system.** There is no escape from innovation and its fundamental ambiguity. We forever journey into the unknown, and there are always new moves to be played.

If the rules-based order is to adapt, we need those who hold power to be committed to restoring a cooperative equilibrium. There are three "equilibria":

- absence of commitment (or commitment only to the self), anchoring anarchy and Thomas Hobbes' "warre of all against all" (1647/1983, p. 34);
- commitment born of structural payoff dependence, anchoring authoritarianism; and
- commitment born of other-regarding preferences, anchoring democracy (and stabilising authoritarianism).

Liberal institutions and liberal other-regarding preferences are essential complements, in the vein of Arrow's speculations. This is a special case of the general theory developed in Part 2.

Commitment alone is impotent; adaptation requires that the committed rule-maker is also competent. Competence requires abandoning systems that are allocatively efficient under empirically invalid assumptions—central planning and "deregulated" free markets—and confronting the concrete, and changing, reality of actors' capacities. In market societies, we must be sensitive to market complexity and asymmetries in the power to innovate. Markets only work if they are structured so that market and institutional actors can actually accomplish the regulatory tasks we allocate to them.

Competent management of democracy requires institutional structures that harness the professional motive. Professions provide essential public goods that can neither be supplied in free markets, nor safely or effectively codified in a system of incentives. To place the judiciary or political journalism in the market, in a world of Schumpeterian innovation, is to invite the end of democracy.

Finally, humans seek group affiliations. We can choose to create a system of rules that is worthy of trust—because it delivers broadly-based prosperity—or we will inadvertently choose parochial backlashes and authoritarianism. The rules we create must reinforce commitment, so the democratic capitalist society must be a trust-building machine.

Liberal societies have slept at the side of the track for a little too long—but put on your running shoes, it may not be too late yet.

# 2

# Building the framework: On the use of sociobiology

This thesis is concerned with long-standing questions about the origins of social order in general, and the conditions that produce stable and prosperous democratic capitalist societies in particular. One aspect of the method merits a preface: I conduct a few targeted raids on the neighbouring discipline of sociobiology. Fortunately, given the non-rival nature of human knowledge, they will make no complaints. For many readers, however, the interdisciplinary nature of this thesis presents some challenges. It requires understanding and working with some unfamiliar concepts, and it raises methodological questions about how exactly we can apply lessons from a field concerned with a different universe of actors.

There is plenty of precedent, including in Milton Friedman's (1953) and Alchian's (1950) evolutionary models of market optimisation, Herbert Simon's (1968/2019) study of the principles of design, and Joel Mokyr's (1998) evolutionary models of innovation. More than two centuries of exchange between these fields, beginning with Smith, are discussed in Section 1. Given the extraordinary advances since, we can make much richer use of parallels between the disciplines today. The approach quite naturally raises puzzlement; Joel Mokyr's readers, for example, wondered whether he was using evolutionary theory as "an analogy, a simile, a metaphor, or a purely intellectual game" (Mokyr, 1998, p. 2). Similar questions were raised by readers of early versions of this thesis, and this short chapter aims to provide an answer.

My answer is broadly the same as Mokyr's and Friedman's. In their cases, it is that "Darwinian selection", in contrast with "natural selection", is a highly general theory that has no necessary connection with biology, fitness, or genes. It is a process that takes place wherever there is some mechanism for generating variation (whether genetic mutation or firm innovation), some selection amongst this variation (whether organism or firm survival), and some means of transmission and retention of the selected variants (whether reproduction of genes or of a firm's management strategy). As Friedman (1953) writes,

> *Confidence in the maximization-of-returns hypothesis is justified by evidence of a very different character… unless the behavior of businessmen in some way or other approximated behavior consistent with the maximization of returns, it seems unlikely that they would remain in business for long… The process of "natural selection" thus helps to validate the hypothesis – or, rather, given natural selection, acceptance of the hypothesis can be based largely on the judgment that it summarizes appropriately the conditions for survival.* (p. 158)

In fact, Friedman here misunderstands the implications of Darwinian selection, which helps explain *adaptation* but does not imply *maximisation* except under conditions that are unlikely to hold in practice.[46] Regardless, he is correct that Darwinian selection helps to explain why entities may be better adapted to their surroundings than we would expect from their level of rationality alone: selection gives us fitter organisms in biology than we would expect from blind change without selection, and better adapted firms than we would expect from bounded rationality without selection.

Economic and biological evolution are each *special cases of a broader dynamic theory* that, in Friedman's (1953) terms, "summarizes appropriately the conditions for survival" (p. 158). As such, it would be perfectly possible to develop an evolutionary theory of economic innovation in a world in which biological evolution had never been discovered, or in which biological evolution did not even exist.[47] Indeed, intimations of Darwinian selection appeared in the social sciences well before Darwin developed natural selection (including in Smith, who Darwin read and regarded favourably; see Darwin, 1875).

The point is further made by the application of evolutionary theory to purely artificial systems, such as simulations of artificial agent populations. These agents maximise neither genetic fitness nor economic utility, but whatever the programmer specifies—they may, for example, vie to provide the best solution to a complex mathematical problem. Again, the programmer relies on some process that generates variation in

[46] Maximisation requires a rate of innovation that perfectly keeps up with external change, and a way to avoid getting stuck on *local* optima in a fitness landscape. The latter is especially difficult, given all search is local to some degree, and the design space is vast—most of it being distant.
[47] Biological scholars might, in such a world, wonder if applying evolution to biology would be making an analogy with economics (it would not).

agents' strategies, selects among that variation (i.e. against performance at solving a problem), and transmission (i.e. the winners copy themselves or otherwise influence the next generation). Needless to say, these computer programmers are not biologists, nor need they have any knowledge of or interest in biology. Yet the invention of evolutionary computation may have been considerably delayed without inspiration from biology.

This process—taking a body of theory in a particular field, abstracting it into a general form, and re-applying it to other fields—is illustrated in Figure 2.1. This thesis does this not for Darwinian selection, but for another body of theory. I will be uncovering a general theory pertaining to abstract interactions between *innovation* and the *forces that align private returns with social returns*. It too can be applied to any system—economic, biological, or artificial—that has (i) genuinely innovative agents, (ii) returns to cooperation and defection, and (iii) forces that may align interests by modifying individual payoffs. All economic systems, from the household to international markets, have these properties, and so the theory is of broad application. Being related to the basic problem of aligning individual with social returns, it has profound implications for sustaining markets, political systems, or any other social order in a changing world.

## 1 A history of intellectual exchange

The social sciences and sociobiology have a long history of interaction. Adam Smith regarded our "propensity to truck, barter, and exchange" (1776/1976, p. 14), and the moral instincts that underpin social order (1759/2009), as innate to the human species—a consequence of our being "fitted by Nature" for the social condition (pp. 103-104). Charles Darwin's (1759) evolutionary biology would later articulate this "fitting" process. Darwin's account was inspired by Thomas Malthus' (1798) study of population growth in a finite environment, which led to scarcity, competition for resources, and survival of those with advantages in resource acquisition. Darwin drew other parallels, noting, for example, that divisions of labour and specialisation within organisms and environments increased productivity (Darwin, 1859, pp. 115–116). In his discussion of the evolution of human morality, Darwin (1871, p. 78) approvingly cites Adam Smith's *Theory of Moral Sentiments* and its model of human "sympathy".

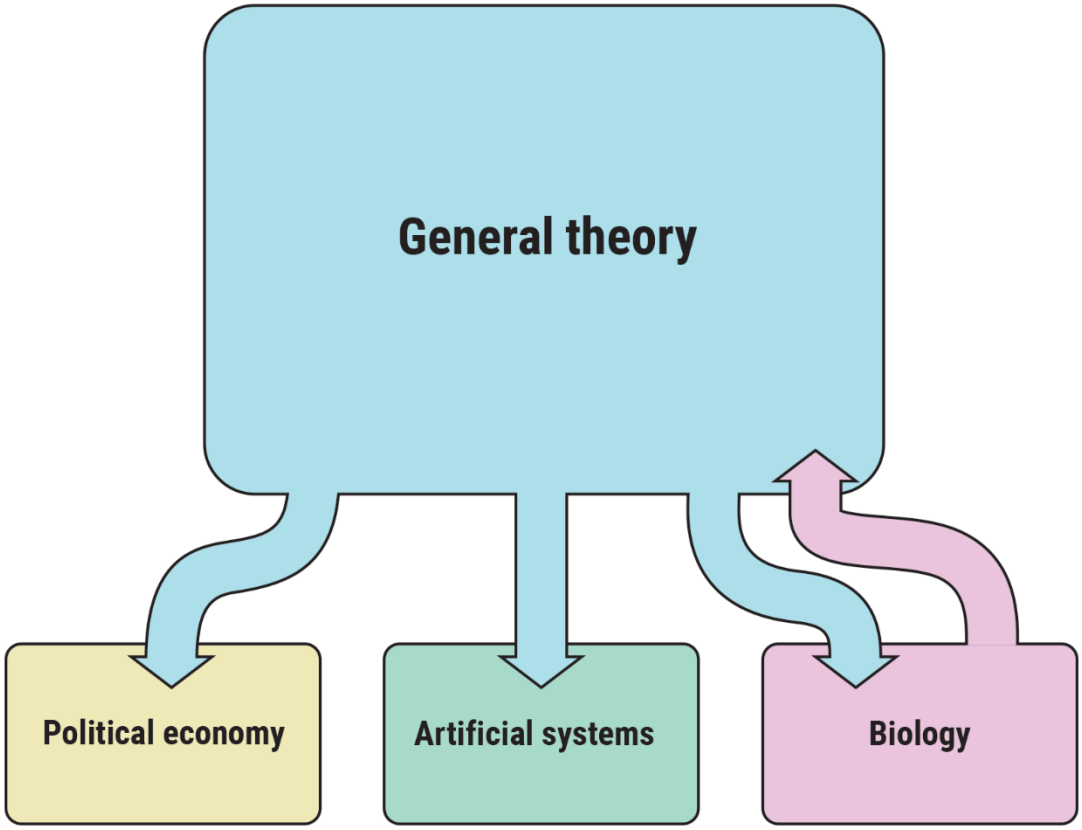



**Figure 2.1** The transfer of Darwinian selection, for example, by Veblen (1898), Alchian (1950), Friedman (1953), and Mokyr (1998). Biology is the source discipline, but selection theory can be applied to any field where the problem involves (i) variation in strategies pursued by actors, (ii) selection among that variety, and (iii) retention of the selected.

Earlier than Alchian and Friedman, Alfred Marshall (1920), one of the fathers of neoclassical theory, argued for a Darwinian economics in which "[t]he tendency to variation is a chief cause of progress" (p. 295). He believed economics could only make it so far with "[m]echanical analogies" such as equilibrium, which were at best "temporary auxiliaries" for simplifying a complex reality. Economics "is a branch of biology broadly interpreted", he wrote, and "[t]he Mecca of the economist lies in economic biology" (p. xiv). What he appears to have meant by this abstruse statement is that economics is ultimately concerned with non-equilibrating processes, with qualitative change driven by innovative agents—in his words, a science of "living force and movement".[48] Thorstein Veblen (1898) was perhaps the first economist to possess a truly sophisticated understanding of Darwinian

[48] Marshall adds that "economics, like biology, deals with a matter, of which the inner nature and constitution, as well as the outer form, are constantly changing".

selection, as the general theory we understand it to be today (e.g. see Hodgson, 1998). He anticipated today's dominant cultural evolution theories, explaining human progress through the selection of cultural and institutional forms. Some of the ideas of Marshall and Veblen found expression in Schumpeter's (1994/1942) "evolutionary" theory of innovation—of a

> *Process of industrial mutation—if I may use that biological term—that incessantly revolutionises the economic structure from within.*[49] (pp. 83)

For economics in particular, two zones of interaction have borne interesting fruit. They are also the two most important areas in this thesis. Naturally they relate to key aspects of the intellectual territory of economics:

- first, as a field dealing with *optimising and innovative agents*; and
- second, as a field concerned with *social interaction*.

This thesis suggests that these two threads can be woven together, and offer us a great treasure from sociobiological theory. If Marshall's Mecca exists, it is to be found in the *interaction between these factors*—between innovation and optimisation on the one hand, and the causes of cooperation on the other. But we will get to this in the body of the thesis.

## 2 Particular differences, general commonalities

Before that, the above discussion raises a question: Why is it that such apparently disparate fields can speak to one another? Paul Krugman, addressing a room of evolutionary economists, batted away the popular idea that economists are bent on imitating physicists. Rather, they are already much closer to evolutionary biologists. The economist who believes otherwise, Krugman suggested,

> *should do two things. First, read a text on evolutionary theory, like John Maynard Smith's* Evolutionary Genetics. *You will be startled at how much it looks like a textbook on microeconomics. Second, try to explain a simple economic concept, like supply and demand, to a*

[49] Schumpeter was cautious about drawing such analogies, which were unpopular at the time—but the commonalities are irresistible.

> *physicist. You will discover that our whole style of thinking, of building up aggregative stories from individual decisions, is not at all the way they think*. (Krugman 1996, section 2)

He concluded that, "to a remarkable degree", evolutionary biology looks like, "dare I say it?—neoclassical economics." To understand why, consider the basic similarities between the methods and concepts used in the two fields. Each involves:

- Optimising actors in a world of scarce resources and imperfect information, producing competition for survival, returns to efficiency, and returns to information and optimisation capacity.
- Returns to cooperation in the provision of public goods, and sometimes returns to exchange in "markets" with supply and demand dynamics. This supports extraordinary divisions of labour and specialisation, and the exploitation of comparative advantage.
- Defector strategies that produce private gains but erode social efficiency by undermining public goods provision, establishing monopoly and monopsony power, or exploiting behavioural biases.
- Strategies and structural arrangements that suppress defection, including punishments and rewards, relationships of dependence, and structures that reduce public good spillovers.
- Processes of strategic change or innovation that are radical and wholly unpredictable in the long-run. This includes innovations in the process of innovation itself, and innovations in "domesticated" forms of competition that use distributed knowledge to accelerate adaptation.
- Productivity growth arising from both proximate and ultimate causes. The proximate causes are extension (i.e. doing the same at larger scale), specialisation (i.e. focusing on a task and engaging in exchange), and innovation (i.e. doing something new). The ultimate cause is the selection environment (institutional, physical, etc.), which rewards some strategies of extension, specialisation, or innovation and not others.

Of course, there are also obvious differences between the fields, especially with respect to the *payoff being optimised* (utility or fitness) and the *process of optimisation* (blind or rational, or in all real-world anthropic and biological cases,

somewhere in between). At the level of abstraction in the general theory developed in this thesis, however, these are generally differences that make no difference; what matters is that there are payoffs and some means of optimisation, the two crucial factors that distinguish the animate from the inanimate world.

This is most obvious in the case of game theory, where the same game forms and dynamics are useful for explaining and predicting behaviour in the social sciences and sociobiology. Game theory requires interacting individuals with some kind of payoff and some means of optimisation, but, like Darwinian selection, has no necessary relationship with biology, fitness, and blind selection, or with economics, utility, and rational choice.

This thesis can be said to be concerned with a particular class of game: "innovative games". The fact of innovation, in the Schumpeterian sense, means that the parameters of any game scenario may change radically and unforeseeably over time. There is, in short, creative destruction of the game form. I suggest this is the only kind of game that exists in the real world. In the case of innovative games, we cannot adopt the standard game theoretic approach of solving for Nash equilibria. It will, however, be possible to identify equilibria of a more fragile sort—and this will be the key to understanding our fragile social orders.

### 3 Comments on the use of sociobiological cases

It is worth remarking, finally, on the presence of sociobiological cases in the thesis. In earlier drafts, these cases were woven more deeply into the argument. Given space constraints, and perhaps as a mercy to the social scientist reader, most have been shifted to the appendices. They are not essential for following the argument, but they are nonetheless useful because they are so *mechanical*—they illustrate cause and effect, stripped of the habits of thought, beliefs, ideologies, and assumptions that colour our understanding of the human world. They may, therefore, allow us to see the forces underlying our own social systems in a new light.

I have retained these appendices and associated citations for two reasons: first, because doing otherwise would be academically dishonest, in that they were key to developing the argument; and second, because they will be rich pickings for anyone who is interested in the underlying unity of these fields.

# PART I

# Introducing the two “keys to paradise”: Cooperation and innovation

# 3

# The fundamental social problem and its solution: The optimality line

## 1.1 Introducing the social strategy space

"Social" behaviours have consequences not only for the acting individual, but also for other individuals who are members of the same group. Consider the set of possible strategies available to an individual in a given period. These strategies exist as points within the "social strategy space" of Figure 3.1. The y-axis represents the expected private benefit of a given strategy for the acting *individual*—let us call this "$B_i$". On the x-axis we have the expected returns to that strategy for a *group,* of which the individual is a member. That group may be as small as a dyad, may be a large firm, may be a national community, or whatever other group we might define. Let us label this axis "$B_g$", for the expected benefit for the group. Every possible strategy is associated with a point in the social strategy space, and some combination of individual and group payoffs.

Most of what is discussed in the rest of this thesis concerns the creation of, and transformation of, strategies in the strategy space. The social strategy space describes the strategies available to a single individual. It differs from game-theoretic matrices in that it does not provide for the representation of the payoffs and strategies of other parties; rather, the effects of other parties' strategies are captured in the expected or actual values taken by $B_i$ and $B_g$. This limitation is by design and reflects similar choices by the cooperation theorists and biologists from whom I have borrowed this graph. It is straightforward to link the social strategy space to game-theoretic analyses.[50] However, as noted at the close of Section 1.2 of the previous chapter, I will not analyse games as payoff matrices or seek Nash equilibria, because doing so requires detailed specification of the game being played; it gives us precise knowledge, but about a toy world. The goal here is to make broad generalisations that apply when the game form is changing unpredictably, so cannot be specified.

[50] Each player's choices are represented in their own social strategy space, with expected or actual payoffs determined by the expected or actual strategy pursued by the other player.

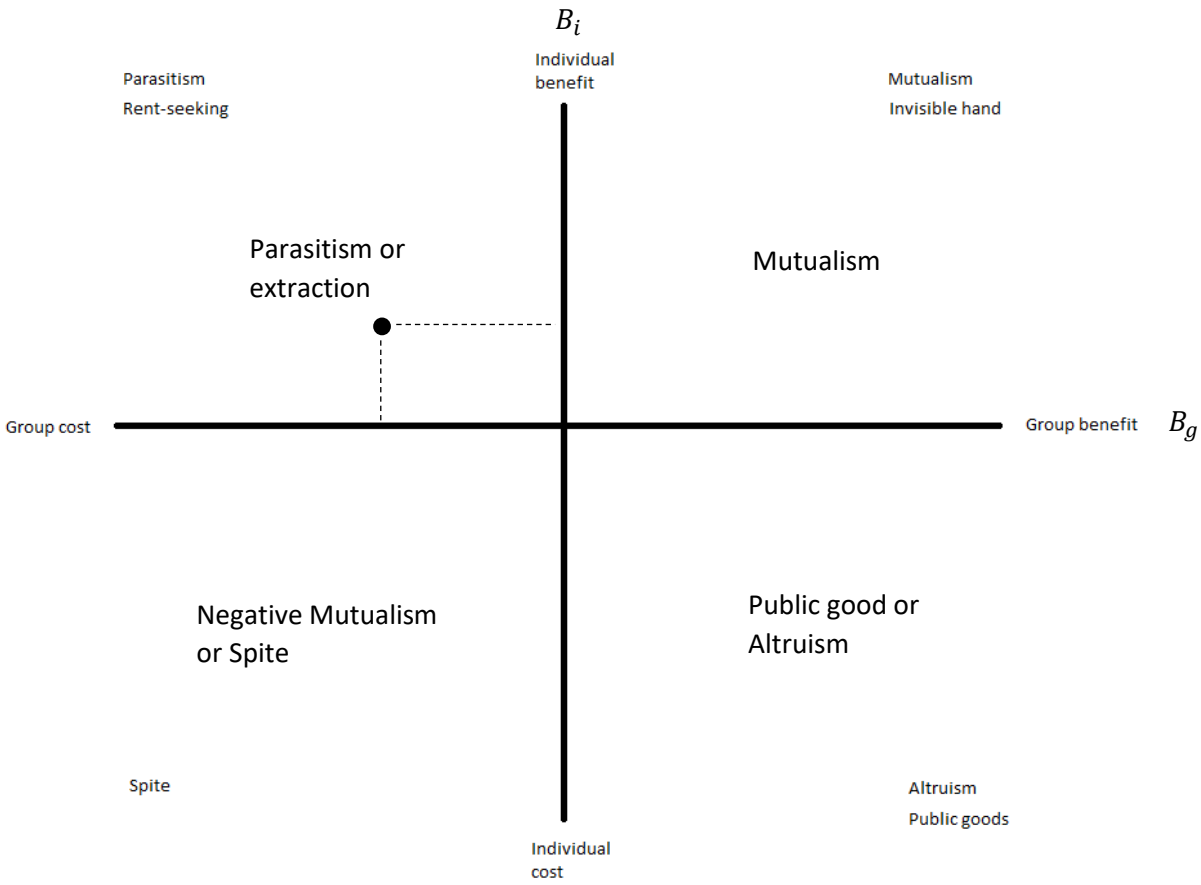



**Figure 3.1** The social strategy space, for analysing social behaviour from the primordial soup to modern civilisation. The y axis indicates individual benefit ($B_i$) and the x axis group benefit ($B_g$). Points in the space are possible actions undertaken by individuals, and the point given is an example of parasitism. The top right is the mutualistic or "M" quadrant, the bottom right is the public goods or "PG" quadrant, the bottom left is the negative mutualism/spite or "NM/S" quadrant, and the top left is the parasitism or "P" quadrant. Payoffs may be expected or actual, depending on the analysis.

The axes $B_i$ and $B_g$ define four quadrants: the "mutualistic" quadrant where payoffs to the actor and group are both positive; the "parasitic" quadrant where the actor benefits and the group is harmed; the "negative mutualism" or "spite" quadrant where payoffs to actor and group are both negative; and the "public good" or "altruism" quadrant when the actor pays a cost but the group benefits. The example strategy in Figure 3.1 benefits the individual but harms the group, and so is in the parasitism quadrant ("parasite" is derived from an ancient Greek term meaning "a person who eats at another's expense").

## 2.1 The "fundamental social problem"

Optimal group welfare requires the execution of strategies that fall within the two rightmost quadrants, those of mutualism and public goods. This is indicated by the green oval in Figure 3.2. Yet optimising self-interested individuals will select actions in both of the top two quadrants, those of mutualism and parasitism (the blue oval). This produces two areas of overlap between individual optimisation and group benefit:

- the mutualism quadrant, where the individual self-interestedly pursues behaviours that also benefit the group, and
- the negative mutualism quadrant, where the individual self-interestedly abstains from activities that also harm the group.

These two quadrants reflect the benign side of Smith's "invisible hand", where individual optimisation produces group optimisation. At the same time, individual optimisation produces two zones of conflict:

- the parasitism quadrant, where the optimising individual produces outcomes that harm the group; and
- the public goods quadrant, where the optimising individual neglects strategies that benefit the group.

These two quadrants are the "dark sides" of the invisible hand, where individual optimisation and group benefit are at odds. This only partial alignment is what DS Wilson (2002) calls the "fundamental social problem".

We must distinguish *expected* or subjective evaluations of payoffs, which include mental models that forecast others' behaviours and responses (e.g. Denzau & North, 1994), from *actual* or objective payoffs that would be available under conditions of complete knowledge and rationality. Self-interested individuals optimise according to expected private payoffs, while social welfare depends on actual group payoffs. Smith's "invisible hand" therefore requires that actual private and group payoffs are aligned *and* that individuals are able to accurately predict private payoffs. I will set aside the complications of limited knowledge and rationality until later chapters that include Schumpeterian innovation. Just as a simple, rational model most powerfully

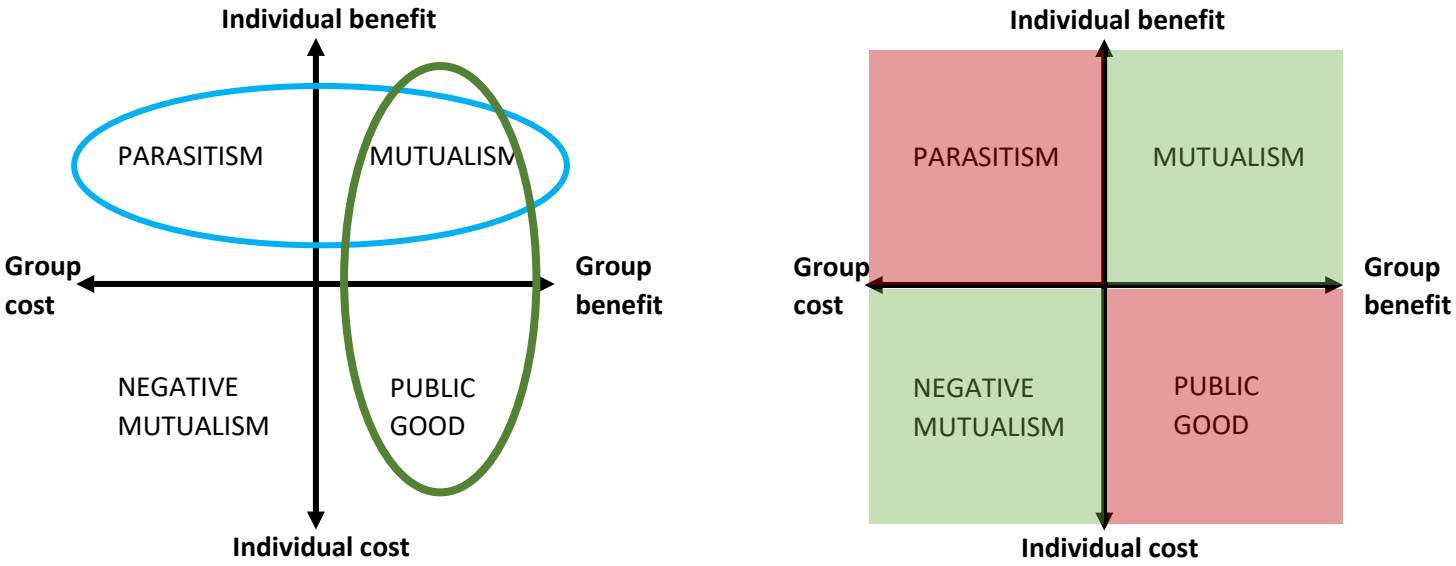



**Figure 3.2** In the strategy space on the left, circled in blue are the quadrants where individuals will select actions because they have positive individual returns. Circled in green are the two right quadrants, where actions have positive payoffs for the group. The result is the quadrant shown on the right: individual and group interests overlap in the mutualism cell and negative mutualism cell (the latter because individuals will not pursue actions that harm themselves), but they are in conflict in the parasitism and public good cells. This is DS Wilson's (2002) "fundamental social problem".

illustrates the basic logic of the invisible hand, so the perfect rationality approach will—for the time being—help convey the logic of the fundamental social problem and its solutions.

Strategy payoffs may be *transformed*, to borrow the mathematical term, by various factors that affect actors' incentives or motivations. For example, the parasitic strategy of theft can be suppressed by shifting private payoffs downward such that theft is unrewarding. I turn to the two main means of payoff transformation in the next chapter.

The fundamental social problem means that—in the absence of knowledge about both the distribution of strategies in the space and the effects of payoff transformation (for example, incentives imposed by institutions)—there is no *ex ante* reason to expect individual optimisation to have, on average, a negative or positive effect on the group. If strategies are randomly distributed in the space, then individual optimisation will produce an *average group return of zero*. These

problems have defined economics as a discipline since its beginnings in Smith. His concern was, as Rosenberg (1960m p. 560) summarises, to design market institutions such that they "cut off all avenues (and there are many) along which wealth may be pursued without contributing to the welfare of society".

The absence of any natural bias favouring cooperation means that growth is more or less non-existent in the Hobbesian pre-institutional condition of Warre, and was extremely slow under the sub-optimal institutions that existed for most of human history. Jones (1988/2000) writes that economic history

> *may be thought of as a struggle between a propensity for growth and one for rent-seeking, that is, for someone improving his or her position, or a group bettering its position, at the expense of the general welfare.* (p. 1)

This is the problem of taming parasitism. Periods of growth in recent and ancient history have followed the appearance of favourable payoff-transforming institutions, as well as other factors I will introduce shortly. As North and Thomas (1973) argue, growth requires institutional arrangements

> *that create an incentive to channel individual economic effort into activities that bring the private rate of return close to the social rate of return.* (p. 1)

Growth requires institutions that suppress the private payoffs of parasitic strategies (e.g. theft, rent-seeking, or more generally the externalisation of costs) and raise the private payoffs of public goods (e.g. teaching children or building roads).

### 3.1 The optimality line: Ideal absolute or rank-ordered alignment

Where resources are finite, group optimisation requires that individuals invest in available strategies in an optimal sequence—beginning with the rightmost actions in the strategy space, i.e. prioritising strategies with the highest group return ($B_g$), and moving progressively leftward, stopping where $B_g = 0$. Individuals, however, will

actually start at the topmost actions—those with the highest individual return ($B_i$)—and move progressively downward, stopping when $B_i = 0$.[51]

With this in mind, we can be even more specific about the ideally efficient alignment between individual and group payoffs: we want a set of payoff-transforming institutions and values that generate a *rank ordering of individual returns* to activities that is aligned with the *rank ordering of the group returns* to those activities.[52] This alignment is indicated by the 45-degree line in Figure 3.3, where the x and y axes are rank-ordered payoffs rather than payoff magnitudes. I call this the "optimality line", with a nod to Jones' (1988/2000, p. 187) "optimality band". Jones' optimality band is a different but related concept that describes a city or nation state that was neither so strong as to be oppressive and extractive, nor so weak that it could not restrain extraction by individuals—similar to Acemoglu and Robinson's (2019) "narrow corridor". The optimality line represents social welfare maximisation or perfect economic efficiency. The optimality band and narrow corridor may be understood as theories about what kinds of distributions of power between governments and civil society will tend to favour the optimality line.

If all available strategies lie on the optimality line, when self-interested individuals start pursuing the topmost strategies (highest $B_i$) and move progressively downward, they will inadvertently produce an identical ordering to starting with the rightmost strategies (highest $B_g$) and moving leftward. Individual optimisation maximises social welfare, as with Smith's "invisible hand". In both human societies and evolutionary sociobiology, cooperation is the result of *correlated returns*. I will give this correlation between $B_i$ and $B_g$ the label $r^*$. Where $r^* = 1$, we have the optimality line.

[51] Note this is true whether we assume self-interest or not. In the other-regarding case the actor still maximises $B_i$, where $B_i$ includes others' payoffs to some degree.
[52] I.e. among strategies available to the individual, the highest-return for the individual is also the highest-return for the group, the second-highest for the individual is the second-highest for the group, and so on. See A.3.1 for additional detail, though it is enough if the reader understands the effects of payoff-correlation.

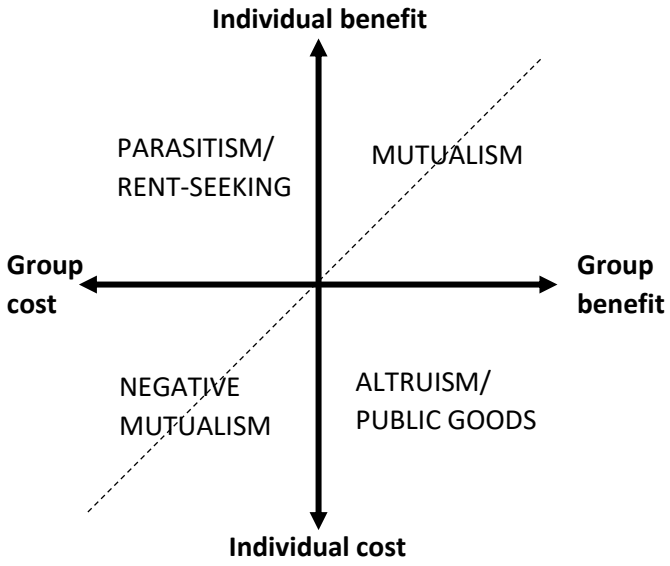



**Figure 3.3** The optimality line, where individual and group returns are perfectly aligned ($r^* = 1$). Note social and private returns do not need to be identical to produce perfectly aligned behaviour, only have an identical sign and rank-ordering—although in many biological and economic cases of ideal cooperation they are also identical in absolute terms. Perfection is of course not necessary for growth, but the better the alignment, ceteris paribus, the faster the growth.

## 3.2 Perfecting "direction" and maximising "pace"

We can thus distinguish two broad tasks for a cooperative system, whether biological or anthropic:

1. Transform payoffs such that individual returns are correlated with group returns on the optimality line (to the extent that the benefits of alignment outweigh the costs of transformation).
2. Subject to #1, maximise the rate of individual innovation and optimisation, i.e. the creation and exploitation of new positive-return opportunities in the strategy space.

These are the problems of *direction* and *pace* respectively.[53] As noted in the Part 1 introduction, they are the twin "keys to paradise" that allow social order to be maintained and prosperity grasped.

The payoffs to all enacted strategies can be expressed as a vector of individual and group payoffs. "Direction" is the direction of this vector, and "pace" is its magnitude.

[53] I have since found these two terms used in similar ways by others, e.g. Mazzucato and Skidelsky (2020).

If the vector points in a parasitic direction, increasing its magnitude (via individual innovation and optimisation) would increase harms to the group. If it points in a mutualistic direction but with meagre magnitude, there is stagnation. “Paradise” is a consequence of optimising both direction and pace, to achieve a vector with the highest feasible total payoff $B_g$.[54]

For the interested reader, I connect the above analysis to standard supply and demand curves, and the investment curve, in A.3.2.

[54] We can expect some trade-off between direction and pace, in the case that achieving $r^* = 1$ (i.e. ideal direction) requires a costly, information-intensive regulatory apparatus and so reduces the magnitude of the vector.

# 4

# Two causes of cooperation: Rules and commitment

The next task is to understand the ways in which the optimality line, i.e. $r^* = 1$ between $B_i$ and $B_g$, may be formed. It could, in theory, arise if the universe were happily structured so that cooperation was the only option—if individual and group returns were always arranged on the 45-degree line.[55] Again, history tells us that growth arrives only with favourable institutional and cultural settings (Jones, 1988), and it is these settings I investigate.

Section 1 begins by defining the functions $B_i$ and $B_g$. The same formalisms appear in the social sciences and sociobiology, although the terms have field-specific interpretations. These formalisms point towards a simple *taxonomy of the forces of payoff transformation*, i.e. the causes of alignment between $B_i$ and $B_g$. There are two—rules and commitment—and both appear in human and sociobiological systems.

Section 2 introduces "rules". These are strategies that actors employ to induce interaction partners to cooperate. They come in institutional (e.g. property rights) and individual-level forms (e.g. Tit-for-Tat). The key properties of rules include that:

- they involve *detection and enforcement* sub-strategies. The rule-wielder must detect that behaviour X has taken place (i.e. some form of cooperation or defection), and enforce the punishment or reward Y;
- they are *adversarial*, in that actors affected by rules have incentives to evade or exploit them, and will innovate to this end;
- *completeness is costly*, given each rule is suited to a narrow set of target behaviours, such that complex, large-scale cooperation requires a complex structure of rules; and
- they may be *public goods* (e.g. most or all institutions) or *private goods* (e.g. Tit-for-Tat in the repeated Prisoner's Dilemma).

[55] E.g. if all games were coordination games with full information.

Section 3 introduces "commitment". These are non-strategic factors that cause actors to place some value on others' payoffs. They include *structural payoff dependence*, where one actor's payoffs structurally depend on another's, and *motivational payoff dependence*, where one actor intrinsically values another's payoffs. They differ from rules in that:

- they are *non-strategic* because they arise from structural relations or actors' preferences;
- they are therefore *non-adversarial*, in that actors generally cannot, or will not, evade them;
- they are *complete* within their domain, producing what is described as "lockstep" transformation of the payoffs of many or all strategies; and
- they are not goods at all, but preferences and structural constraints.[56]

I define two kinds of actors: *mercenaries*, who are actors without any commitments and can only be induced to cooperate via rules (i.e. punishments and rewards); and *committed actors*, who are motivated by some degree of payoff or motivational dependence. Section 4 closes with a summary of the taxonomy.

### 1.1 Defining $B_i$ and $B_g$

Optimising agents cooperate when they stand to gain:

- *direct private benefits;* and/or
- *some indirect share in others' benefits*.

Much of this thesis is predicated on the usefulness of this general formulation in the social sciences. Box 4.1 discusses its ubiquitous use in evolutionary sociobiology. Individuals effectively weight their own payoffs with a value of "1", and others' payoffs with a value usually less than one and often zero. This construction is hardly controversial, and has a long history of use in the social sciences for describing the behaviour of both purely self-interested and other-regarding actors. In Kollock's (1998) terms, much of the modelling of human cooperation "has concentrated on various linear combinations of individuals' concern for the outcomes for themselves and their partners" (p. 192).

[56] Or "primitives" in modelling jargon.

Purely self-interested utility functions sometimes include weightings of others' payoffs, as where the actor depends on, or receives a share of, the returns to others—as in Becker's (1976) model of familial cooperation and Olson's (1993) model of dictatorial investment in society. It can equally be applied to explain the alignment of managerial and shareholders returns, where the former are paid in stock.

Weightings of others' payoffs are of obvious application in models of other-regarding preferences (e.g. Edgeworth, 1881, p.51, Liebrand, 1984, Fehr & Schmidt, 1999, Bolton & Ockenfels, 2000, and Dimick et al. 2016). As Edgeworth (1881) puts it,

> *between the two extremes Pure Egoistic and Pure Universalistic there may be an indefinite number of impure methods; wherein the happiness of others as compared by the agent … with his own, neither counts for nothing, nor yet "counts for one," but counts for a fraction.* (p. 16)

An informal statement of this utility function for the other-regarding agent has been present in the economic literature since Adam Smith (1759/1976), and is found in the work of his great friend David Hume (1739/1975), who maintained a theory of human motivation that included valuation of others' welfare (both influenced by Joseph Butler, 1726). Other-regarding preferences only enter the argument meaningfully from Chapter 15; before then, I will usually assume self-interest.

The individual payoff axis of the social strategy space is defined as follows:[57]

$$\boldsymbol{B_i = rb - c} \qquad \textbf{Eq. 4.1}$$

Every strategy available to the actor has some payoff $B_i$ defined in this way. Here $c$ is the direct private payoff (represented as a cost), $b$ is the payoff for other agents, and $r$ is a weighting that gives the individual some indirect share in others' payoffs. The value of $r$ will be constrained between 1 and -1, and of course it may often be zero.[58]

[57] Equation 4.1 can accommodate purely "strategic" cooperation by self-interested actors via mechanisms that modify $c$ (e.g. institutions) or raise $r$ (e.g. long-term dictatorial extraction), which are discussed later in the chapter. More complex models of social preferences, such as Fehr & Schmidt's (1999) and Bolton & Ockenfels' (2000) models of inequality aversion, can be accommodated by allowing the welfare of the focal actor and others (i.e. $c$ and $b$) to respond to the distribution of income.

[58] Cases of valuing others' payoffs more than one's own (i.e. $r > 1$) are rare. They may be common in families, but certainly not among strangers in large-scale societies. In biology, $r$ values above 1 are generally nonsensical.

Direct private payoffs are framed as a cost, so that negative values represent a benefit and positive values represent a cost. As shown in Box 4.1, this is the convention in sociobiology, and following it will make transferring some formalisms a little easier. Sociobiologists frame private payoffs as a cost rather than a benefit because this allows some economy of expression in the main cooperative cases of interest, i.e. the providing of public goods or abstention from parasitism, as I will now show.

For the individual actor to benefit from any strategy, the following must hold:

$$rb > c \qquad \textbf{Eq. 4.2}$$

If providing a public good or abstaining from parasitism is to be attractive, the weighted benefit to others ($rb$) must be larger than the private cost ($c$). This inequality also holds in cases where $r = 0$ so that the individual has no stake in others' payoffs. If $rb = 0$ so that others do not matter, an individual will only pursue a strategy if $c$ is negative, i.e. the behaviour is privately beneficial. Thus, *every* privately favourable strategy must satisfy $rb > c$.

The group payoff function is:

$$\boldsymbol{B_g = b - c} \qquad \textbf{Eq. 4.3}$$

Given we are dealing with actual group payoffs, the variable $r$ disappears—the focal actor is not special in any way, so that returns to the focal actor and to others are all weighted "1".

Interpretations of these variables and equations differ across human and biological cases. In the politico-economic context, the direct private payoff $c$ and others' payoff $b$ are measures of *welfare* or *utility*. The variable $r$ either reflects the degree to which the actor's payoffs depend on others', or is a preference about others' welfare, national glory, or whatever else may be relevant. I will clarify and label these different forms of $r$ shortly. In sociobiology, $b$ and $c$ instead refer to fitness, which is discussed further in Box 4.1.

**Box 4.1 Sociobiology: Fitness and Hamilton's rule**

The same equations are found in evolutionary sociobiology, although actors optimise "fitness" rather than utility. To be clear, the argument in this thesis in no way attempts to explain economic behaviour in terms of fitness. Understanding fitness, however, will help us to borrow some concepts from sociobiology. Wilson (1975) defines the fitness of a particular gene as:

> *the change in relative frequency in genotypes due to differences in the ability of their phenotypes to obtain representation in the next generation*. (p. 67)

A "phenotype" is the pattern of biochemistry, bodily organisation, and/or behaviour encoded by a "genotype", or gene variant. As Wilson observes, when organisms holding a particular gene are fitter than the average organism in the population, due to some aspect of bodily function or behaviour, they will reproduce more than the average. Fitter genetic variants proliferate and less fit variants disappear. This is the cause of evolutionary adaptation over time.

The central theorem in sociobiology for describing the conditions for cooperative behaviour is Hamilton's rule (Hamilton, 1964). It states that a gene's abundance in the population will increase given:[59]

$$rb > c$$

The formalism will be familiar—it is the same form as Equation 4.2. We again have private payoffs ($c$) and benefits to others ($b$), although now they refer to fitness rather than utility. And again, the variable $r$ indicates the stake that the actor has in others' payoffs—the share of the benefit for other agents that is relevant to themselves. Conventionally, $r$ refers to *genetic relatedness* with other affected actors, and the more related they are, the greater they weigh others' benefits.[60]

Its logic as an explanation for cooperation with kin is neatly captured by Haldane's famous witticism that he would gladly die for "two brothers or eight cousins" (Maynard Smith, 1976, p. 247), given average $r$ values of 0.5 and 0.125 respectively. Its appeal as a possible explanation for economic altruism within

[59] Equation 4.2, repeated in Box 4.1, is a common simplification used in biology. See A.4.1 for a more technical summary.
[60] Relatedness is in practice a little more complex: positive relatedness indicates that interaction partners share more genes in common than the average interaction partner, while the opposite is true for negative relatedness. See Birch (2017). For us, nothing important is lost in the simplification.

families,[61] where $r > 0$, is obvious (e.g. Becker, 1991, p. 16). However, the general form of Hamilton's rule is extremely flexible, and can be used to describe all the various conditions that shape cooperation (Birch, 2017 and Lehtonen, 2020).

It is so generalisable because it simply separates out "direct" versus "indirect" (Birch, 2017, p. 50) influences on cooperation, and that encompasses all the possibilities. This is also the role of Equation 4.2 in the politico-economic case. Actors cooperate because of *direct private benefits* and/or *some indirect share in others' benefits*.

With these equations in place, the fundamental social problem can be restated. It is that the selfish and non-dependent actor ($r = 0$) purely optimises over direct private returns.

$$\text{If } r = 0 \text{ then } B_i = -c$$

As discussed earlier, this gets us into the top two quadrants in the social strategy space, which will include the parasitism quadrant and exclude public goods. On average, we expect zero gains from cooperation.

### 1.2 Two modes of payoff-transformation

Now we are ready to return to the formation of the optimality line. Again, the goal is $B_i \propto B_g$. Given Equations 4.1 and 4.2, there are two obvious mathematical means of modifying the private payoff $B_i$ such that it is correlated with group payoffs $B_g$.

1. Modify direct private payoffs $c$, so that cooperation is directly privately beneficial.
2. Modify the stake an individual has in others' payoffs, i.e. via the weighting $r$ applied to $b$.

The first is the topic of Section 2, and the second the topic of Section 3.

[61] I.e. a gene that promotes generosity towards likely carriers of the same gene, i.e. genetic kin, may increase its prevalence in the next generation.

### 2.1 "Rules": Building the optimality line by modifying $c$

In this section, assume that $r = 0$. Individuals are self-interested and do not have any other kind of meaningful stake in others' payoffs. Thus we have

$$B_i = -c$$

$$B_g = b - c$$

There is an obvious way to modify $B_i$ such that it is ideally correlated with $B_g$. For any given strategy, simply apply a reward or penalty to $B_i$ that is equal in magnitude to $b$ (the strategy's payoff for others). This reward/penalty alters direct private returns to pursuing a given strategy. Let us call this reward or penalty $b^*$. This gives us

$$B_i = b^* - c$$

Clearly if $b^* = b$, then $B_i = B_g$ and we get the optimality line. This is the logic of Pigouvian subsidy and taxation, and we will turn to that and other mechanisms, including in sociobiology, shortly. Because this approach to generating cooperation works by altering the individual actor's direct private payoffs, these are "$c$-based mechanisms". Rather than using this clumsy term, however, I will generally refer to these mechanisms as *rules* or *rewards and punishments*.

As these equations suggest, modifying $c$ via rules may therefore substitute for $r > 0$ (soon I will term $r$ a measure of "commitment"). They are two different ways of bringing the value $b$ into $B_i$. Later, innovation will upend this result spectacularly; they become only partial substitutes and essential complements.

### 2.2 Rules as strategies: Identification, enforcement, and counterstrategies

A crucial property of rewards and punishments is that they are *strategies* that must be executed by some actor, and they generally take the form of "if X, then do Y". They involve two sub-strategies: first, identifying that X has taken place; and second, enforcing Y.

Here we can see more convergence between political economy and sociobiology: the institutional economist Douglass North (1988, pp. 17-19; 1990, pp. 27-32) refers to these two capabilities as "measurement" and "enforcement",[62] while sociobiologists refer to the same but the first term is varyingly replaced with "discrimination", "detection", or "identification". For our purposes, North's "measurement" is a problematic term. Measurement implies perceiving dimensional attributes, and is clumsy in contexts where qualitative or categorical judgement is involved—for example, to speak of a state regulator "measuring" whether property rights have been violated. "Discrimination" is perhaps the most accurate and flexible term, and it is closely associated with perceptual processes in psychology and sociobiology. It is, however, unfamiliar for many social scientists and carries heavy baggage from its association with chauvinism (i.e. discrimination against this or that sub-group).

I will instead use "detection", which usefully implies that there will be adversarial efforts to evade detection, and is a natural pairing with enforcement. Thus, "detection and enforcement" are the sub-components of rules. In sections where detection and enforcement are referred to frequently, I will sometimes use the abbreviation "D&E".

As just implied, rules are *adversarial*, in the sense that the targets of regulation almost always have incentives to innovate counterstrategies for evading, exploiting, and capturing the rules—if they can—to escape due punishments or receive undue rewards.[63] This is achieved by concealing or misrepresenting X, and evading or defending from Y. We will turn to these problems in later chapters, once I have properly introduced innovation.

### 2.3 Institutional reward and punishment

Rules may be enforced by lone individuals, or they may be enforced by individuals within institutional hierarchies. Let us first turn to institutions, which only exist in human societies and are a core part of politico-economic theory. The addition of $b^*$ to $B_i$ is the basic logic of Pigouvian taxation and subsidy. Pigou (1920) saw that if an economic activity is associated with an external social cost or benefit—equal to $b$ in

[62] North places these alongside an "ideological" force, which is set aside until Chapter 15.
[63] They may prefer that systems of rules exist and apply appropriately to others, but not to themselves. This is a classic free-rider problem.

our analysis—then a publicly imposed tax or subsidy that ideally sets $b^*$ will internalise the cost for private transactors. If ideal, the taxation and subsidy of various activities will produce the tight correlation of the optimality line as shown by the black dots in Figure 4.1. Parasitisms are penalised or taxed, so that their private payoffs are adjusted downwards and the strategy is pushed into the negative mutualism quadrant. Public goods are rewarded or subsidised, so that their private payoffs rise and lift them into the mutualism quadrant. With that, the fundamental social problem is solved.

Example strategies labelled A to D, and their transformation via institutions, are shown in Figure 4.1. To give some examples, "A" might be criminal justice in the case of theft; "B" might be a tax applied to an externality-generating activity such as carbon emissions, or the removal of a perverse subsidy; "C" might be payments to teachers so that they provide public education, or to soldiers who would otherwise not accept the costs of specialising in warfare; and "D" might be reducing a subsidy for a public good such as road-building because it may otherwise be overprovided. Detection and enforcement are naturally critical in all these cases. For example, a state must successfully identify that the individual strategy of labouring to build roads or educate children is group-beneficial but privately costly and underprovided, identify the subsidy $b^*$ that will encourage efficient provision, and be able to enforce the appropriate delivery of that subsidy. The same goes for a public harm, such as theft, where $b^*$ is a tax or punishment.

Note that the approaches of Coase (1960) and Ostrom (2000a) still employ c-modifying rules to deliver the optimality line, though they differ from Pigou in the ways that the penalty or subsidy $b^*$ is negotiated and enforced (see A.4.2). Ostrom's scheme, however, also explicitly depends on $r > 0$—forms of commitment, such as cultural affiliation and trust-building communication (e.g. Poteete et al., 2010)—and this places Ostrom beyond the pure use of $c$-modifying rules.

If ideal, punishment and reward regimes induce purely self-interested actors to behave "as if" they intended to maximise social welfare. This establishes the conditions of Smith's (1776) butcher, brewer, and baker, who, under a regime of property rights and competition, serve others' needs out of self-interest.

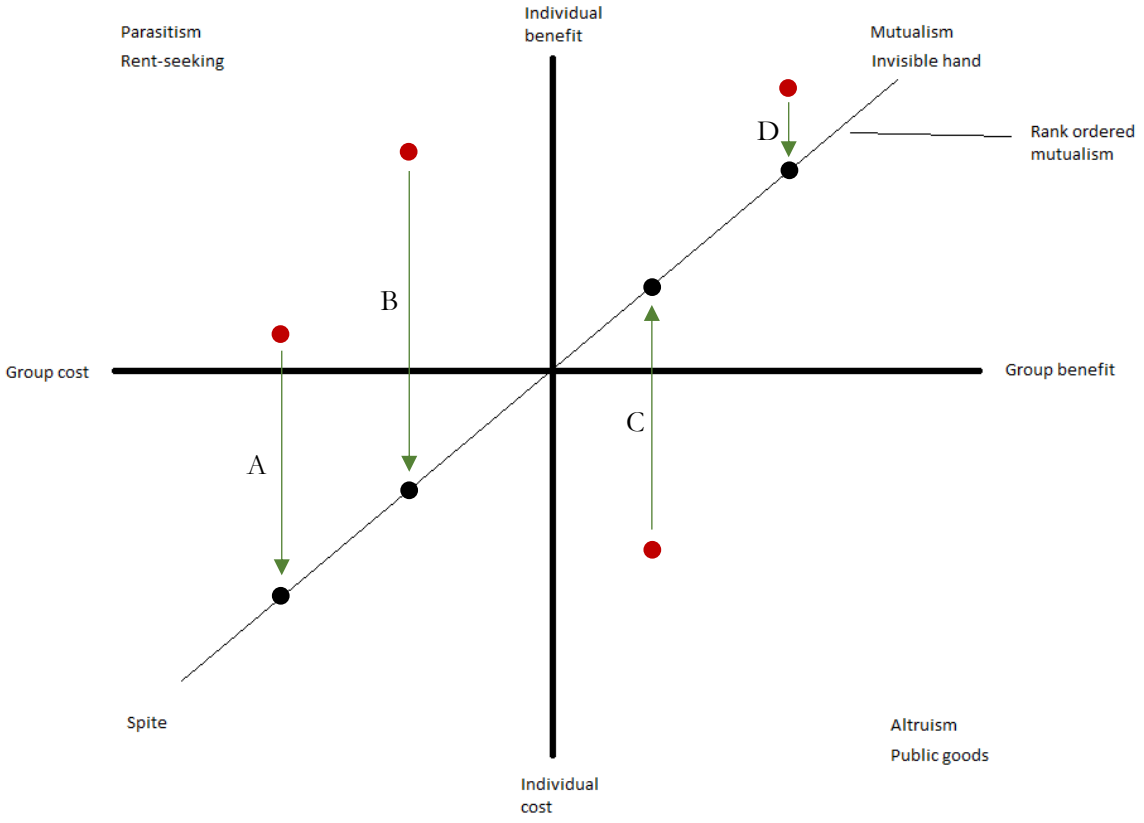



**Figure 4.1** The optimality line achieved via rules. Ideal rules modify individuals' private returns, shifting them towards the optimality line (green arrows). The "raw" strategies, prior to rules, are indicated by the red dots, and the post-intervention "transformed" strategies are indicated by the black dots.

## 2.4 Individual-level reward and punishment

The second class of rules is non-institutional. It involves *lone individuals* altering the private payoffs of their interaction partners. They still require strategies of detection and enforcement: reciprocity strategies, for example, involve detecting interaction partners' behaviour in past rounds of play, or signals that indicate intentions, and depend on capacities to execute punishments and rewards.

Like institutions, individual-level rules can increase cooperation. They can, for example, transform the iterated Prisoner's Dilemma into a pure mutualism, as shown in Axelrod's tournaments held between AI players programmed by various game theorists (Alexrod & Hamilton, 1981). The most successful AI strategy was Rapoport's simple Tit-for-Tat program: cooperate if the partner cooperated and

defect if they defected. Tit-for-Tat produces the desired tight correlation between actors' payoffs; as Kollock (1998) writes, the Tit-for-Tat strategy

> *means that the only long-term possibilities are mutual cooperation and mutual defection—there is no hope of exploiting this strategy in any kind of sustained way.*[64] (p. 196)

Following Chapter 3, perfect payoff correlation ($r^* = 1$) entails perfect cooperation, with payoffs on the optimality line. This is a crude form of the optimality line, given the choice is binary, but Pigouvian patterns appear in experiments where penalty size is allowed to vary continuously: penalties increase with the size of the harm inflicted (e.g. Fehr & Fischbacher, 2003). I discuss various other varieties of individual-level rules in A.4.3, which again appear in both human societies and biological systems.

Finally, while we may tend to think of markets as primarily institutional, individual-level mechanisms of detection and enforcement are essential complements to institutions. We will explore this further in later chapters. Particularly fascinating research on the appearance of markets in biological systems, which arise in circumstances that provide for voluntary choice, is presented in A.4.4.

### 3.1 "Commitment": Building the optimality line via raising *r*

Now I turn to the second cause of cooperation, which occurs where $r > 0$ so that the actor values others' payoffs ($b$). I give this force the label "commitment", for reasons that will become clear shortly. Commitment arises for two reasons:

1. *Structural payoff-dependence:* an individual's payoffs structurally depend on others' payoffs, so helping others is a self-interested strategy; or
2. *Motivational payoff-dependence:* an individual places intrinsic value on others' payoffs, so will assist others without any redounding benefit.

These subtypes are discussed in detail in Sections 3.3 and 3.4. First let us turn to their differences from rules.

[64] In practice, Tit-for-Tat strategies are problematic. Information problems, disagreements about interpretations of events, and interactions with social norms can lead to spiteful spirals of punishment and counter-punishment (e.g. Andersson & Pearson, 1999).

### 3.2 Self-enforcing, non-adversarial, and "lockstep" payoff transformation

The properties of structural and motivational dependence are, roughly speaking, opposite to those of rules discussed in Section 2.

First, they are *non-strategic*, in that they (usually) do not require the execution of detection and enforcement strategies.[65] They are self-enforcing. That is, $r > 0$ is usually a consequence of the arrangement of facts outside of the strategic game—properties of the world, or of preferences that I will, initially at least, treat as given.

Second, to the extent that payoff transformation is self-enforcing, it is also *non-adversarial*. The person whose payoffs are shifted to favour cooperation does not develop counter-innovations to resist that payoff alignment.

Third, they tend towards *completeness*, at least within their domain. Rules, as discussed, are composed of concrete detection and enforcement strategies that are tailored to solve specific cooperation problems—to modify $c$ where a specific behaviour X is detected. In contrast, the presence of just one commitment mechanism can transform the payoffs of most or even *all* strategies. We can see this formally in the "lockstep" nature of payoff transformation by $r > 0$. This is illustrated in Figure 4.2, which shows the payoffs for eight strategies and how they are transformed when $r$ takes values of (A) zero, (B) 0.5, and (C) 1.

The left-hand side and right-hand side of the graph show the same phenomenon in two different ways. In both cases, the area shaded green indicates the strategies that are favourable for the actor.

- The left-hand side of Figure 4.2 shows the "raw" individual payoffs for these eight strategies, prior to transformation by $r$. That is, it shows only *direct* private payoffs. As $r$ increases, the actor appears to become a $B_g$-maximising altruist: the green shaded region changes shape, rotating around the origin, until at $r = 1$ actor pursues all strategies in the mutualism and altruism quadrants.

[65] There are some significant exceptions, mainly in the case of structural dependence, to be discussed later.

- The actor is, of course, just optimising their payoff $B_i$—and as $r$ increases, there are increasing returns to benefiting others. They may win a share in the returns to others, or may place intrinsic value on others' returns. The right-hand side of Figure 4.2 shows these transformed payoffs, and so *actual* individual payoffs, both direct and indirect. Accordingly the green shaded area always covers the part of the graph with positive individual payoffs ($B_i > 0$).

Thus, the $r = 1$ actor is a group-maximiser (left-hand side) but is of course maximising their own payoff function (right-hand side). Our preferred representation will be that on the right-hand side, showing the agent's actual payoffs.

To summarise the "lockstep" effect, increases in the value of $r$ shift all strategies, and they move towards the optimality line in proportion to the value of $r$. If $r = 0.5$, private payoffs close half the distance to the optimality line. If $r = 1$, they close the full distance. It can be described as a sort of automatic and universal Pigouvian mechanism, in that the shift in individual payoffs is proportionate to group benefits—but, of course, it occurs without the use of rules.

I have mentioned that this lockstep effect may apply to all, or only some, strategies. That is, the *domain* of an $r$-based mechanism may be complete or partial. A complete domain is naturally preferable for more perfect cooperation, though a partial mechanism may also do good work. Examples will follow shortly.

Fourth and finally, while rules can be private or public goods, commitments are neither. They are primitives—preferences or structural facts that actors generally cannot change.

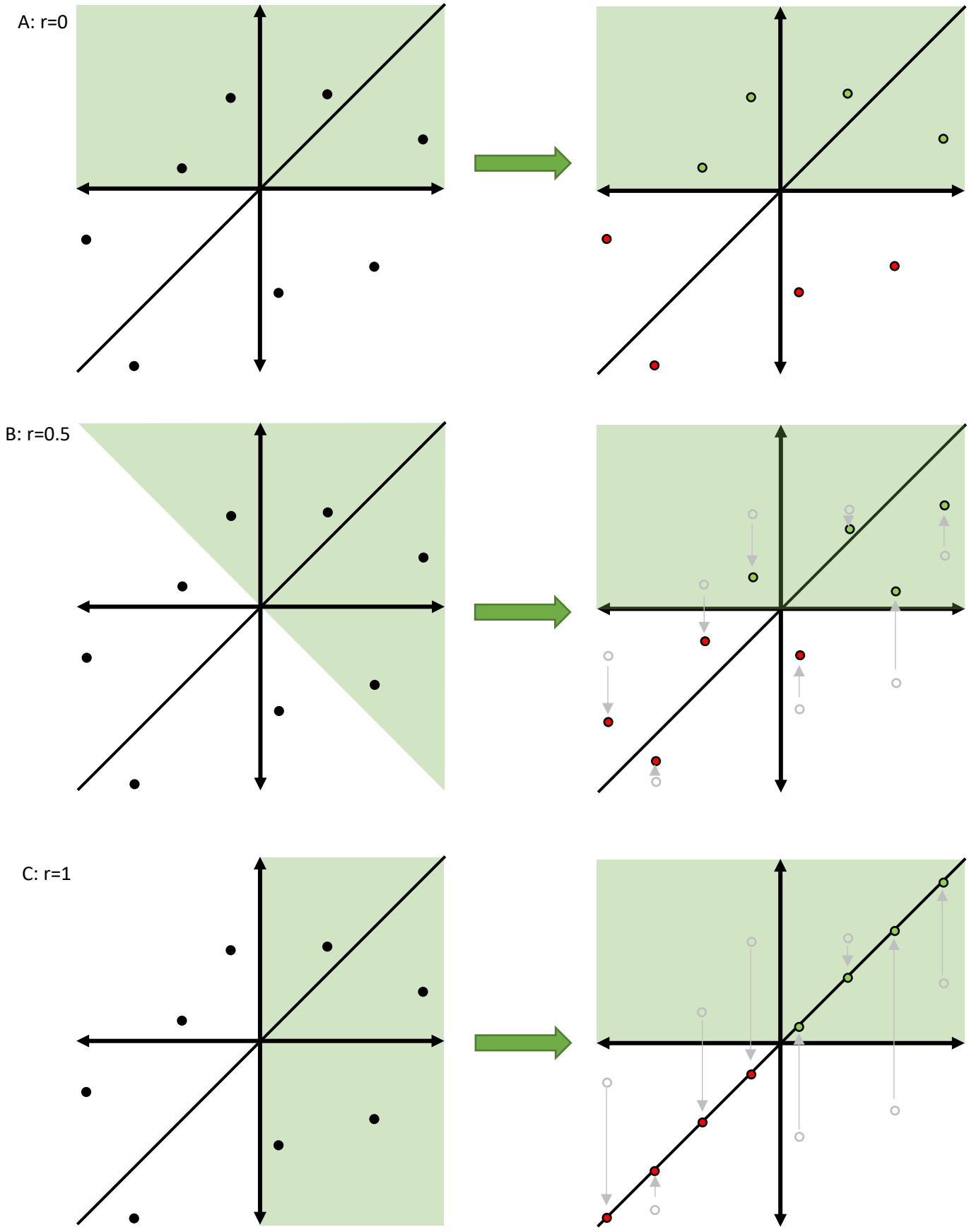



**Figure 4.2** The same eight strategies under different values of $r$. The green area indicates strategies favoured by evolution, or an economic actor, given the value of $r$ in the payoff function. Graphs on the left show payoffs pre-transformation, and show that as $r$ increases actors will increasingly favour altruistic strategies and disfavour parasitism. Graphs on the right side are post-transformation, and show that these altruistic actions are favoured, and parasitic disfavoured, because the actor's "real" individual payoffs have been shifted to the optimality line.

From here, these dynamics, and the specific effects of the different sub-types of commitment, are illustrated through some cases. I define one sub-type of $r$ for the self-interested case:

- $r_s$, indicating structural payoff dependence. It works the same way in anthropic and sociobiological cases: individuals come to depend on others, or on groups, such that it is worthwhile investing in those others or groups.

There are two sub-types of $r$ where others' payoffs are valued in themselves, rather than because they redound to the focal actor:

- $r_c$, or commitment arising from other-regarding preferences (Smithian sympathy, "love thy neighbour") or preferences over group phenomena, e.g. national glory, equality, or the spread of a religion.
- $r_g$, or genetic relatedness. It is a sociobiological analogue to other-regarding preferences, as the actor *intrinsically* values' others payoffs.[66]

Our politico-economic argument will only require three modes of payoff transformation: rules, $r_s$, and $r_c$. We will not have much use for $r_g$, though it will occasionally turn up as we perform our "raid" on sociobiological theory.

### 3.3 Cooperation via other-regarding preferences: $r_c$

Let us start with other-regarding preferences. Obviously if the actor fully values $b$, then $B_i = rb - c = B_g$ and we get the optimality line. I have termed this form of intrinsic commitment $r_c$, and it represents a psychological preference. Later we will examine a large body of evidence that suggests this preference is intimately related to perceived cultural distance, or cultural relatedness (which is neatly consonant with the term $r_c$).

Other-regarding preferences share most properties in common with ordinary "selfish" preferences. The nature of all preferences is one of intrinsic commitment; we do not hold some preference over chocolate ice cream for strategic reasons, or because rules tell us to do so. Indeed without preferences, without some intrinsic pursuit of utility, it would be impossible to reward or punish an actor and rules

[66] See A.4.5 for brief discussion of the question of whether "altruism" has any sensible meaning in a world of apparently selfish genes (it does).

would be meaningless. Preferences, whether selfish or other-regarding, are non-adversarial in the sense that the actor cannot strategically "evade" those preferences, nor coherently "prefer different preferences".[67]

The domain of other-regarding preferences—the set of strategies affected by lockstep payoff transformation—may be complex. In the simplest case, the other-regarding actor places some value on others' welfare. This produces a commitment that is domain-complete ($r_c$ applies equally to all values of $b$). Interest in others' welfare may take the richer pattern of Adam Smith's "sympathy", such that the domain of payoff valuation is constrained by cultural norms and human psychology (discussed in Chapter 15). Or people may value national glory, or hold moral preferences over particular kinds of people, behaviours, and sources of welfare. The domain of the other-regarding preferences of the Smithian "fanatic" may be limited to the welfare of a narrow cultural in-group; indeed, they may delight in the destruction of immoral non-believers (negative $r_c$ in the domain of outgroup welfare).

Other-regarding preferences will only become significant from Chapter 15 onward. There we will examine empirical evidence on their strength and structure, and begin to see why Smith regarded them as the necessary ground of liberal social orders (a position held by many others including Hirsch, 1976, and to some degree Arrow, 1972, and Schumpeter, 1942).

### 3.4 Structural payoff dependence: $r_s$

The other form of commitment I have labelled $r_s$. It applies where self-interested actors can reap indirect benefits by raising the rewards reaped by others or a broader group. If $r_s = 1$, then $B_i = B_g$ and we get the optimality line. This purely strategic form of payoff dependence is profoundly important in biological cooperation; indeed it is necessary to explain every organism that has ever lived, given organisms are built from different genes that do not have intrinsically aligned interests (Box 4.2). After introducing innovation, we will later see why $r_s$, rather than rules, is the necessary ground of complex social orders (e.g. organisms) throughout the kingdom of life.

[67] Addiction presents some challenges. In the standard framework, the choice to partake reflects their preference at the time. Explanations include optimisation problems and time-inconsistent preferences.

In politico-economy, $r_s > 0$ operates in the same kind of conditions outlined in Box 4.2 on sociobiology. It is a spur to cooperation wherever individuals have some share in group returns, moving between groups is difficult, and where within-group competition is muted and between-group competition is intense. A classic example is an employee with firm-specific skills, whose access to wage premia depends on firm survival, and who—if within-firm competition is restrained—can be confident in winning a share of improvements in firm performance. Such an arrangement acts as a substitute for "rules", i.e. the individual will work for group interests without surveillance and associated rewards/punishments (or detection and enforcement). We can already glimpse the advantages of the non-adversarial nature of such transformation: the firm does not need to discover and thwart novel strategies for employee shirking, as employee cooperation is self-enforcing. Another example is the attempt to align managerial incentives with shareholders via payment in stock. These kinds of incentive regimes differ from intrinsic motivations ($r_c$) in that they are not wholly innovation-proof, and they are seldom if ever domain-complete.[68] These shortcomings will be of more interest later.

As we are concerned with large-scale social orders,[69] the most important case of $r_s > 0$ will be the dictator. Consider Mancur Olson's (1993, p. 567) account of the motives of "roving bandits" versus the motives of the dictator as a "stationary bandit".[70] Roving bandits have incentives for total extraction from the communities they pillage, because they move from one to the next and gain no benefits from leaving any community intact. The world of roving bandits undermines incentives for communities to accumulate wealth, which will only attract the bandits' attention, and so communities remain poor. The stationary bandit, however, extracts from a single society for an extended period. If taxes are set at 100 percent, this will destroy economic activity and opportunities for further taxation. Taxation of zero percent also provides nothing. The rational stationary bandit sets an intermediate tax rate

[68] Because this extrinsic commitment is brought about by particular strategic conditions, it is vulnerable to very specific forms of change in individual strategies (e.g. innovations in moving between groups) or the strategic environment (e.g. increased within-group competition). Worker incentives, for example, are domain-limited because they may have incentives to engage in within-firm competition too, e.g. preventing the rise of more talented others. To the extent that these weaknesses can be closely guarded, as they are in organisms, the lockstep alignment effect over a vast range of other strategies remains very useful.
[69] I.e. especially democratic, but also authoritarian, capitalist societies.
[70] Note the logic also applies to the local warlord or anyone who extracts from a community on the long-term.

that maximises long-run extraction. Moreover, they will even provide public goods where they increase economic activity, and so tax income, so long as the cost to the stationary bandit ($c$) is less than the public benefit ($b$) times the tax rate ($t$). Thus Olson offers the following equation:

$$tb > c$$

This will be familiar: $t$ has taken the place of $r$ in Equation 4.2 (and in its equivalent in sociobiology, Hamilton's rule). Specifically, $t$ is a form of $r_s$, with the dictator structurally dependent on market activity for tax income, and having no incentive to innovate a way out of this agreeable situation.

Are authoritarian motives well-described by Olsonian payoff dependence? In practice, dictators too are products of acculturation and are likely to have some degree of commitment to a cultural in-group (whether narrow or broad) or national glory that goes beyond pure considerations of material income. This as a case of $r_c > 0$, operating alongside $r_s > 0$. Olson's caricature of the purely self-interested dictator is, however, useful for illustrating the effects of structural payoff-dependence, illuminating insofar as authoritarians depend on tax income to achieve their goals, and later will help us reveal what varieties of order can survive in a world of pure material self-interest (i.e. where $r_c = 0$).

### 3.4.1 Structural payoff-dependence versus rule-based reciprocity

Can the authoritarian's investment in society be explained as a rule-like system of reciprocity, where citizens detect whether the ruler invests, and strategically reward or punish in response? The $r_s$ structural dependence mechanism operates in the total absence of such strategies. This is significant and obvious in biology too.[71] If the dictator fails to provide public goods such as roads or property rights, structural payoff-dependence means that he/she will suffer *automatically* from lower economic output. Payoff-dependence is *self-enforcing* across a wide range of strategies, given the lockstep effect. Of course, citizens may additionally punish the dictator through

[71] For example, we can distinguish between an organism that strategically commits suicide to kill a parasite or a cheater (an extreme punishment strategy that some ant queens employ, see Rissing et al. 1996) versus a host dying because of high parasite or cheater virulence. Both regulate parasite fitness, but only the former is a *strategic* response. Incidental host death instead regulates the parasite purely to the degree of its structural dependence on the host.

**Box 4.2 Structural dependence in sociobiology**

In evolutionary sociobiology, $r_s \approx 1$ is behind the basic structure of every single form of microscopic or macroscopic life, from the simplest virus to the tallest tree. It is closely analogous to the modes of structural payoff dependence in human societies discussed in the body of the text.

In sociobiology, two varieties are ubiquitous:

- the organisation of genes into the *genomes* that make up all life. From the gene-centred perspective that dominates biology, genomic cooperation is a form of group selection. As Haig (2020, p. 21) puts it, genes are themselves "members of social groups", and we call those groups genomes; and
- the organisation of organisms into *cross-species symbioses*, where a higher-order individual is composed of a group of two or more species. Such symbioses are almost as foundational to life on Earth as the genome, and can be just as cooperative—mitochondrial, chloroplastic, and bacterial symbioses, among many others, are essential for all life visible to the naked eye.

All biological "individuals" in existence are, from a genetic perspective, actually groups composed of complex, multi-level alliances. The members of these groups share no intrinsic genetic interest (i.e. $r_g = 0$), but form "individuals" only because structural payoff dependence is so well-developed that the interests of their constituents usually cannot be divided. The only viable pathway for reproduction of a group member is reproduction of the whole group. Thus, $r_s = 1$ and the players involved behave "as if" they were a unitary interest.

This payoff dependence, described extensively in Birch (2017), requires:

- *partner fidelity,* a tendency for group members to be stuck with one another over the life course and to reproduce together over many successive generations (e.g. Vautrin & Vavre, 2008); and
- competition *between* these well-bounded groups rather than *within* them (Maynard Smith, 1988). Forms of within-group competition that do not serve the group allow private returns to become decoupled from group returns, leading to collapse of structural payoff dependence.

These echo the conditions for $r_s > 0$ for the dictator, for the employee with firm-specific skills, or any other anthropic case: actors must be stuck together without an alternative, and not caught up in destructive internal contests, for $r_s$ to be high.

Szathmáry and Maynard Smith's (1995) "major evolutionary transitions"—great leaps in the complexity of "individuals"—arose the structural creation of conditions that deepened payoff dependence.[72] Again, complex social orders arose from the deepening of commitments, not advances in rules. The explanation, we will see later, is intimately linked with innovation.

detection and enforcement strategies, such as protesting the absence of public goods, and this may raise the total correlation between ruler and societal payoffs ($r^*$) above the level achieved by $r_s$ alone.

### 3.4.2 Structural versus motivational payoff-dependence: The Leviathan and long-term parasite

Say the dictator imposes a tax rate of 50 percent, and optimises over $tb > c$. At this tax rate, is the dictator's behaviour analogous to that of an altruist who applies a 0.5 weighting to others' welfare? Superficially, we might conclude the answer is yes: the dictator will pay a private cost wherever doing so produces at least twice the increase in societal output.

However, cooperation based in $r_s$ has additional potential for conflict (see footnote 65, and Box 4.3 for some sociobiological cases), and this case is no exception. To begin with, the dictator's payoff alignment is domain-limited. If purely self-interested, the dictator will have an interest in strategies that raise *total economic output*, but no commitment to *social welfare*. The dictator has no interest—beyond secondary effects on expected income—in civil and human rights, citizens' actual standard of living, or society's economic performance following his or her death or deposition. Further, although the dictator's structural payoff dependence is non-adversarial, it may nonetheless be weakened by exogenous shocks. As Olson (1993) notes, dictators become more parasitic when their expected length of tenure shortens, which brings a decline in the expected returns to long-term investments in public goods. The insecure or short-term dictator may accept extensive economic

[72] Maynard Smith (1988) ties such suppression of internal conflict to jumps in the complexity of life: *"One can recognize in the evolution of life several revolutions... In each of these revolutions, there has been a conflict between selection at several levels. The achievement of individuality at the higher level has required that the disruptive effects of selection at the lower level be suppressed."* (pp. 229-230)

destruction if it helps secure their capacity to rule and tax into the future. Again, one cause of this is within-group competition: it motivates wasteful expenditure of resources on that conflict, and creates incentives to pursue shorter-term extractive strategies in a sort of tragedy of the commons.

We can connect this to Hobbes' preference for a secure Leviathan—i.e. absence of internal competition—to which we will return in later chapters. These incentives are analogous to those of the long-term parasite in sociobiology, discussed in Box 4.3; the best Hobbes (1651) thinks we can do, one might say, is install and protect a single long-term parasite.

**Box 4.3 Conflicted structural dependence in sociobiology**

Olson's stationary and roving bandits appear in sociobiology too. Take parasites, which depend to varying degrees on the host's survival, so that $r_s$ may be positive. Fascinatingly, this can lead parasites to evolve into partial or complete mutualists over time (e.g. Jeon, 1972; Roughgarden 1975; Boucher 1985; de Mazancourt et al. 2005; Fellous & Salvaudon, 2009). Such a transition has been observed occurring in as little as twenty years (Weeks et al., 2007). This effect is meaningful where parasites' fates are partially tied to their hosts: they must be *long-term* parasites, living in the host for a substantial part of the host's life-cycle, and ideally transmitted to the host's offspring. The parasite becomes, like Olson's (1993) dictator, "stationary" with respect to the host. In short, long-term parasites depend on host performance (Read, 1994; Dawkins, 2004), so that $r_s \gg 0$.

As in political economy, $r_s$ in sociobiology retains some potential for conflict. The appearance of within-group competition changes strategic calculations and favours virulent parasitism (Hamilton, 1972; Bremermann & Pickering, 1983; Frank, 1992 and 1996). When a symbiont pursues a strategy of restraint, it effectively invests in the public good of host efficiency. Competitor endosymbionts can exploit and outgrow them by pursuing a short-term extractive strategy. The consequence of within-group competition is therefore a rush to exploit the common resource—a classic tragedy of the commons recognisable to any economist.[73]

Similarly, if parasites evolve shorter-term strategies or means of moving between hosts (e.g. Le Clec'h et al., 2013), this reduces partner fidelity and therefore

[73] As Birch (2017, p. 101) writes, the formation of individuality also requires "the presence of mechanisms that suppress selection within the groups", i.e. this kind of within-system competition. See footnote 68.

dependence, increasing virulence—just as businesspeople who find success in countries made prosperous by public goods may shift their affairs to the Bahamas. As Dawkins (2004) writes,

> *to the extent that parasite genes pass to their own next generation via some sideways route which is not shared with those of the host genes, to that same extent the parasite will tend to be vicious and dangerous*. (p. 383)

Finally, long-term parasites may have domain-incomplete interests in their hosts. They may benefit from and even enhance the host's long-term *survival*, but have no interest in its *reproduction*, as discussed by Dawkins (2004, p.327).[74] Like the dictator who has an interest in economic output but not welfare, the alignment may be partial and consistent with considerable harm for the host.

If one could, in the manner of Hobbes, ensure that a sociobiological parasite was long-term, without rivals, and had guaranteed "succession", one could confidently predict the extensive development of mutualism. Guaranteeing such conditions is no easier among sociobiological actors than it is among humans.

### 4.1 The taxonomy in summary

Table 4.1 summarises the different modes of payoff transformation. Both *c*-modifying rules and payoff-dependence ($r_s$) generate cooperation among purely self-interested actors. Other-regarding preferences and genetic relatedness ($r_c$ and $r_g$) operate via intrinsic valuation of others' payoffs. All forms of $r > 0$ create lockstep payoff movement across some broad domain, though the domains of different mechanisms will differ. Rules, including institutions, rely on specific technologies of detection and enforcement that transform the payoffs of specific target strategies.

[74] Cheng (1973, cited in Dawkins, 2016, p. 327), for example, notes that "[a]lthough one generally considers parasites to be detrimental to their hosts and cause the loss of energy and poor health, instances are known where the occurrence of parasites actually induces enhanced growth of the host." Dawkins (2016) counters that "detriment" should be "defined in terms of reproductive success rather than survival and 'health'", and a parasite may nonetheless be "harming the host's reproductive success, even if it is at the same time promoting host survival."

**Table 4.1** Four modes of payoff transformation.

| | **Mechanism** | **Payoff transformation** | **"Motivation"** | **Biology** | **Society** |
|---|---|---|---|---|---|
| $c$ | Detection and enforcement | Targeted to particular cases or classes | Extrinsic | Individual-level reciprocity and similar | Individual-level reciprocity and similar, plus institutions |
| $r_s$ | Structural dependence | Lockstep, often domain-limited (e.g. long-term parasites, dictator), sometimes domain-general (e.g. genome) | Extrinsic | Vertical transmission, partner fidelity | Vertical transmission, partner fidelity |
| $r_g$ | Identity via vertical reproduction | Lockstep, universal (e.g. fitness between clones) | Intrinsic | Genetic relatedness | N/A |
| $r_c$ | Other-regarding preferences | Lockstep, stronger under between-group conflict | Intrinsic | N/A | Cultural relatedness or distance |

*The known is finite, the unknown infinite; intellectually we stand on an islet in the midst of an illimitable ocean of inexplicability. Our business in every generation is to reclaim a little more land.*

Thomas Henry Huxley (1887, p. 204)

*Innovation means a bad change, presenting to the mind, besides the idea of a change, the proposition, either that change in general is a bad thing, or at least that the sort of change in question is a bad change.*

Jeremy Bentham (1824, p. 143-144)

*Innovation has become the emblem of modern society and a panacea for resolving its problems… Before recommending of favouring innovation, it is good to weigh its consequences.*

Benoît Godin (2015, p. 133)

# 5

# Innovation: Exploring the design space

The discussion so far has related to the problem of crafting the first of the two keys to paradise: "direction". This is the problem of aligning individual and group payoffs so that the vector of payoffs for available strategies points up the optimality line. The second key, "pace", is no less essential. It is the magnitude of the vector in the social strategy space, and it is increased by *innovation*. Here "innovation" is defined broadly and encompasses most of its historical meanings: it covers the inventions and imitations of both political and economic entrepreneurs.

Section 1 examines the scale of the "design space", the space of all possible strategies that may one day become available to individuals. It introduces the concept of "innovation capacity", or the power to find useful strategies in this vast design space.

Section 2 places innovations within the social strategy space according to their private and social returns. They may be mutualistic, parasitic, public goods, or negative mutualisms. Ineradicable uncertainty about the innovations that will appear

in a given period, and their payoffs, results in an *uncertain cooperator-defector bias in strategic change*. In lucky periods of strategic change cooperation may advance; in unlucky periods it may retreat.

Section 3 places beliefs about the cooperator-defector bias in strategic change in historical context. Prior to the Industrial Revolution, innovation was a pejorative, regarded as largely undermining of, and parasitic on, social order—in Bentham's (1824) terms, a "bad change".[75] The twentieth century brought a complete reversal, such that innovation referred to "unforecastable improvements" (Awrey, 2012, p. 258). In the middle of this transition in the nineteenth century, Bentham argued against any prejudice about the effects of innovation. For Bentham and for this thesis, innovation simply produces "novelty"—for better or worse. In itself, the welfare implications of innovation are *ambiguous*.

### 1.1 The design space

The formula $rb > c$ can tell us something about whether a strategy is likely to be adopted and spread, once discovered. It cannot, however, tell us whether or how it might be discovered—how it is that novelties appear and populate the social strategy space, from some landscape of unrealised possibility.

First let us introduce that landscape as the "design space"—the space of all possible strategies. For the social scientist, the design space includes all possible culturally-transmitted strategies, each with its own utility payoffs. For the evolutionary theorist, it contains all possible genetic (and epigenetic[76]) designs, each with its own fitness consequences. These design spaces are unfathomably large and multi-dimensional, and can be treated, for practical purposes, as "infinite" (Eshel, 1996, p. 486). The Nobel prize winner Paul Romer (2016, para 13) explains this as the consequence of "combinatorial explosion". As Solée et al. (2013, p. 7) write, "pairs of components can be combined with some probability" so that "the number of new designs will increase with the number of already present designs."

To grasp the consequences of combinatorial explosion it is useful to consider some analogies. Herbert Simon (1972) and von Neumann and Morgenstern (1953) use the

[75] See also Francis Bacon's (1597) remarks on innovators in the epigraph to Chapter 1.
[76] Changes in gene expression that are heritable, but do not involve alterations to the DNA sequence.

example of chess, which, in the scheme of our universe, is an extremely small game world. There are on average a little over 30 moves available at one time, and we have the luxury of complete information about possible game states. Even so, the game is unsolved. Just three turns per player already gives 119 million possible games. For the average game of about 40 turns each, there are an estimated $10^{120}$ possible games—more than one trillion trillion trillion games for every elementary particle in the universe (Shannon, 1950).[77]

The design space for anthropic and biological strategies, meanwhile, makes the game world of chess infinitesimal. Daniel Dennett (1995) offers an engaging illustration with a metaphor borrowed from Jorge Luis Borges' short story, *The Library of Babel*. Given space constraints I leave that for A.5.1, and instead turn to Paul Romer's (2016) discussion of combinatorial explosions. He uses the example of exploratory chemistry, where chemists experiment with mixing some of the 100 or so stable elements together and observe the results. He notes that if the mix is of four elements, there are 94 million possibilities. With five elements, there are 9 billion—and so on. Then there are the possible proportions, pressures, temperatures, and configurations to consider. He continues:

> *To see how far this kind of process can take us, imagine the ideal chemical refinery. It would convert an abundant, renewable resource into a product that humans value. It would be smaller than a car, mobile so that it could search out its own inputs, capable of maintaining the temperature necessary for its reactions within narrow bounds, and able to automatically heal most system failures. It would build replicas of itself for use after it wears out, and it would do all of this with little human supervision. All we would have to do is get it to stay still periodically so that we could hook up some pipes and drain off the final product. This refinery already exists. It is the milk cow. Nature found this amazing way to arrange hydrogen, carbon, and a few other miscellaneous atoms by meandering along one*

[77] Even Rubik's cube, a simple system with only six faces of 3x3 coloured stickers, has $4.3 \times 10^{19}$ possible configurations. If we were to explore one new configuration per second it would take nearly ten times the current age of the universe to find them all.

> *particular evolutionary path of trial and error (albeit one that took hundreds of millions of years).* (para. 13)

As Romer summarises, "[t]here are incomprehensibly many discoveries yet to be found."

This view must be tempered, however, by the rapid decline in research productivity over the past century (Gordon, 2016; Bloom et al., 2017). This does not appear to reflect the limits of the design space—as Romer's example illustrates, it contains endless marvels—but rather the limits of our capacities to explore it. For now, at least, we might have plucked much of the low-hanging fruit.

### 1.2 Rational evolution: A product of quantity and quality

The challenge of innovation, both biological and social, is to find the exceedingly rare needles within these colossal haystacks. Hayek's (1979) model of optimisation and innovation, particularly in the context of institutions and social progress, makes for a useful foil. He argues that there are two opposite ways to locate useful designs: evolution and rationality. Social progress, he writes, is the result of a "trial and error" process, so that change in culture and institutions is "guided not by reason but by success" (Hayek 1979, p. 166). That is, Hayek argues human innovation is evolutionary rather than rational. Yet if we call every instance of learning from success "evolutionary", then we have defined all processes of discovery to be evolutionary unless they always find the best needle the first time. Hayek reaches his conclusion—that institutional progress is non-rational and therefore evolutionary—by largely defining rationality out of existence (see Steele, 1987, for a comprehensive dismantling).

"Evolution versus rationality" is a false dichotomy. The evolutionary model involves the generation of variation, selection via some filter, and retention of the selected variants (Lewontin, 1970). Directed problem-solving, via bounded rationality, is simply another imperfect means for generating variation (Popper, 1972; Zak & Denzau, 2001; Nelson, 2007; and Turchin, 2009). From there, designs are filtered based on their success or failure in practice, and selection over results produces evolution. Human innovation can be (boundedly) rational and evolutionary at the same time, and the cause of this is the insufficiency of rationality in light of the size and complexity of the design space—just as human rationality is unequal to the task

of solving chess. We make guesses and learn from experience. As Campbell (1974) writes,

> *increasing knowledge or adaptation of necessity involves exploring the unknown, going beyond existing knowledge and adaptive recipes. This of necessity involves unknowing, non-preadapted fumbling in the dark.* (p. 147)

Or in Schumpeter's (1934, p. 85) terms, for the innovator, "surveying all the effects and counter-effects of the projected enterprise" is an "impossibility".

Alchian (1950), in a paper admired and extended by Friedman (1953), applied the logic of Darwinian selection to the case of the firm to show that high degrees of market optimisation may occur even with bounded rationality. In Alchian's model, firm survival is based in *ex post* profits. The complexity of the firm's optimisation problem makes it impossible to solve, and so firms make educated guesses. Whatever the source of the variation in firm strategy, be it bounded rationality or random chance (e.g. the kind of accident that led Alexander Fleming to discover penicillin (Gaynes, 2017), George de Mestral to invent Velcro (Snell-Rood, 2016), and herdsmen to discover yoghurt by storing milk in animal intestines (Fisberg & Machado, 2015)), the process of selective retention on the basis of profitability leads the *population* of firms gradually towards optimisation.

Alchian (1950) suggests that even the completely random guesses may be sufficient. We cannot go this far. Truly random exploration would have to search through so much nonsense that it would require an impossibly large quantity of experiments.

In short, human progress is:

- necessarily evolutionary, because of the limits to rationality; and
- partly rational, because the design space is too large to explore randomly.

### 1.3 Innovation capacity

We can therefore define "innovation capacity" as the outcome of:

- *trial quantity;* and
- *trial quality*.

Such descriptions already exist in the economics literature: for Bloom et al. (2017), research output is the product of "the effective number of researchers and their research productivity" (p. 2).

With respect to trial quality, the lowest quality algorithm is random exploration without recursion—the random selection of volumes from the shelves of the library, and the retention of the fittest in a pile on the library floor. The challenges of random exploration are preposterously large, but they are solvable if we posit an equally preposterous quantity of trials. The "infinite monkey theorem" is a famous thought experiment illustrating the emergence of order from massively-repeated stochastic processes (Borel, 1913 being the first known instance of the idea). It holds that infinite monkeys randomly typing on infinite typewriters will eventually type Shakespeare's *Hamlet*. Indeed, they will type any finite text infinite times, with a probability of 1. Generating order via low quality trials requires a very large quantity: it would take around 10 billion monkeys just to get a roughly 50 percent chance that one of them will type "banana" in the first six letters (Isaac, 1995, pp. 48-50).[78]

At the other end of the spectrum, trial quality is ideal, and only one trial is required. The explorer simply picks the best book off the shelf and is done with it. We might call this a "single monkey theorem", if we can grant that the monkey is a sort of all-knowing god. While sometimes economists model *Homo sapiens* as intellectual god-monkeys, this is only a useful approximation when problems are simple. The rationality assumption may also apply to "average" behaviour, in the special case that errors balance out (i.e. the appropriate domain of rational expectations theory[79]).

Between the omniscient agent and the random explorer is the middle ground, occupied by all real-world systems. The simplest forms of biological evolution employ more sophisticated algorithms than random exploration. Fascinatingly, the "innovation capacity" of biological systems has increased over time, because faster and more sophisticated innovators out-compete the laggards. Evolution itself evolves (see Box 5.1).

[78] An average typewriter has 44 keys, giving a probability of $(1/44)^6$ for the word "banana".
[79] Sometimes biases do balance out. E.g. Galton (1907) found that the wildly varying estimates of the weight of an ox at a country fair averaged out to within one percent of the actual weight. The same is not true of average expectations about political and economic variables, though this differs by country (e.g. Grönlund & Milner, 2006).

Human rationality brought another step-change in trial quality. The earliest cognitive algorithms appear to have been best suited to imitation, with occasional accidents leading to novel behaviours. This facilitates the slow cultural development observed in our primate relatives, as in the macaques that famously learned to wash sweet potatoes (Karamura, 1954, as cited in Kawai, 1965). The sophistication of mental algorithms has grown over time, with large step-changes as the birth of complex civilisations brought writing, mathematics, and eventually scientific reasoning.

**Box 5.1 Evolution evolves: Biological innovations in trial quality**

The basic algorithm of biological innovation is the amplification of successful designs through natural selection and reproduction—increasing their quantity—and then the recursive mutation of these successful designs. This is many orders of magnitude more effective than random search.

But the mechanisms of evolution also evolve (Wimsatt & Schank, 1988; Dawkins, 1988; Stein & Lipton, 1989). What biologists term "evolvability", the capacity to generate useful innovations, has increased over time (Payne & Wagner, 2019). Organisms actively direct mutations to parts of the genome where change tends to be more favourable—for example, in the genes that encode immune responses—and away from areas where change is unfavourable. Stress increases the mutation rate, so that organisms "innovate" more when novel strategies are most needed (Rosenberg 2001; Foster 2007). Still more sophisticated is sexual reproduction, which involves the intermixing of successful designs to create novel combinations of modules that are already known to work (Maynard Smith, 1978). Horizontal gene transmission between organisms allows already-discovered genes to combine in novel ways with the genomes of other individuals or even other species (Koonin, 2016). So-called "jumping genes" allow genes to move around within genomes in a variety of complex ways. In some species whole genomes may intermingle (Roper et al., 2011).

As trial quality has increased, so biological evolution has squeezed fitter innovations from fewer trials. In short, even evolution is far too "rational" to try out most of the nonsense of the design space.

*Homo sapiens*' special talent is cognitive modelling or simulation (Popper, 1994). Humans use mental models to perform "experiments" internally, allowing them to rule out trials without testing them in practice. According to Popper (1994), the role of consciousness is

> *to anticipate success and failure in problem solving and to signal to the organism… whether it was on the right or wrong path to the solution of the problem.* (p. 17)

Human innovators can rule out most nonsense designs in advance. We need not experiment with a bicycle where the rider sits upside down so their head drags on the pavement, where the handlebars are set behind the rider, or the wheels are square and made of gooseberries. Popper's mental models allow us to forecast the consequences of these absurd designs, and so rule them out without trying them. They also allow single-trial innovations where problems are sufficiently simple—for example, the mathematician Robison (1960) predicted that a bicycle with square wheels will be able to deliver a perfectly smooth ride over terrain made of inverted catenaries of the right period, and such a vehicle was demonstrated by Stan Wagon in 1997. The impressive powers of cognitive simulation, armed with the sophisticated algorithms of scientific reasoning, have allowed scientists and engineers to build devices such as the Large Hadron Collider despite only limited experience with similar projects—although the process still involves considerable trial and error (e.g. see Collier, 2015). This is evolution at its most rational yet.

### 2.1 Innovations in the social strategy space

All innovations can be placed within the social strategy space introduced in Chapter 3. They may be mutualistic, parasitic, public goods, or negative mutualisms. In any period, the distribution of innovations in this space shapes the direction and pace of change.

If innovation was conducted with perfect foresight, individuals would only ever discover strategies that deliver positive returns ($B_i > 0$). Such strategies may be parasitic or mutualistic. However, because innovation is a trial-and-error process occurring in a design space overwhelmingly dominated by maladaptive strategies, most novelties will be unfavourable—that is, biased towards being in the bottom half

of the social strategy space (negative $B_i$). The basic principle is that there are always many more ways to become disordered than to become ordered.

Box 5.2 discusses the biological case. For human innovation, with the advantage of Popperian mental models, this rate of useless or harmful innovations is much lower—though it still dominates. We must sift through many hopeless ideas and projects in order to find one treasure. In present-day advanced economies, more than two-thirds of start-ups fail to achieve profitability (Eisenmann, 2021), and most of the rest fail to deliver more than a modest income. A much greater number of new ideas receive some investment, at minimum the investment of exploratory effort, but are stillborn and never progress to the stage of business formation.[80]

**Box 5.2 Sociobiology: The costs of trial-and-error**

Biologists universally agree that most mutations are harmful. Dennett's (1995) "Good Tricks" discovered in the evolutionary process reduce this bias compared to a counterfactual world with purely random exploration, but the fit "needles" are still very rare (pp. 109, 222-223). Of meaningful[81] point mutations, an estimated 99.9 percent are deleterious or fatal (Remine, 1993). The same logic applies to innovations that occur within a cooperative group: they will usually be unfavourable for that group, so that payoffs are biased towards the left side of the social strategy space (negative $B_g$). But progress occurs because the scant few innovations that provide positive returns tend to be selected, retained, and further refined, by either boundedly rational or natural selection.

Thus, the enormity and complexity of the design space ensures that most innovations are negative mutualisms—strategies that are good for no-one. They rarely present risks to cooperation, because they will tend to be weeded out by individual optimisation. They are complexity's tax on innovation. The main consequence is that innovation can impose quite considerable costs on species, societies, and firms in the short-run, even though innovation is the only pathway to survival and growth in the long-run (e.g. Solow, 1956 and Swan, 1956 in economics). In Chapter 6 and 7 we will see that in firms, societies, and indeed biology, this sometimes leads to perverse

[80] A popular statistic in writing courses is that 97 percent of started books are never finished, and around 50 percent of doctoral candidates fail to complete their programs (Bowen & Rudenstine, 1992).
[81] Some DNA mutations are non-meaningful or "silent", because they do not produce protein changes.

disinvestments in innovation capacity—producing short-term gains but long-run decline.

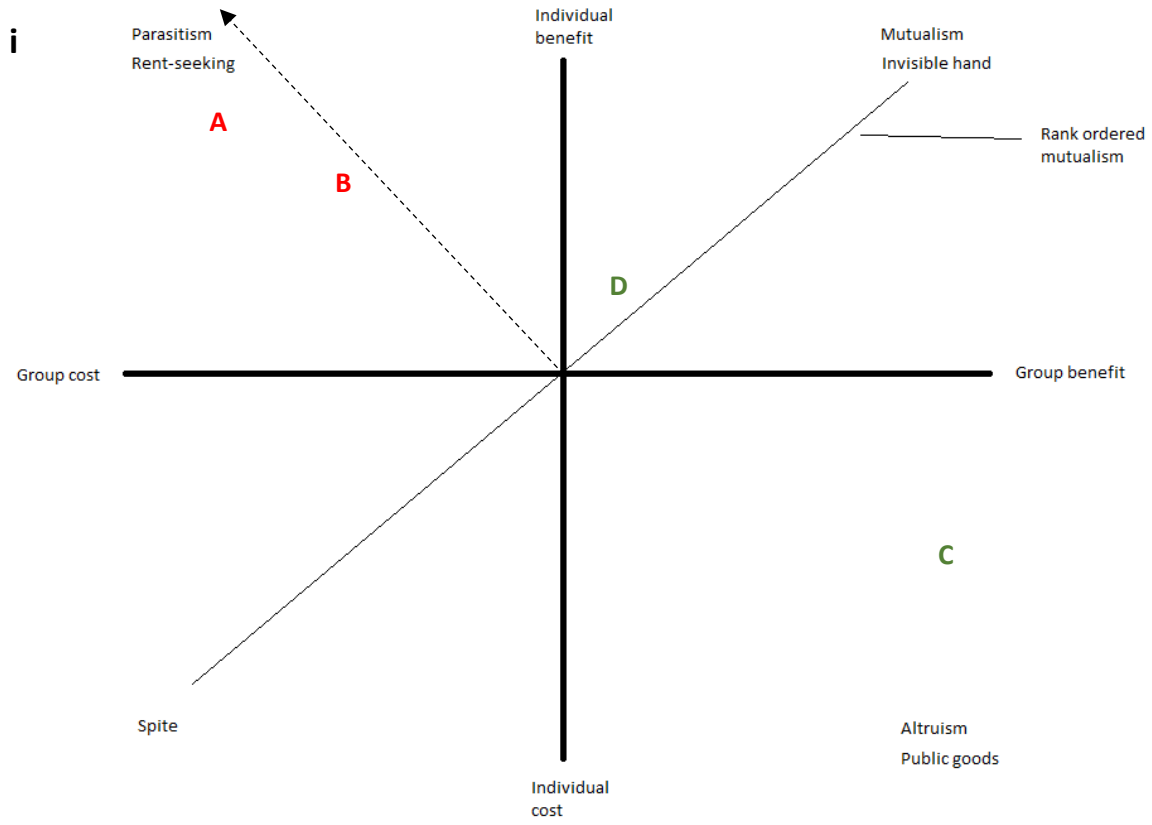
### 2.2 The cooperator-defector bias in strategic change

Our focus will be on strategies with some staying power—those that are favourable for individuals, and so lie in the top half of the social strategy space ($B_i > 0$). Again, they may be parasitic or mutualistic. Following Baumol (1990), we do not know whether the entrepreneur will allocate their efforts to "productive", "unproductive", or "destructive" activities. Murphy et al. (1991, p. 506) similarly observe that "talent goes into activities with the highest private returns, which need not have the highest social returns". As noted in Chapter 3, the paucity of growth in Hobbes' state of Warre, and the tendency for parasitism to snuff out growth historically (Jones, 1988), suggests that there is no inherent design space bias in favour of mutualistic strategies. Here I take the average group effect of individual innovations as varying stochastically around a mean value of zero (though for efficient societies, this is an unlikely best case).[82]

This brings us to an important concept: the *cooperator-defector bias in strategic change*. This terminology draws on an analogy to factor biases in technical change in the economics literature, where a period of change may unexpectedly favour the returns to labour, capital, or skill. In the same way, periods of technical or strategic change may be biased towards mutualism or parasitism, or may be unbiased.

The cooperator-defector bias in strategic change is reflected in the direction of the vector of innovation payoffs in a given period. Figure 5.1.i illustrates a case where, by chance, individual innovation favours parasitism and a decline in cooperative efficiency. The parasitic strategies might be complex, poorly-regulated derivatives that increase systemic financial risk, new ways for monopolists to erect barriers to the entry of competitors, or "deepfakes" that undermine information markets. Figure 5.1.ii is a case where the bias happens to favour mutualism and growth.

[82] For the reasons discussed in Section 2.1—that there are more ways to become disordered than ordered—individual innovations are on average deleterious for the group. For reasons discussed in A.5.2, innovation becomes increasingly biased towards parasitism as societies approach the efficiency frontier. This may be important for understanding innovation in capitalism today, and is worth further investigation. But the arguments of this thesis do not require it.



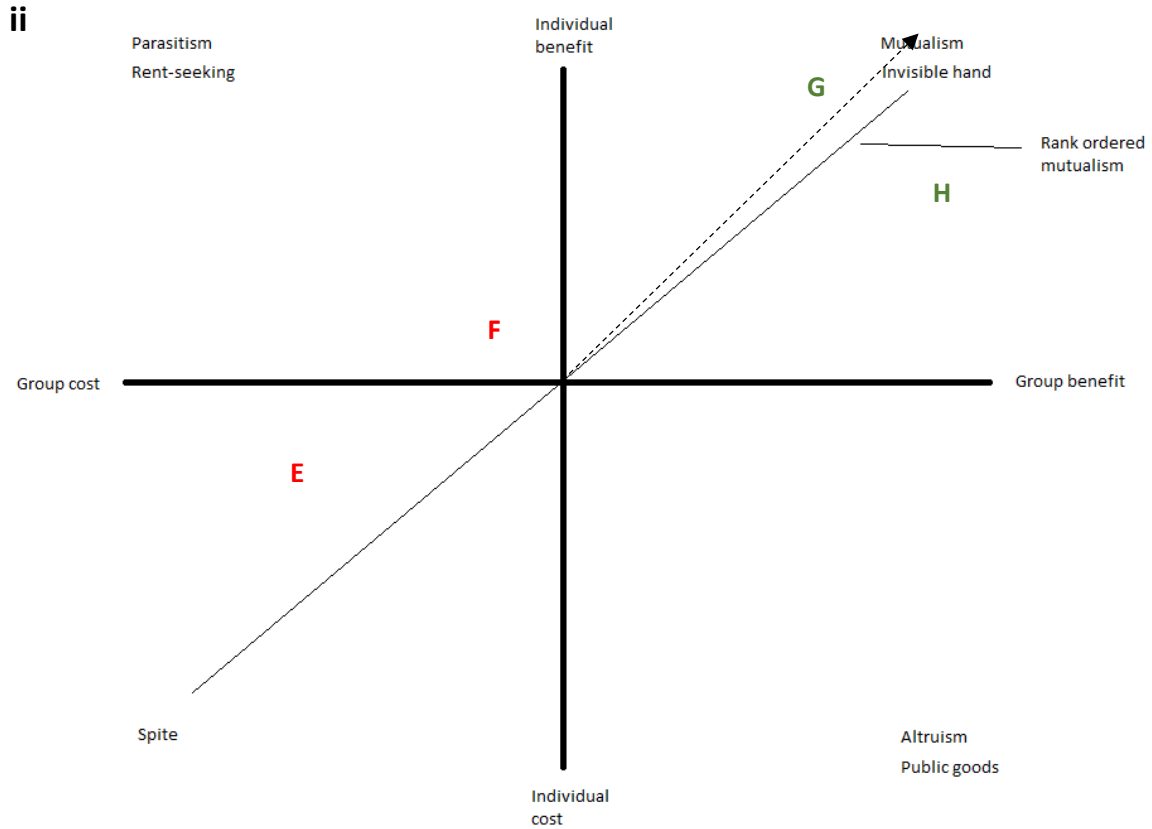




**Figure 5.1** Innovation in two periods, leading to novel strategies A to H. Group payoffs are negative where the letters are red, and positive where green. ***(i)*** Payoffs in this period favour parasitism (A and B especially). Note that innovation C is a beneficial public good, but the transformational environment is insufficient to make it privately favourable. ***(ii)*** Payoffs in this period favour mutualism (especially G and H). The dotted lines indicate the average direction of individually beneficial strategies—nearly optimal in ii, but unfavourable in i.

These might be more efficient manufacturing techniques, new products that satisfy changing consumer preferences, or a novel public good made privately viable by a Pigouvian subsidy. The dotted lines show the average direction of the vectors of the strategies that individuals will pursue in these two periods (i.e. those where $B_i > 0$).

We are necessarily uncertain about the cooperator-defector bias of the next period and can treat it as stochastic. We lack knowledge about the strategies that will be discovered, and how they will combine with myriad existing strategies (including, in the human case, the political context), and what the resulting payoffs will be. Again, as Schumpeter (1942, p. 83) writes, their "true features and ultimate effects" can only be understood, and often dimly, in retrospect.

### 2.3 Transforming the cooperator-defector bias

The cooperator-defector bias in strategic change, as discussed so far, is a property of the *raw* payoffs of innovations. These innovative strategies may, of course, be transformed by rules or commitment, which affects whether they are privately viable and so retained.

Take the case that $r = 1$ for whatever reason. All the strategies A to H in Figure 5.1 will be shifted to the optimality line. Despite 5.1.i being unfavourable in raw terms, and 5.1.ii favourable, under $r = 1$ they in fact favour similar amounts of growth. In Figure 5.1.i, the two parasitic strategies will be rejected and the public good, H, will become favourable. The same applies under ideal rule-based transformation. In Baumol (1990, p. 894), this is the "reward structure of the economy" that directs entrepreneurial allocation to productive or destructive activities.

## 3.1 Half an economics? The assumption of mutualism in innovation theory

The terms "innovation" and "entrepreneurship" are, in their most technical sense, "shocks" to the strategic landscape of the economy. But Awrey (2012) writes that

> *beneath this veneer of academic objectivity there survives a marked tendency within the literature to view these unanticipated shocks as being more in the nature of "unforecastable improvements."* (p. 258)

In the terms of this thesis, theorists tend to assume a *perpetually favourable* cooperator-defector bias in strategic change.

Fascinatingly, the opposite was once true. Prior to the nineteenth century, the term "innovation" was largely restricted to the political domain and tended to be regarded with suspicion, because it entailed the upsetting of established orders (Godin, 2015; 2016).[83] Innovation was an unforecastable hazard. This negative valence became especially acute in the context of the Reformation and Counter-Reformation. In England during the reigns of Edward VI and Charles I, explicit laws were introduced against innovation and innovators put on trial. This extended to technical innovations, as in the case of the 1551 prohibition of the gig-mill, which was not repealed until 1809 (Mokyr, 1992).[84] Strong echoes of this attitude persisted into the early nineteenth century, per the quote from Bentham (1824) in the epigraph to this chapter. To the modern ear, the definition of innovation as a "bad change" is shocking. Bentham's purpose was to repudiate this view:

> *to say all new things are bad, is as much as to say all things are bad, or, at any event, at their commencement … Whatever is now establishment was once innovation*. (p. 144)

Since the time of Bentham we have been dazzled by the miracle of long-run growth, driven by the "*A*" residual in the Swan-Solow production function. Innovation and the entrepreneur are the heroes of progress, ever expanding the pie. They are seldom the villains plotting to divide it. When the economist or layperson thinks of innovation, they tend to think of technological advance—of the printing press, the internal combustion engine, the personal computer, and so on. Countless books and papers wax on the economics of industrial innovation, "soft" innovation in the design and creative industries, the optimal "innovation policy mix", and so on—an endless march towards greater choice, higher quality, and lower cost. In the financial sector, for example, Frame and White (2004)

[83] Some remnants of this use in Smith's *TMS*, where he refers to the "often dangerous spirit of innovation" (p. 273) with specific reference to political innovations.
[84] The link between religiosity and the suppression of innovation persists today, in the United States and across the globe, and manifests in lower rates of patenting, less creativity, and negative attitudes to science and technology (Bénabou et al., 2015).

> *define financial innovation as something new that reduces costs, reduces risks, or provides an improved product/service/instrument that better satisfies financial system participants' demands.* (p. 118)

For Garcia-Macia et al. (2019, p. 1) it is the means by which firms generate productivity growth, either by creating "new products" that "displace the products of competitors", or by inventions that "improve their existing products" (see Aghion et al., 2021, for another recent contribution). For Schumpeter (1942), who describes innovation as "creative destruction", the short-term disruption brought by new technologies is expected to deliver gains in the long-run.[85]

> *The opening up of new markets, foreign and domestic, and the organizational development from the craft shop and factory to such concerns as U.S. Steel illustrate the same process of industrial mutation—if I may use the biological term—that incessantly revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one.* (p. 83)

This tendency to accept "direction" as solved and focus purely on "pace" extends even to the main threads of evolutionary economics, beginning with Nelson and Winter (1973). Their critique of conventional economics, as well as the views of various authors compiled more recently (Nelson et al., 2018), can be understood as retaining the general assumption of cooperativity while exploring interventions that might increase the rate of innovation, such as subsidies for R&D.

To say these quotes and snapshots encapsulate these theorists' *complete* view of innovation would be unfair. Nonetheless, the attitude to innovation has transformed over the centuries such that the primary question in innovation economics has been about how to maximise pace. This is partly why economics has struggled to explain stagnation; it is largely a theory of health rather than a theory of sickness. There are obvious exceptions, including in the literatures on rent-seeking (e.g. Tullock, 1967) and monopolistic competition (e.g. Chamberlin, 1933), and in much of the practical work of economists that deals with ground facts rather than high theory. Such work is taken as the study of exceptions to the rule, and seldom modifies the metaphors

[85] Schumpeter (1911) does make some allowance for innovation in establishing monopolies, which can cover some—but not all—of the parasitic innovations with which we are concerned.

and intuitions that guide our thinking about markets and regulation. Volcker (2009) famously suggested that most innovations in finance over the last few decades are unproductive and primarily related to rent-seeking; it is an important (and in this thesis, very likely) possibility that in markets left to themselves, it is productive innovations that may be the exception.

In short, the dominant economic theories of innovation give us only half a political economy. It is a *political economy of the mutualism quadrant*, and of the optimality line in particular. This thesis argues for a political economy with agents who generate innovations across *both* of the quadrants in the top half of the social strategy space. At first glance, it may not be clear what exactly this means. The rest of the thesis aims to draw out the implications.

**Box 5.3 Innovation as novelty, for better or worse**

It may be suggested that we restrict the term "innovation" to changes that enhance social welfare, and "entrepreneurs" to agents that develop such innovations, while reserving other terms for activities and actors in the parasitic quadrant. This proposal is unattractive. Whether an activity is mutualistic always depends on the welfare function selected and the particular economic and political conditions in which that activity occurs. The same actor and novel strategy may be considered an "entrepreneur" and an "innovation" under one welfare framework and within certain conditions that give it positive group returns ($B_g > 0$), but not an innovation under another framework or conditions that made its group return negative ($B_g < 0$). It is better to regard the term as a concrete description of the process at hand: an innovation is a new strategy in the social strategy space. Without knowing more, we cannot say where it is good or bad. Indeed, this was the approach Bentham advocated in 1824:

> *The idea of novelty was the only idea originally attached to the term innovation, and the only one which is directly expressed in the etymology*. (p. 218)

In this view, increasing the pace of innovation is neither desirable nor undesirable, until we know something about its direction.

# PART II

## Running the Red Queen's race:
## Rules and commitment in a world of innovation

*Life is a Sisyphean race—run ever faster toward a finish line that is merely the start of the next race.*

Matt Ridley (1994, p. 174)

# 6

# Introducing the Red Queen's innovation race: The problem of maximising "pace"

This chapter introduces a concept that is essential for understanding competition and cooperation in our innovative world, and that plays a core role in the thesis: the innovation race. The individuals of our world, and the groups they form—e.g. firms, industries, states, and so on—compete with one another for limited resources. Relative performance in this competition is not static, but dynamically changes as new strategies emerge.

This race takes complex forms. In practice, it occurs at multiple levels: individuals may compete within teams, teams within firms, firms within industries, industries within national economies, and nations within an international system. Performance at one level affects outcomes at other levels. Forms of individual competition within a team, for example, may support or undermine that team's performance.

This chapter starts with the simplest, one-level innovation race, to draw out some universal features of such races. In this simple model, the problem of "direction" is solved—that is, there is no possibility of conflict within a group, and the groups doing the racing are perfectly cooperative. The only remaining problem is that of the group's "pace": to compete most successfully, the group must maximise the rate of innovation. In the terms of Chapter 5, we are concerned here with the familiar *economics of the mutualism quadrant*, where innovation is benign and more is always better.

Section 1 examines the forces that drive ideal groups to innovate. We pursue prosperity for its own sake, but group survival also depends on keeping up in competitive innovation races against other groups. Each racer responds adaptively to the changing strategies of others, in a perpetual battle for competitive advantage.

This is Van Valen's (1973) *Red Queen's race*, a concept of great significance in sociobiology and, I suggest, political economy. Performance, and survival, in the race is predicted by innovation capacity. This has produced a tendency for innovation capacity to progressively increase over the history of cooperative systems, both biological and anthropic.

Section 2 examines the convergence of both biological and anthropic systems on one ingenious tool for maximising pace: the *Darwin machine*. Darwin machines solve extremely hard allocation problems by using *domesticated competition*, where competition between individuals is allowed to occur in constrained ways that serve the group. Markets are an obvious and important example. In biology, analogous systems are found in organisms. In both cases, they accelerate group optimisation by enabling massively parallel experimentation, utilising distributed knowledge, and avoiding the need to encode knowledge in some centralised repository.

Domesticated competition is essential if the group is to compete, but this chapter has assumed that the domestication of innovation is automatic and ideal. The next chapter lifts this assumption, creating a new within-group Red Queen's race.

### 1.1 The Red Queen's race

We take direction as solved. Given the vector of available strategy payoffs points up the 45-degree optimality line, we can maximise group benefit by maximising the magnitude of that vector—by maximising pace.

Ideally cooperative groups may control all internal conflicts, but they still face the problem of adapting to a changing external environment.[86] Exogenous changes in factors such as prices and preferences (for firms in markets) and the climate and available resources (for organisms in biology) may threaten survival.[87] Figure 6.1 illustrates a case where exogenous change shifts the adaptive zone for a particular variable over time—for example, a change in consumer preferences requires

[86] Environmental changes tend to fall on a power law distribution (Reed & Hughes, 2002), so a wide range of shocks to both social and biological systems appear with exaggerated frequency compared to a normal distribution.
[87] Of course, there are also advantages to benign innovations even in the absence of external environmental change, given they allow larger populations to be supported and/or raise the average standard of living.

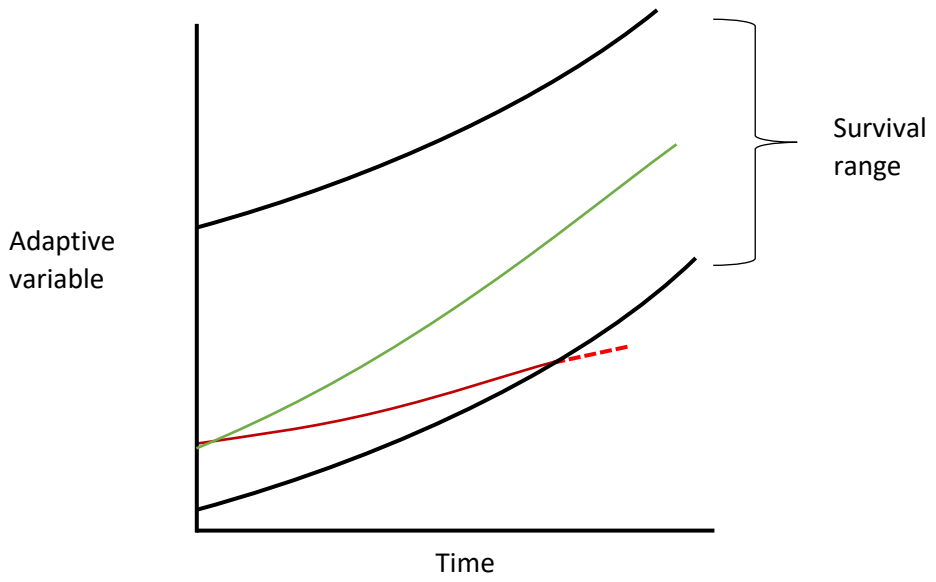



**Figure 6.1** Movement of, and within, the adaptive zone. The group or population innovates in the right direction in both cases, but the group or population following the red pathway fails to keep up with environmental change.

adaptation by a firm, and a change in the climate requires adaptation by a species. The group that follows the red adaptation pathway is unable to keep up with change, while the green pathway with a higher innovation capacity stays within the new adaptive zone and survives.

More important in this thesis, however, is the problem of adapting to changes in competitor strategies. Competitor groups, i.e. other firms, states, organisms, etc., are perpetually innovating and adapting. Two groups are engaged in an antagonistic relationship each develop innovations that raise their relative competitiveness. If one actor's innovation gives them a fresh advantage over competitors, this raises the payoffs to, and favours survival of, competitors with counter-innovations that curb or reverse that advantage. This produces a co-evolutionary arms race that Van Valen (1973) termed the "Red Queen's race", and that was introduced in Chapter 1 as one of the core concepts in the analytic framework. Again, Van Valen draws it from Lewis Carroll's novel, where

> *it takes all the running you can do, to keep in the same place.* (1897, p. 50)

In a co-evolutionary race, competitors must run as fast as they can in order to stay in the same position *relative* to one another. Bergstrom and Lachmann (2003) write that, "[a]s a result of this Red Queen process, each species is forced to evolve ever more rapidly just to break even" (p. 593). I also noted in Chapter 1 the widespread use of the Red Queen's race in works of economics and political economy over the last two decades.

Van Valen did not intend to suggest that the racers will actually stay in the same relative positions in the long-run, as did Alice and the Red Queen. Instead, competitors' fitness is variable over time—he suggests normally distributed—which translates into *variable gaps in the relative performance* of competitors. Note this is related to the notion of the uncertain and stochastic cooperator-defector bias in strategic change, except here we are dealing with a given period's bias in the favourability of innovations to one competitor or another. There is no cooperation or defection. The next chapter turns from a race between two antagonists to a race between cooperators and defectors, so returning us to the cooperator-defector bias described in Chapter 5.

In any case, Van Valen notes that if the distributions of the two racers' competitiveness are the same, then the expected competitiveness gap between them is also normally distributed but with twice the variance. If, by chance, the gap between the racers reaches a critical threshold in a given period, as for racer A in Figure 6.2, the result is extinction. Van Valen suggests that stochastic variability in the relative performance of a given racer over time explains why, in his famous finding, the age of a species and its extinction likelihood are independent—i.e. there is no correlation between the period a given species has existed so far and the likelihood of its extinction in the next period.[88] Species that last longest are not those with "unbeatable" strategies. Rather, species run the Red Queen's race until they have a sufficiently unlucky period, and the longest-lived species are simply the luckiest.

[88] Analogous to the flipping of a coin: extremely large samples will produce long strings of heads or tails, but this does not affect the probability of the next flip.

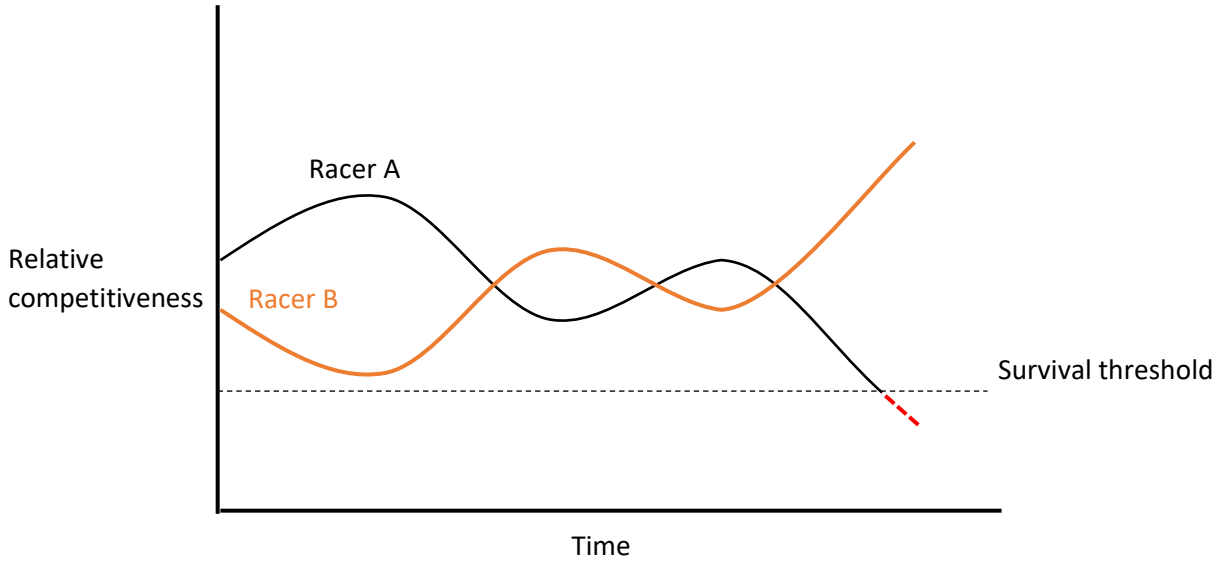



**Figure 6.2** Competition between groups or populations, where the competitiveness of one reduces that of the other. Over long periods within a stable environment, strategic change leads to oscillations in competitiveness, eventually driving Racer A to extinction. In biology this may be, for example, competition between a tree and grass species, ending as the development of C4 metabolism in grasses allowed the replacement of forest with savannah (Beerling & Osborne, 2006).

### 1.2 Unpredictability and precarity: The absence of "unbeatable" strategies

The race is unpredictable in part because of the general absence of "unbeatable" strategies (and the impossibility of knowing that a strategy is unbeatable, so that it is always hubristic to presume that one has identified such a strategy). This reflects the enormity and multi-dimensionality of the design space: almost all strategies producing some advantage have a set of possible counter-strategies, so a racer cannot simply develop an unbeatable strategy, stick with it, and be assured of winning forever. In the biological case, an estimated 99.9 percent of species that have lived on our planet are now extinct—the strategies they pursued became uncompetitive and they lost the race. Even if some rare strategy is in fact unbeatable, there is no way of identifying that this is so in advance.[89]

Identifying unbeatable strategies, and the Nash equilibria they produce, is the goal of game theory in both the social sciences and biology. It is possible where the range of

[89] There are rare species, "living fossils", that appear morphologically unchanged for hundreds of millions of years—yet this conceals ongoing microphysiological adaptation that is no less significant (Lindholm, 2014). In any case, we cannot predict which relatively new species today will remain unchanged for hundreds of millions of years into the future.

strategies is fully specified. In biology, identifying unbeatability is a feasible goal of *short-term* evolutionary studies (e.g. Hamilton, 1967). But this is impossible in long-term studies that allow for genuine innovation, which reaches into parts of the design space, and novel interactions, that we have no means to predict. As the evolutionary biologist Ryan Gregory (2007) suggests, belief that a given strategy is unbeatable simply reflects a lack of imagination. Nobel laureate Francis Crick, who co-discovered the structure of DNA, called this Leslie Orgel's Second Rule of evolution: "Evolution is cleverer than you are" (see Dennett, 1995, p. 74).

Individual optimisation will be cleverer than any theorist or designer. This is an essential fact to carry with us as we study any of the fields to which the Red Queen's race applies—political economy, biology, and artificial agents. A key argument for the market society is that individuals innovate in ways that theorists and regulators cannot predict, although this observation has profound implications that we are not yet ready to explore.

### 1.3 Predictability via innovation capacity: The Tortoise and the Hare

The absence of unbeatability sometimes makes the Red Queen's race somewhat predictable, in respect to the broad question of who will win the race. Winning the race is largely a matter of luck, but the game of chance can be tilted in one's favour. The racer who can innovate more rapidly and effectively increases the probability of being lucky—of discovering a useful strategy in any given period. The Red Queen's race therefore generates a *second-order race over innovation capacity*—pressure, as Bergstrom and Lachmann (2003, p. 593) write, to run "ever more rapidly". Following Box 5.1 in Chapter 5, evolution evolves. The importance of innovation systems for long-term performance of firms, militaries, nations, and any other kind of competing social group is broadly understood and needs little further explanation.[90]

Disinvesting from innovation is therefore suicidal. I call this the *Tortoise and the Hare principle* after the fable: whatever the size of a racer's lead may be, if it falls asleep at the side of the track, then the race becomes highly predictable. The Hare is guaranteed to lose the Red Queen's race. Withdrawing from the race results in a

[90] As Pierre & Fernandez (2018, p. 140) write, "[i]nnovation capacity represents a firm's ability to innovate continuously ahead of its competitors."

monotonic decline in relative performance, as one's competitors gradually—or sometimes rapidly—accumulate advantages.

The problem of the sleeping Hare will be of great interest for two reasons. First, because large asymmetries in innovation capacity allow predictions despite the uncertainties of the Red Queen's race. Second, because neglecting innovation is a common error in both anthropic and biological systems, and has severe consequences in both. I explore some biological cases in A.6.1 of species committing "evolutionary suicide" (Ferrière, 2000) by playing the Hare. In brief, individuals that withdraw from innovation may win higher short-run returns, but they undermine the common good of collective innovation and the population goes extinct.[91] In markets, one example is the firm that is stripped of R&D capacity to maximise the short-term rents that can be extracted by its managers (e.g. Bushee, 1998), with predictable effects on long-term performance. This may be a product of perverse managerial incentive structures that weight short-run over long-run returns, or it may reflect optimisation errors (i.e. foolishness or excessive caution). Another example is a citizenry that votes to reduce tax rates, boosting short-run incomes but undermining the public goods that made them rich in the first place. Or we may think of the guileless dictator who feels naïvely secure in their power and fails to creatively play palace games.

Later I will argue that this helps us to explain the problem with many economic theorists and regulators believing in self-regulating markets, and so choosing hands-off regulatory strategies. This is similarly naïve—they play the Hare falling asleep in the middle of a fierce institutional Red Queen's race. Theorists and regulators believed that, in free markets, they had discovered an "unbeatable" strategy—a fundamental misunderstanding of the nature of the world we inhabit. Creative market actors, meanwhile, busied themselves with exploiting the simple designs of classical liberals and libertarians.

Understood within the framework of a Red Queen's race, the consequences of belief in the unbeatability of free markets are wholly predictable. This will be made

[91] Interestingly such findings are taken to refute "adaptationism", the biological equivalent of the "invisible hand" theory.

concrete, and we will explore its deeper implications for regulatory design, via some case studies in later chapters.

## 2.1 Maximising pace: Domesticated competition and the Darwin machines

The nearly universal adoption of markets is a continuation of the process of growing innovation capacity throughout human history. Those states that failed to adopt markets played the Hare, choosing suicidal strategies with lower rates of innovation.[92] The race continues, of course—there are many possible ways to govern markets, some adaptive and some grossly maladaptive—and this is a matter for later chapters. For now, I will build on Hayek's views about their indispensability.

Markets employ *domesticated competition*. Fascinatingly, the general properties of this solution are common to biology. Before getting into the details, I must clarify the terms "cooperation" and "competition", given there is some confusion about the nature of cooperation and competition among both laypeople and social scientists.

Social scientists employ a common-sense interpretation of competition that emphasises agent psychology and behavioural processes. In this view, it seems contradictory to say that ideal markets harness competition and yet are perfectly cooperative. Hayek, for example, frequently wrote as if cooperation and competition were dichotomous. Cooperation, he says, is "agreement" on means and ends. It "makes sense in a small group whose members share particular habits, knowledge and beliefs", but is senseless in "the extended order" of the market society. In the latter, he asserts, it is "through further competition, not through agreement [that] we gradually increase our efficiency" (Hayek 1988/2013, p. 19). Like Hayek's dichotomy between evolution and rationality in Chapter 5, this too produces confusion. In practice, competition may take forms that harm the group—notably the abundant theft, violence, and rent-seeking during the long stagnation of human history—or in constrained forms that produce enormous benefits for the group. *Ideal* constraints on competition rest on an extraordinary degree of cooperation, indeed *ideal* cooperation. Hirshleifer (1978) writes that

[92] E.g. the Austro-Hungarian empire in the 19th century, and later the USSR and China. North Korea remains a holdout, but like the USSR still relies on markets cryptically.

> *[t]he Invisible Hand… requires a severely constrained form of competition; vying to engage in exchange with third parties, and doing so only by offering better terms under an ideal system of property and law.* (p. 240)

To understand this, it is helpful to be clear about whether we are talking about *interests* or *processes*.

The biologist's sense of cooperation and competition is defined in relation to individual and group *interests*. Szathmáry and Maynard Smith (1995) and Frank (2003) identify the total subservience of within-group competition to between-group competition as one of the defining features of ideally cooperative organisms. Crucially, there are no individual interests distinct from the group interest, and any changes in individual success are consistent with their effects on group competitiveness.[93] This is fully commensurate with the ideal market: If market rules are permanently perfect, there can be no individual interests distinct from the group interest.[94]

The systems we are discussing—ideal markets and organisms—are cooperative with respect to group interests, but involve harnessing competitive *processes*.

Let us therefore define two types of competition.

- First, "A-competition", where there is competitive antagonism between individual and group *interests*. This involves various forms of zero and negative-sum conflicts over resources.
- Second, "O-competition", referring to cases where individual and group interests are aligned, but the group uses a *process* of competition between lower-level elements in the system. I give it this name because its purpose is group-level *optimisation*, i.e. the maximisation of pace.

Most of human history has been held back by A-competition, while it is recent innovations in O-competition that have brought about the Anthropocene. A

[93] E.g. whether a particular gene proliferates or is extinguished within a genome—its individual success in the group—is determined by the consequences for genome (i.e. group) fitness.
[94] Analogies can be drawn with sport: the rules that govern competition that ensure that private striving serves the relevant societal values, such as entertainment, the safety of the audience and players, and so on. The private interest in victory is made consistent with the societal interest, and wherever competitor innovations threaten to undermine the game, rules are remade.

perpetual challenge is maintaining O-competition: lower-level elements may discover ways of competing that violate or escape the constraints of competition, and that are contrary to group interests. This turns O-competition into A-competition, and group performance declines. We assume cooperation in this chapter, so this is a matter for the next.

### 2.2 Darwin machines vs. central planning: Solving hyper-adaptive problems

The most sophisticated driver of innovation and optimisation found in biological and social systems is what Calvin (1987) terms the "Darwin machine". This is to be distinguished from the Turing machine. While Turing machines apply a set of given rules to an input to compute an output, Darwin machines employ an iterative process of diversity generation, transmission, and selection to produce adaptive outcomes. As DS Wilson (2002) writes, they are "machines" in the sense that the constraints on this process must be highly managed, and "Darwinian" because within those parameters they allow unguided, and competitive, adaptive change. That is, they harness O-competitive processes for some collective purpose. This form of optimisation has been exploited by biological systems for hundreds of millions of years. Although it may have been possible for human societies learn this trick from biology, instead this is a case of convergent evolution—of different systems stumbling upon the same strategy independently.

There are exquisite examples of O-competition in *Homo sapiens* biology. During brain development hundreds of billions of neurons compete to extend axons, find partners, and establish active synapses. If they succeed they are rewarded with brain-derived neurotrophic factor and other molecules that instruct them to stay alive (Chan et al., 2002; Becker & Bonni, 2004; Deppmann et al., 2008; and Je et al., 2012). As many as 50 percent of neurons fail to compete and are selected against, undergoing programmed cell suicide (Dekkers et al., 2013), where they neatly liquidate their assets like a rule-obedient bankrupt firm. Finely-tuned O-competition is critical to normal mental function (Kim et al., 2010), and excess neuron and synapse survival is associated with diseases such as autism (Courchesne et al., 2011) and neuronal cancer—the latter a form of A-competition between the organism and rogue parasitic genes.

A second example is the adaptive or acquired immune system (as opposed to the lesser-known, and more primitive, innate immune system); the interested reader can find a discussion in A.6.2 of this, along with some other examples.

As DS Wilson (2002, p. 31) writes, biology employs Darwin machines to solve "hyper-adaptive problems", when "physical and social environments become sufficiently variable". The simplest lifeforms possessing nervous systems live in such predictable environments that they have fixed neural structures—as in the nematode *Caenorhabditis elegans*, whose 231 neurons are always organised into the same pattern (it is the first organism to have its neural "connectome", or wiring diagram, fully mapped—see Cook et al., 2019). This might be termed a "centrally planned" nervous system, given most of its specifications are transcribed in the genome rather than emerging from individual adjustments to local conditions. For humans, the range of possible behavioural adaptations are, like the range of pathogens encountered, so diverse and unpredictable that blind genetic selection cannot provide adequate ready-made solutions. A fixed neural architecture for behaviour, or a fixed suite of antibodies for detecting pathogens, would prove fatal.

Darwin machines have three advantages over conventional natural selection:

- First, they greatly accelerate the pace of innovation via the *massive simultaneity of "trials"* within the lifespan. They allow adaptive change to occur not only in less than a single generation, but, in the case of the immune system and the brain, on the order of days, hours, or even seconds.
- Second, these trials occur within and respond to local conditions, and can utilise information about those conditions that is impossible to centralise in the instructions of the genome. In short, they access and use *distributed knowledge*.
- Third, they avoid the need to take this huge volume of accumulated knowledge and *encode it in a plan*. Even if blind evolution were fast enough or the environment sufficiently stable for conventional natural selection to work, the arrangement of a hundred billion neurons, or the $10^{12}$ possible antibodies, could not possibly be encoded into the genome.

### 2.3 Darwin machines in the Anthropocene

All of these advantages of domesticated competition will be familiar to economists. The market is a Darwin machine employed to solve a hyper-adaptive resource allocation problem—how to satisfy an extraordinary multiplicity of human preferences via an extraordinary multiplicity of possible means, under scarcity, and with both ends and means in perpetual flux. The market similarly allows massively parallel exploration of the design space, the use of distributed knowledge that is unavailable to a central planner, and avoids the need to track all economic activity in some central government log. These advantages are variously described in Barone's (1908) early warnings about the impossibility of central planning, the better-known contributions of von Mises (2022), Keynes' (1936; see also Keynes, 1940, quoted in Skidelsky, 2007) view of the advantages of decentralisation, and Hayek's (1945) elaboration of the distributed nature of knowledge that helped push this understanding into the mainstream.

Many non-market institutions also exploit forms of domesticated competition for the purpose of optimisation. In the academy, scholars compete within a selective environment in which the fittest are those who can persuade their peers of the value of their contributions to the body of knowledge. Firms and bureaucracies construct selective environments that, ideally, favour individuals who make the greatest contributions to the efficiency of their respective systems. Democracies create a selective environment in which myriad political entrepreneurs compete, ideally, over the satisfaction of voter preferences.

Yet domesticated competition creates natural tensions: if innovative actors can exploit or capture the rules that govern the allocation of resources, they can reap extraordinary returns and undermine the group. Here we have assumed benign innovation, but from the next chapter we turn to its ambiguous reality.

# 7

# The Red Queen's race over rules: The contest between cooperators and defectors

Idealised market societies and idealised organisms have a few basic properties in common. The payoffs of all strategies, both existing and newly emerging, lie on the optimality line. Innovation is unambiguously favourable for the group, indeed it is Awrey's (2012, p. 258) "unforecastable improvement". Competition within the group is of the domesticated variety only (O-competition), while all antagonistic or parasitic forms of competition (A-competition) are suppressed. There is a long history of theorists likening human societies to organisms—with early forms in Plato and Aristotle (see Walter, 1960), Comte (1855), Durkheim (1895/2014), Spencer (1875), Virchow (1859, 124), Haeckel (see Reynolds, 2008), Hayek (1933), and DS Wilson (2003) in the case of religious groups. The validity of this metaphor has long been unclear.

One fruit of the theory developed so far is some clarity on this old problem: the metaphor is profoundly misleading. The key peculiarity of the organism is that it has achieved *individuality* (Box 4.2), meaning that lower-order (gene) interests can be taken as identical with the group (genomic) interest. This is a consequence of the mode of payoff transformation: powerful commitments in the form of structural payoff dependence ($r_s = 1$). It is the lockstep and non-adversarial nature of this mode of payoff transformation that means innovation will be overwhelmingly benign, such that there is usually no internal conflict at all.[95] To make this distinction is not trivial. Hayek's (1933) picture of the market society is of "an organism and not an organisation",[96] and indeed he supposes that it is

> *an organism in which every part performs a necessary function for the continuance of the whole without any human mind having devised it*. (p. 130)

This is consistent with, indeed essential to, Hayek's belief in benign innovation and his preference for the minimal use of rules to govern markets. Yet while the

[95] With exceptions that prove the rule, which are discussed in A.7.1.
[96] Hayek's views are discussed further in Chapter 13.

organismic *ambition* of the ideal market is reasonable,[97] realising this ambition in the absence of $r_s = 1$ requires using the only alternative means of payoff transformation available: *rules*. This, we will see, entails overcoming some extraordinary challenges.

In any case, human societies are clearly not individuals, but are *conflicted groups*. They are meaningful *groups* because there are returns to cooperation, but are *conflicted* because there are private returns to defection. With levels of $r$ low or zero (as in the self-interested model of society), cooperation is a consequence of rules.[98] But rules are *adversarial*—they are strategies vulnerable to counterstrategies—and this brings us to a new Red Queen's race.

Thus, we move on from the single-level Red Queen's race of Chapter 6 to a two-level race: the fight for survival occurs both without and within.

Section 1 turns to the *within-group* Red Queen's innovation race, a race between cooperators and defectors. In unlucky periods, the bias in strategic change is unfavourable: the techniques of cheating and defection outpace those favouring cooperation. Instead of Hayek's "spontaneous order", we get "spontaneous disorder". Uncooperative groups are less competitive in the between-group conflicts of Chapter 6, so if cooperators play the sleeping Hare—if they fail to keep up in the internal Red Queen's race—then group collapse is certain.

Section 2 turns to a puzzle this raises: in the race between cooperators and defectors, why would *individuals* innovate and invest in rules that enhance *group* cooperativity? I distinguish two kinds of rules:

- rules-as-private-goods; and
- rules-as-public-goods.

I observe that actors have incentives to innovate in the improvement of rules-as-private-goods, but such privately-rewarding rules can only sustain cooperation in small-scale groups (i.e. below 20 members).

[97] I.e. the ambition to align individual with group returns, so individuals behave as Hayek supposes.
[98] Excepting the case of the dictator, whose value of $r_s$ is roughly set by the rate of output taxation.

Cooperation in complex social orders depends on rules that are public goods. This includes institutions. To run the internal Red Queen's race on the side of cooperation in large social orders *is to provide public goods*. Chapter 8 turns to this final problem, of why individuals supply such public goods, and completes the general theory of social order.

### 1.1 Innovation in conflicted systems: The internal Red Queen's race

Lifting the assumption of benign innovation returns us to the fundamental social problem: strategies are distributed across the four quadrants of the social strategy space. The process of innovation adds new ones over time, with a stochastic cooperator-defector bias in strategic change.

If we assume r is fixed, at some low or even zero value (and r=0 in the market society of self-interested actors, excepting the dictator), then cooperation must be raised by rules. Some individuals employ rules to suppress the use of parasitic strategies by others, and these rules, of course, are strategies. This brings us to the Red Queen dynamics introduced in Chapter 6: there is an adversarial innovation contest, in this case occurring within the group, and it is a race between cooperators and defectors.

Take a classic Public Goods game. Initially, the standard result holds: some players cooperate, but defectors are more successful and cooperators disappear or give up. The public good is wholly neglected. But then a novel strategy for punishing defectors emerges, or becomes cheap enough, so that defection no longer pays. The public good is fully provided. Next emerges a strategy for evading detection, nullifying the punishment, or for cheap counter-punishment—and cooperation again collapses. Perhaps cooperators then devise a strategy for co-locating, and find ways to reduce public good spillovers. Or they discover a mechanism for providing the public good only in the case that a critical proportion of group members decides to contribute. Or they find a way to ensure that individuals must co-produce the public good whenever they produce a private good that they cannot live without. And on it goes.

What is the equilibrium level of public good production? There is no long-run equilibrium—only a level that oscillates with the stochastic bias in strategic change. Random changes in the relative effectiveness of cooperator and defector strategies produce "evolutionary cycles of cooperation and defection" (Imhof et al., 2005, p.

10097), and oscillations in group productivity.[99] Sometimes we get Hayek's "spontaneous order" (e.g. with the fortunate innovations of Figure 5.1.ii), but equally we may get "spontaneous disorder" (e.g. those of Figure 5.1.i).

If the group depends on public good production for survival, then a sufficiently unlucky period of strategic change will produce extinction. In this unhappy case, extinction is the long-run equilibrium. For human civilisation, certain public goods have this characteristic; non-zero rates of global nuclear war or severe environmental catastrophe, for example, may guarantee failure in the long-run.

The within-group Red Queen's race has same general properties emphasised in Chapter 6:

- The stochastic bias in strategic change means that the gap between cooperator and defector competitiveness varies over time. Where the gap reaches a critical threshold, cooperator or defector types are wholly eliminated.
- The complexity of the design space means that Orgel's rule applies, and we cannot rule by fiat that a given strategy is unbeatable.
- The outcome of the race is strongly influenced by relative innovation capacities, producing predictable Tortoise and Hare patterns where one racer fails to invest in innovation sufficiently (see A.7.1 for a biological example).

### 1.2 A two-level race: Cooperate to compete

In all real-world orders, Red Queen's races occur at multiple levels. They occur, for example, between individuals within teams, between teams within firms, between firms in different sectors, between sectors in economies, and between economies. Performance at one level affects that at others.

In the simple two-level model I explore here, the *within-group* Red Queen's race has consequences for the *between-group* Red Queen's race: oscillations in internal cooperativity affect the group's capacity to keep up with external competitors.

[99] In sociobiology, tree-mycorrhizal fungus interactions are an interesting example. Against the popular picture of pure cooperation, they entail regular oscillation between parasitism and mutualism (e.g. Johnson et al., 1997).

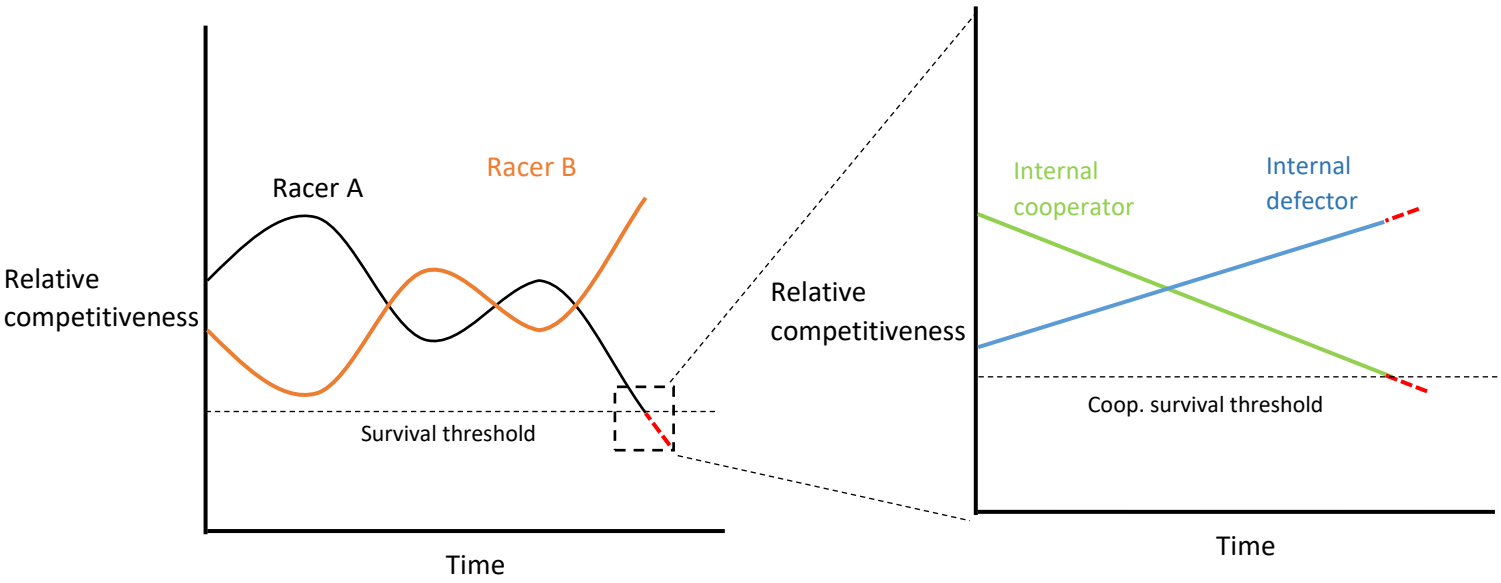



**Figure 7.1** The competitiveness of groups is shaped by the internal performance of cooperators versus cheaters. In the case of Group A, internal cheater innovations lead it to lose the higher-order race against Group B.

Figure 7.1 is a modification of Figure 6.2 in the previous chapter. Now the internal balance between cheaters and cooperators, on the right side of the graph, determines whether a group can keep up with its external competitors, on the left. When internal cooperativity declines, the group becomes less competitive. If competitiveness of the overall system falls below critical thresholds—as for group A—this leads to collapse, displacement, or frequently, in the case of societies, the imitation of, or absorption by, more successful others. Thus, keeping up in *between-group* Red Queen's races requires keeping up in the *within-group* Red Queen's race over cooperation.

Given this thesis focuses on capitalist societies, it useful to return briefly to the Darwin machines of the previous chapter. Darwin machines are the product of rules. In the absence of high levels of commitment, the use of rules creates a Red Queen's race: defectors may benefit by evading rules, and cooperators aim to benefit by repairing them.

There are large potential returns to defection within systems of domesticated competition. The more cooperative the system, the greater the social surplus. To access these rivers of resources, domesticated competitors must satisfy onerous conditions: they must search intensively for novel ways to increase group welfare. Defectors, meanwhile, may achieve extraordinary returns by finding weaknesses in

the resource-allocating rules, and so bypassing these conditions. Such patterns are found in all Darwin machines, anthropic and sociobiological (we of course focus on the former, but see A.7.2 on the latter).

To be more concrete, in market economies, individuals and firms may reap large rents by finding novel ways to steal, evade tax, evade competition, turn private financial risk into public financial risk, socialise the costs of production, influence media actors and political entrepreneurs to capture the rules of the game, and much beyond. Economic and political entrepreneurs may collaborate to prevent the Darwin machine from being repaired when it breaks, to defend existing opportunities for rent-seeking. Or they may prevent it from being properly set up in the first place. The most dangerous kinds of parasitism involve positive feedback loops (again important in biology, see A.7.2). For example, market actors' extractive strategies allow them to accumulate wealth and structural power, these resources may be converted into political influence, and political influence can be used to defend and extend opportunities for extraction. Such concerns in political economy go back to Adam Smith, who hoped to overturn the heavy rents of mercantilism—and they will reappear in every age.

### 2.1 Motivating cooperators: Keen racer or sleeping Hare?

To reiterate Chapter 6, it is true that the *ideal* system of rules solves the cooperation problem. As Hayek (1985, 16:47) remarks of the market,

> *the only symbol that tells us where we can make the best contribution is profit, and in fact by pursuing profit we are as altruistic as we can possibly be.*

Again, this is the description of the market society that has permanently achieved the organismic ambition, where payoffs are ideally transformed and all innovation is benign. To describe and redescribe this ideal, however eloquently, does not itself contribute to winning the adversarial race.

A key puzzle raised by Section 1 is why some individuals race on the side of cooperation. Why is there a racer, instead of a sleeping Hare? The motives of defectors are easily understood: evading rules allows the execution of privately-

rewarding parasitic strategies that would otherwise be blocked.[100] In the case of cooperators, it is useful to point to two kinds of rules:

- *Rules-as-private-goods*. These are cooperative rules where the "raw" payoffs are in the mutualism quadrant ($B_i > 0$). Individuals using rules that support cooperation reap a direct private benefit that justifies their investment. The increase in group performance is a happy and unintended side effect.
- *Rules-as-public-goods*. Here the "raw" payoffs for rules are in the public goods quadrant, such that they are privately costly ($B_i < 0$). If individuals are to originate and repair rules-as-public-goods, they must be motivated by some mechanism of payoff transformation—either more rules, or commitment.

### 2.2 Rules-as-private-goods: Micro-scale cooperation

Rules-as-private-goods provide the individual with a direct private payoff, such that we do not need to invoke some transformation mechanism to explain why individuals might invest in them. This category includes, for example, Axelrod's Tit-for-Tat strategy as an optimal response to the standard iterated Prisoner's Dilemma (PD) (see Section 2.4, Chapter 4).[101] Although individuals adopt Tit-for-Tat purely for its private benefits, it has the happy side-effect of maximising group productivity.

Rules-as-private-goods are almost always individual-level rules, rather than institutions. There are three kinds of individual-level reward and punishment strategies emphasised in the literature:

- *Direct reciprocity*, the punishing or rewarding of interaction partners based on past gameplay. Tit-for-Tat is the prime example. It requires iterated gameplay, and breaks down in one-shot games and in the last round of iterated play.
- *Indirect reciprocity,* an informational strategy that involves learning from a partner's gameplay with other individuals. It requires some mechanism for tracking reputation, such as direct observation or gossip.

[100] Including receiving rewards, e.g. subsidies, without paying the cost of providing associated public goods.
[101] Or better still, the strategy of pursuing Tit-for-Tat *unless* one's partner is an unconditional cooperator, in which case defection is most rewarding.

- *Third-party punishment*, where a third-party individual punishes those who cheat in interactions with other members of the group. That is, they police cheaters even when they are not themselves the victim.

Due to space limitations, detailed discussion is relegated to A.7.3.

The decisive problem in each case is scalability. The literature finds that these rules only function in "small" groups, and in this literature groups of *just ten or more individuals* are considered "very large". This alone is enough to ensure that, as Henrich (2004, p. 2) writes, individual-level punishment mechanisms "fail to explain the prosocial preferences that promote large-scale cooperation".

A second limitation is the nature of the Red Queen's race. Henrich (2004) observes, taking indirect reciprocity as an example,

> *if a… defecting-strategy can generate an inflated reputational signal, perhaps by paying other individuals to lie about the mutant's cooperative tendencies, then the predictive value of reputation will corrode and cooperation will collapse. The only way to prevent the entry of such strategies into the mix is to restrict them by fiat.* (p. 9)

To *reliably* produce cooperation within small groups one must assume that these rules are unbeatable, and this is no more justifiable than taking defector strategies as unbeatable. Given the stochastic cooperator-defector bias in strategic change, levels of cooperativity will rise and fall in oscillating cycles.

To conclude, rules-as-private-goods can only support *oscillating* and *micro-scale* cooperativity.

### 2.3 Rules-as-public-goods: Complex, large-scale cooperation

Large-scale cooperation is sustained by rules-as-public-goods, where the returns to cooperation mostly accrue to others. In large-scale orders, when and why do individuals race on the side of cooperation, rather than defect or play the sleeping Hare?

Innovating in the rules that produce domesticated competition in the market, for example, generally provides no direct private returns. Take property rights. An individual can only gain by paying for enforcement that is *particular* to their case.

The benefit is larger the more *partial* that enforcement is to their interests. This can turn policing into a private good, but in doing so it will tend to make it parasitic rather than a public good: ownership of property is determined by who can muster the greatest force. This pattern is common where institutions have broken down, and in history prior to the provision of genuine property rights. Socially efficient property rights must approach, however imperfectly, the ideals of *universality* and *impartiality*. No individual can benefit by making lone contributions to the funding pool for socially efficient property rights (except for the dictator, because $r_s > 0$); all are better off defecting and free-riding.

To reiterate and summarise, we need individuals to run the internal Red Queen's race on the side of cooperators against defectors, and in the large-scale social order, this means contributing to the supply of rules-as-public-goods. Given the framework developed so far, there is a straightforward answer to how this must be done: their payoffs must be transformed. And we are now familiar with the two available mechanisms:

1. $r > 0$, so the individual has some commitment to group performance; and
2. adding an additional layer of rules that impose rewards or punishments (i.e. modify $c$), to induce even mercenary actors to contribute.

In the next and final chapter of Part 2, I investigate the use of these mechanisms in detail.

# 8

# An impossibility theorem for rules-based order: Rules and commitment as essential complements

Now we arrive at one of the core theoretical arguments of this thesis: all complex social orders in our universe must be constructed from a particular combination of commitment and rules. The two are essential complements. Later this argument will be applied to the cases of authoritarian and democratic societies. The long-term performance and survival of democracy—and escape from the current period of instability in the twenty-first century—requires understanding its unique and complex combination of these two fundamental forces for cooperation.

I start with the key question raised at the end of the previous chapter: If group cooperativity depends on individuals supplying rules-as-public-goods, and so running the Red Queen's race against defectors, what motivates these cooperators?

Section 1 analyses the possibility of the purely rules-based order (i.e. $r = 0$). If individuals are to innovate and invest in rules-as-public-goods, they must be motivated by a higher-order set of $c$-modifying rules. This results in a problem of infinite regress: every layer of rules we add requires another. The purely rules-based order requires a *universal repair mechanism*—some uber-rule that somehow says, generically, "reform the rules in the interests of the group". It must universally allocate Pigouvian rewards in proportion to the group benefit, and must be unbeatable and so outside of the strategic game. This is, in short, an impossibility theorem for the purely rules-based order.

Commitment, however, *has precisely the properties of a universal repair mechanism*. Lockstep payoff transformation produces universality,[102] and commitment is non-adversarial and external to the strategic game. Commitment is, in the world of the Red Queen, the only solid ground. As a result, all complex social orders rest on commitment. Sociobiology offers a useful, hard test of the logic: all complex sociobiological orders rest on a commitment mechanism.

[102] As noted in Chapter 4, it is akin to a universal Pigouvian mechanism.

Commitment "anchors" the degree of cooperativity in rules provision. If $r = 0$, then no rules-as-public-goods survive the innovation race. If $r = 1$, all rules-as-public-goods are provided. If $r = 0.5$, then rules-as-public-goods are provided wherever group benefits are at least twice the private cost.

Section 2 analyses the possibility of purely commitment-based orders. Such orders are also impossible. There are two reasons. First, communities of committed actors who do not employ rules are swiftly overrun by defectors. Second, rules are needed for optimisation via domesticated competition. If a society of perfectly committed saints ($r_c = 1$) was so irrational as to reject rules, it would be forever destitute.

To conclude, commitment and rules are essential complements. This is illustrated in the closing Figures 8.1 and 8.2, which show the feedback between these two fundamental causes of cooperation.

The "general theory of social order" states that:

- commitment is the universal repair mechanism that anchors rules;
- rules defend committed actors from defectors; and
- rules support optimisation via domesticated competition.

### 1.1 Rules without commitment: An impossibility theorem

In the purely rules-based order, individuals will provide rules-as-public-goods only if a higher-order layer of rules rewards them for doing so. But we immediately hit a logical roadblock: this higher-order layer of rules is also a public good. Thus, the same problem is simply pushed upward to a higher order of rules. Every layer we add requires yet another, leading to a problem of *infinite regress*: "a series of appropriately related elements with a first member but no last member, where each element leads to or generates the next" (Cameron, 2022, para. 1). As William James had it, it must be "turtles all the way down" (see Ross, 1967, iv). This problem is fatal, and it will be helpful to develop the argument further to show why.

Let us begin with the lesser problem of rule *origination* and its proposed solutions—that is, the problem of how mercenary actors may begin providing rules-as-public-goods, from a starting point where there are no other rules to motivate them. For example, how can a society arrive at a set of rules for an open market and political

system from a condition where they are closed and dominated by economic and political rentiers? The typical origination conditions proposed—a benign dictator, a balance of power, or Coasian bargaining—are discussed in Box 8.1. The first point is that none of these origination conditions involves building up from rules; each involves supposing a *temporary period of commitment* that allows rules to be formed.

The second and more decisive problem is that of *maintenance* after the conditions of origination fade. However efficient the system of rules originated, the Red Queen's race continues, and there must be some way for the rules to adapt as strategic change proceeds—as new externalities, extractive strategies, and public goods appear. Thus, the structure of rules must include a "repair layer": strategies for detecting when the rules become maladapted, detecting the quality of mercenary actors' efforts to restore the system, and rewarding or penalising them in a Pigouvian manner that reflects the importance of the contribution.

It may be clear by now that rules are poorly suited to this task. The limits to rules, discussed in Chapter 4, are fatal:

1. First, a completeness problem. Say that every actor for some reason follows the letter of the law. Nonetheless, with imperfect foreknowledge, the system of rules will be compromised in unanticipated ways. Novel public goods are required that can restore the original equilibrium from any possible corrupted equilibrium, and it is impossible to codify the supply of all these unknown public goods in advance. Even will perfect foreknowledge, it is implausible that a sufficiently flexible and unambiguous system of rules, suited to every eventuality, could be written down. Over time, inaction or imperfect repairs, following the letter of the law rather than the public good, produces a progressively expanding mismatch between rules and the strategies that actors pursue.
2. Second, the repair layer itself will be compromised by defector innovations, and a compromised repair mechanism cannot repair itself. If individuals are to repair the repair mechanism, we need yet another layer of rules that can motivate that repair—and then another, and so on. This is the aforementioned problem of infinite regress.

**Box 8.1 The standard origination arguments depend on commitment: Temporary $r > 0$**

One solution for originating group-favourable rules-as-public-goods is a period of benevolent dictatorship. This solves the public goods problem by recourse to $r > 0$: dictatorial taxation gives us $r_s > 0$, and dictatorial benevolence bolsters this further with $r_c > 0$. When the dictator leaves, this source of commitment disappears.

Another solution is a balance of power. If competing powers are uncertain about whether they will win or lose in future rounds of conflict, and if they are risk-averse, they may prefer to agree to rules that formalise and stabilise that balance of power. The Magna Carta is a classic example: the balance between King John and the rebel barons birthed a system of rules designed to constrain power and limit conflict. To place them within the framework:

- This is another case of $r_s > 0$. The balance of power only produces public goods for society if the competing powers are *stationary* bandits in the Olsonian sense (like King John and his barons). A balance of power between *roving* bandits is of dubious value; it is likely to make the marauders more dangerous.
- These stationary bandits are employing a rule like Tit-for-Tat among themselves. Because competition erodes $r_s$ (see Chapter 4, Section 3.5), if rules can support cooperation this effectively raises $r_s$. Multiple "kings" are able to avoid losing resources in conflict. But Hobbes, of course, would point out that conflict is lowest, and $r_s$ highest, with a single Leviathan.

Note that to get balances of power between *large* groups, we need to explain collective action—which reintroduces a public goods problem.

Coasian bargaining may appear to be a third approach. Coase's (1960) insight is that with frictionless bargaining and a third-party enforcer, it is possible to arrive at any condition that is a Pareto improvement. Coase himself regarded these assumptions as rarely applicable. In any case, it is the wrong tool for the job: it *assumes* an ideal third-party enforcer (which, to be ideal, must be motivated by $r = 1$).[103] Coase's theorem makes a different proposition: that one public good (minimal property rights and contract enforcement) might sometimes be a substitute for another public good (more comprehensive systems of property rights).

[103] The permanently ideal third-party enforcer requires (A) completeness and (B) immunity to innovation. As argued in this chapter, this means this enforcer must be motivated by $r = 1$.

The key problem for *any* origination argument is the Red Queen's race. The balance of power is a temporary condition: the stochastic bias in strategic change (the bias in group competitiveness of Chapter 6, Section 1.1) has, for a moment, led neither party to perceive an advantage. But strategic change continues. The Magna Carta was breached almost immediately, as its parties saw opportunities to upset the balance. The benevolent dictator, should such a creature ever be spotted in the wild, will cease ruling sooner or later. When the origination conditions are inevitably overturned, the system of rules will need to stand by itself against the winds of change—and my argument here is that this is impossible.

The nub is this: Large-scale cooperative systems need individuals to innovate beyond the rules, and mercenary actors will not do so. Because cooperation in the world of the Red Queen requires a perpetual stream of novel public goods, the "origination" and "maintenance" problems converge in the long-run.

What such systems need is a *universal repair mechanism*—a rule that says "reform the rules in the interests of the group", that allocates private payoffs in proportion to the gains of the group, and that is unbeatable or innovation-proof. This is impossible. Such a mechanism would be absurdly complex, require vast amounts of future knowledge, and its complexity would render it vulnerable to an enormous range of possible defector attacks.[104]

### 1.2 Commitment as a universal repair mechanism

There is such a universal repair mechanism: commitment. It solves the two problems affecting rules because:

1. the lockstep nature of transformation by commitment makes it complete. If $r > 0$, then private payoffs are proportional to group payoffs, for existing strategies and for unanticipated novelties. No codification or foreknowledge is required, and individuals will act beyond rules-based incentives to maintain the favourable group equilibrium.

[104] E.g. a rule that says every individual should receive a compensating payment for any provided public good. This involves an extraordinarily complex, indeed impossible, detection task; it would necessarily require vast regulatory discretion, which is the ripest of opportunities for rent extraction.

2. commitment is not strategic, but is rather a preference or structural property of the world, it solves the problem of infinite regress. Commitment is outside of the strategic game and innovation-proof.[105]

This framework thus gives us an *impossibility theorem for purely rule-based orders*. All such orders require either structural or motivational dependence, i.e. an exogenous commitment to originating and maintaining that order. The case of motivational dependence, at least, is related to Arrow's (1972) speculation that morality and the price system may be "essential complements" (pp. 346). That of structural dependence is related to Olson's (1993) account of the dictatorial order.

The impossibility theorem applies to social orders in cognate fields, which share the fundamental social problem and open-ended innovation (Box 8.2). Indeed, my research was partly inspired by the richness of rules, and yet the absence of *purely* rule-based cooperation, in complex sociobiological systems.

### 1.3 Red Queen equilibria: Commitment "anchors" rules

The complex cooperative "equilibria" we see in the world around us—in families, societies, and organisms and their various mutualisms—and the systems of rules they employ, all rest on commitment. Commitment *anchors* the rule structure, because it determines individuals' willingness to provide rules-as-public-goods in the absence of supporting rules. This follows the logic of Equation 4.2, $rb > c$, where positive values of $r$ are required to induce individuals to provide public goods.

The two edge cases are straightforward:

1. If $r = 0$ for all actors, then no rules-as-public-goods will be provided.
2. If $r = 1$ for all actors, all rules-as-public-goods will be provided (why these systems still need rules is a matter discussed shortly).

The first case will be significant later: in political economy, it is a case where actors are self-interested and have no payoff-dependence (i.e. have the incentives of Olson's roving bandits), so that both $r_c = 0$ and $r_s = 0$. The consequence is *anarchy*.

[105] In later chapters, and in relation to democratic capitalism, we will see how $r$ can change values—but it does not do so as a consequence of individual innovation.

**Box 8.2 Commitment grounds sociobiological (and artificial) orders**

Sociobiology is of particular interest, because it falls within this general theory and offers an immense universe of cases. It is a powerful test of the logic of the argument, even if it cannot tell us whether this logic applies to human societies. Over four billion years, evolutionary processes have uncovered extraordinarily complex systems of rules, including those that use domesticated competition to optimise systems such as the human mind. And yet, in nature, *no complex, large-scale cooperative system is rules-based*. All require $r \gg 0$. Certainly no Darwin machines are observed in systems without high degrees of commitment. Because $r = 1$ systems most successfully capture the returns to cooperation, they are also the most competitive—so are everywhere the dominant form of life. See A.8.1 for a discussion of the impossibility of the rules-based, uncommitted "organism".

Similar principles can also be applied to the problem of controlling artificial general intelligences (AGIs). All existing strategies aim to control such agents via rules, by programming utility functions and various constraints. This places us in a Red Queen's race against an innovative intelligence. If it has vastly greater innovation capacity than we do, then we are more or less guaranteed to lose the Red Queen's race. See A.8.2 for a more detailed discussion, including many empirical examples of how existing (and primitive) Ais are already winning innovation races against their programmers. The framework indicates that *rules-based control of AGIs is impossible*. If there is any solution, it lies in bestowing AGIs with innovation-proof commitments (see A.8.2).

The second case is useful for understanding patterns of cooperation in systems such as idealised organisms ($r_s = 1$). For human societies, universal $r = 1$ is, of course, implausible.

The most interesting cases fall into a third category:

3. If actors within the system are shaped by variable levels of $r$, we have a race between actors with different degrees of commitment to cooperation. In a simplified case where some actors are defined by $r = 1$ and others by $r = 0$, we get a race between pure cooperators and pure defectors. With intermediate

values of $r$, we have actors who will cooperate in providing some public goods and defect for others.[106]

In these cases, equilibria are dynamic. Outcomes in any given period depend on the pattern of commitments among agents, whether strategic change is biased to favour cooperators or defectors, and agents' relative innovation capacities. All these factors will be important when we turn to the workings of authoritarian and especially democratic capitalist societies.

## 2 Commitment without rules: Instability and inefficiency

Rules are also an essential complement to commitment. It may initially seem contradictory that, for example, an $r_s = 1$ organism would employ a complex array of $c$-modifying rules, given the job of motivating cooperation appears to be solved by commitment. In practice, however, the use of rules-as-public-goods is often *deepened* as $r$ increases. On one level this is obvious: as just noted, $rb > c$ tells us that public goods will be more extensively provided the higher the value of $r$. The question is why $b > 0$, i.e. why are rules still group-beneficial in committed groups?

### 2.1 Defending committed actors

The first benefit arises if we allow defectors to invade, as is always possible in real-world systems. The cooperative actor who refuses to use rules (i.e. detection and enforcement strategies) is an *unconditional* cooperator, and a well-established result in the literature is that unconditional cooperators are easy prey for defectors. The defector consumes public goods without paying for supply, and devises extractive strategies that unconditional cooperators cannot resist. In biological systems, defectors reproduce rapidly, cooperators die out, and group productivity plummets. In anthropic systems, roving bandits may steal everything—not only the social surplus, but also the groups' means of subsistence—with the same outcome.

Commitment anchors rules. In turn, committed actors use rules to preserve this anchor, by removing the defector advantage. We will see in Chapter 15 that this logic has profoundly shaped the evolution of human morality. It is readily observed in

[106] Following the logic of $rb > c$ from Chapter 4. E.g. at $r = 0.5$, actors will cooperate if public goods give at least twice the group benefit to the private cost, but defect for lower ratios.

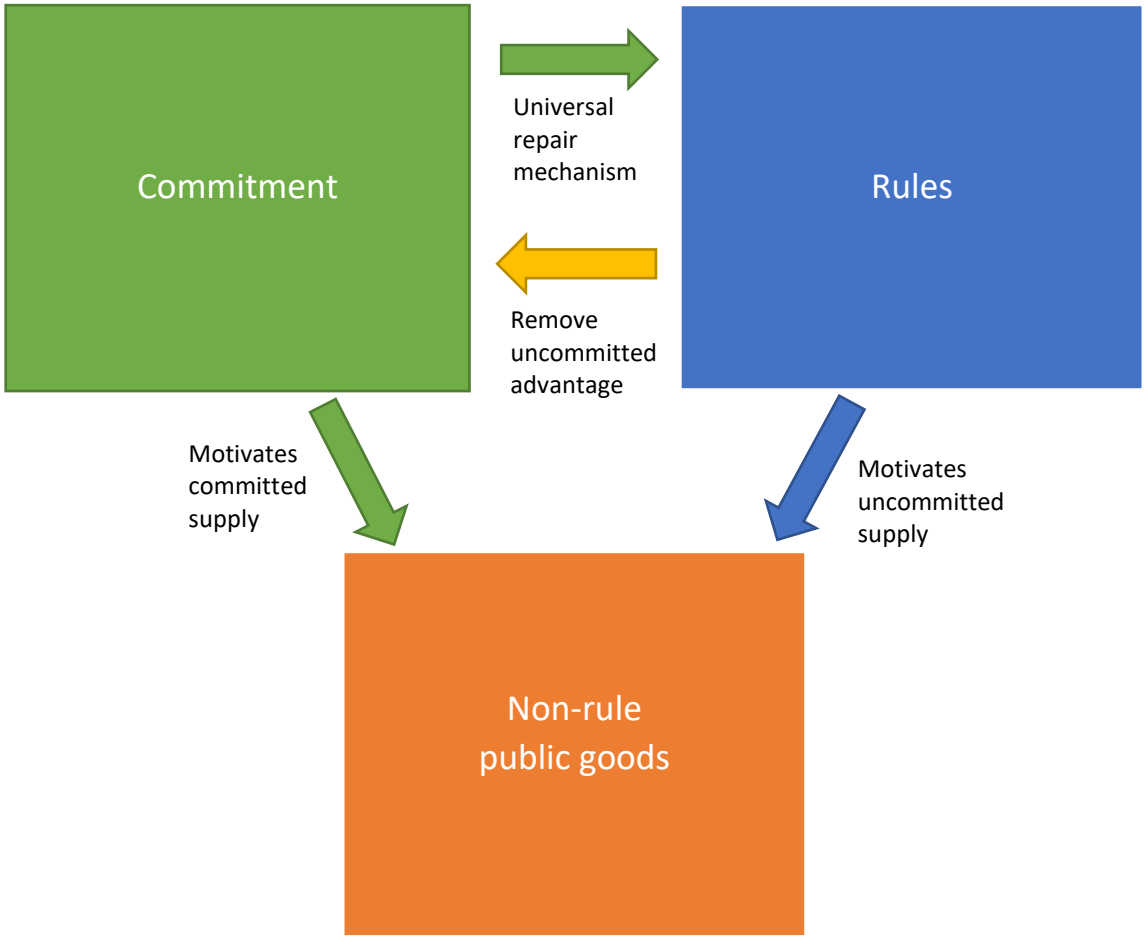



**Figure 8.1** Reproduction of Figure 1.1. Commitment is an exogenous motive for providing rules, given the impossibility of a self-repairing rule-based order. In turn, rules support the maintenance of commitment by penalising defectors. Both rules and commitment motivate the provision of various other economic and political public goods.

biological systems, where rules act as purging mechanisms for removing troublemakers as they appear (Travisano & Velicer, 2004; Strassmann & Queller, 2011), and those groups that are best at purging defectors are fitter—reproducing more and increasing their representation in the next generation.

## 2.2 Optimising via domesticated competition

Rules have a second function that makes them valuable even in a world without invading defectors: optimisation. As discussed in Chapter 6, Section 2, systems of domesticated competition—Darwin machines—are formed from rules that govern within-group resource allocation. Their purpose is purely to harness distributed

knowledge and accelerate the development of group-favourable innovations (Figure 8.2).

Suppose, for example, that every individual is an ideal altruist, transformed by $r_c = 1$, and that defectors never invade. These altruists still need to allocate resources efficiently. If they are sufficiently rational and economically savvy, then for the reasons discussed in Chapter 6 they will regard prices and market competition as an essential contrivance for doing so. So long as they believe market rules are well-tuned, such that profitability adequately reflects social welfare, these altruists will assiduously follow the price signal.

In the ideal market, altruists and self-interested actors become indistinguishable. Yet the distinction is, in practice, of decisive importance. Market rules are public goods, and are subject to defector counter-innovations. The advantage of the perfectly altruistic market society is like that of the typical organism: there is no internal Red Queen's race over the capture or evasion of rules, and no problem of a stochastic cooperator-defector bias in strategic change. There are only perfectly motivated attempts to improve and adapt the rules of domesticated competition over time.

Commitment anchors rules, including systems of domesticated competition. The more committed the group, the more perfect its O-competition. The less committed, and the closer $r$ slides towards zero, the more A-competition takes over. This is highly suggestive about the nature of market societies, though the argument will not be fully developed until the end of Part 4.

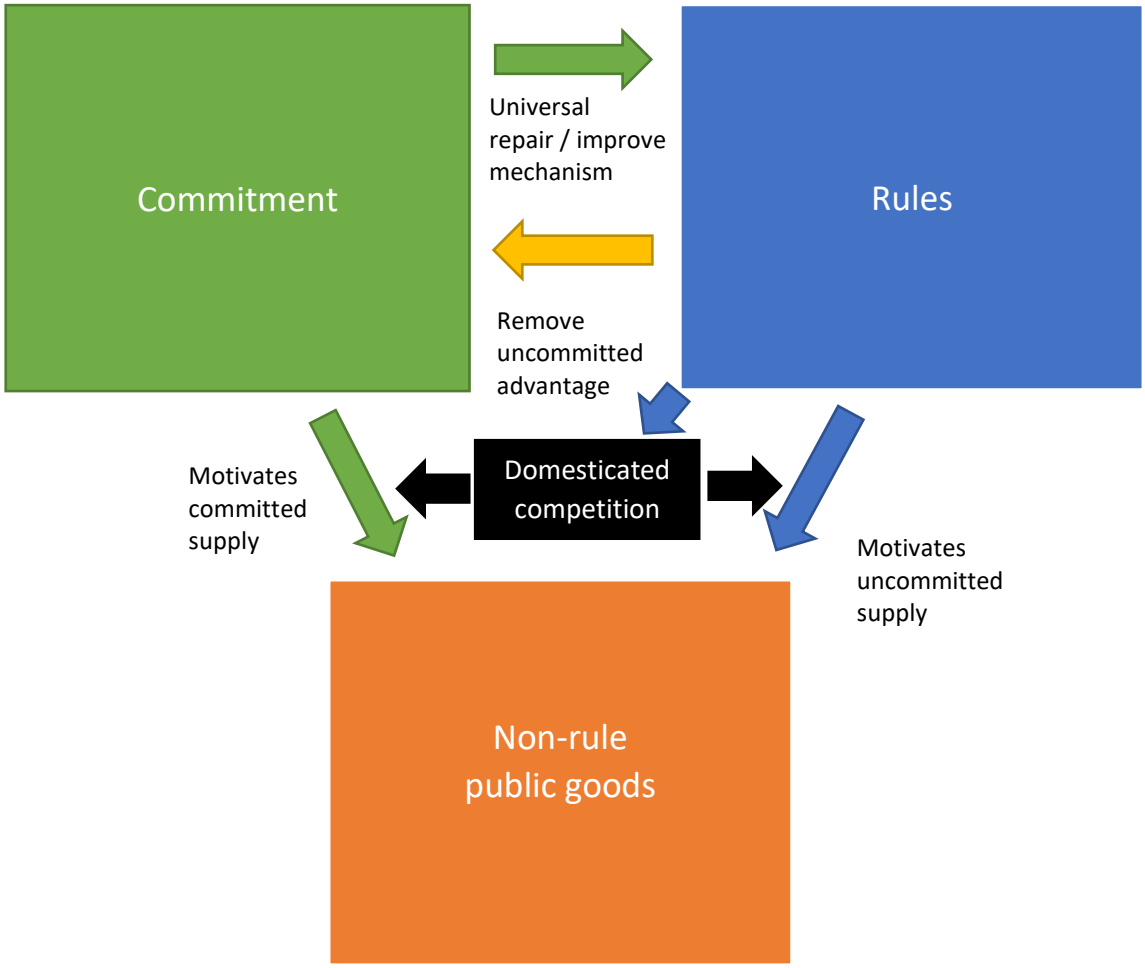



**Figure 8.2** To facilitate optimisation, committed actors may devise systems of rules that produce Darwin machines, loci of domesticated competition in which resources are allocated to individuals who best serve the group. These rules are public goods and may be exploited, with competition returning to its "feral" form; the dynamics presented in Figure 8.1 are a precondition for domesticated competition.

*…the law has everywhere a tendency to lag behind the facts of life.*
Supreme Court Justice Brandeis (1915, p. 464)

# 9

# Interlude:
# Introducing the institutional "Red Queen's relay"

With the abstract framework complete, the next step is to apply it to the cases of democratic and authoritarian market societies. By the end of the thesis, we will have a picture of how the different modes of payoff transformation, and the different innovation capacities of various actors, can be combined into a stable and prosperous system—or how they can come apart.

This chapter serves as an interlude, a bridge between the framework and practice, that sets up the analyses of Parts 3 and 4. In human societies, *institutions* play a critical role. I described them only fleetingly in Parts 1 and 2, to avoid over-burdening the broad analytic framework. This chapter follows up by defining institutions in more detail, and introducing the three innovation races that institutions create.  It closes by summarising the overall analytic approach.

Section 1 begins by defining the term "institution". Following Chapter 4, they are a variety of rule—a means of generating cooperation by modifying private payoffs ($c$). Property rights, for example, modify private payoffs so that individuals cannot benefit from theft. The hierarchical nature of institutions allows them to solve some of the problems of individual-level rules discussed in Chapter 7, and so maintain larger and more complex social orders. They are, at least when efficient, almost always public goods.

Section 2 introduces the institutional Red Queen's race. Institutions are composed of detection and enforcement strategies, and they invite counterstrategies. The framework predicts the usual cycles of adaptation and counter-adaptation. I give the problem some further structure by distinguishing three Red Queen's races in the market society:

1. the race between market actors;
2. the race between market actors and the regulator; and
3. the race between all actors for control of the regulator.

Together, these races form what I call the *Red Queen's relay*. The prosperity and competitiveness of any market society, whether democratic or authoritarian, rests on setting a good pace in each of the three legs of the relay. In essence the problem is how to generate a regulator that is both *competent* and *well-motivated* (i.e. solves problems of "pace" and "direction"), and so effectively runs the institutional Red Queen's race in the interests of the group. The three legs of the relay are explored, in turn, in Chapters 11, 12, and 14.

Section 3 sketches out the analysis of the Red Queen's relay over the remainder of the thesis. In Part 3 of this thesis, we focus on the first two legs of the relay, both of which relate purely to the matter of regulatory competence. It explains what is involved in effective economic management in the world of the Red Queen, and the recent underperformance of liberal market economies. Part 4 focuses on the third leg, the race for control of the regulator. It is concerned with the forces that make authoritarian and particularly democratic social orders possible, and what this tells us about recent errors in the theorisation and management of democratic capitalism.

**Box 9.1 "Institutions" in sociobiology**

Given we have established the framework and now move into the discussion of institutions, sociobiology will from here play almost no role. Yet there are uncanny parallels between the "social technologies" that stabilise cooperation in human and biological systems. We may be the only form of life to consciously promulgate rules of social behaviour, to argue over them in old stone buildings, and to inscribe them on sheets of cellulose for later reference, yet even the simplest of cellular lifeforms—bacteria—voting on whether to provide public goods ("quorum-sensing", e.g. Travisano & Velicer, 2004, pp. 74-75), policing production of public goods and indeed of voting behaviour (e.g. Dandekar et al., 2012), forming closed communities of public goods-sharing cooperators (Hol et al., 2013), and enforcing exclusive relationships between transactors (Currie et al., 1999 and Clark et al., 2000)—and often multiple such mechanisms at the same time (Wang et al., 2015). Many of the mechanisms found in biological systems have parallels in the institutional order. Thus, analogies may occasionally remain useful.

### 1.1 Defining institutions

I begin by briefly clarifying the use of the term "institution". The nature of semantic debates is that one cannot "prove" that a term necessarily includes this or excludes that. Nor, in a complex world, can we eliminate the problem of edge cases that are difficult to categorise. The ambition of this section is to reach a definition that is sufficiently clear and precise that the term can play its necessary theoretical role in this thesis. The definition offered is useful if we wish to study cooperation in boundedly rational, conflicted systems, and I argue we must do so if we are to understand human societies. It is largely consistent with Hodgson's (2015) definition of institutions as "integrated systems of rules that structure social interactions" (p. 501), where, as North (1990, p. 2) emphasises, those rules "structure incentives".

First, institutions are a sub-type of *c*-modifying rule. They work by shifting private payoffs—for example, penalising property rights violations. Unlike individual-level rules, they comprise a *hierarchy* of detection and enforcement strategies. These strategies are still devised and implemented by individuals (see A.9.1 on how this "individualist" account is in fact compatible with structuralist accounts) and like other strategies, all institutional strategies can be located somewhere within the social strategy space according to their payoffs.[107] The payoffs for institutions, or at least *efficient* institutions, are usually in the public goods quadrant—they are costly to supply and their benefits diffuse. If individuals are to contribute to the institutional structure—if they are to "structure incentives"—then they must first have their own incentives structured. This transformation occurs via a series of higher and higher-order institutions, which seems to invite the Chapter 8 problem of infinite regress, so that commitment must enter the picture.

The incentives of actors in institutional hierarchies may be modified by threats, bribes (e.g. the payoffs of regulatory actors may be transformed by cash payments), and other informal strategies. If corruption and violence become systemic and predictable, they may be considered part of the institutional hierarchy. Institutional actors may also be motivated by the varieties of commitment ($r > 0$), although the

[107] They may be parasitic, mutualistic, public goods, or harmful for all parties.

assumption of self-interest limits this to structural payoff dependence ($r_s > 0$) mainly in the case of the dictator.[108]

Second and more specifically, institutional strategies represent an extended, hierarchical form of *third-party punishment*[109] (mentioned in Chapter 7, Section 2.2 and discussed further in A.7.3). Each actor's payoffs are structured by another level of the hierarchy. For example, say an individual decides not to steal from another individual, for fear of arrest by a third-party police officer. That officer is in turn motivated by incentives enforced by a supervisor, the supervisor's performance is monitored by the police chief, the chief answers to the Minister for Police, and so on. All are third parties, none being the victim of the original theft. The next sub-section examines what distinguishes institutional hierarchies from individual-level chains of third-party punishment.

Third, recall that the purpose of rules is (A) to adjust others' payoffs, and/or (B) to allow the use of distributed knowledge.[110] Thus, our definition of institutions excludes all public goods with *direct* payoffs for the group, such as educating a child, vaccinating a person, or constructing a highway. These are orthogonal to institutions. To see why, notice that the latter public goods would retain their direct payoffs in a world of *omniscient* and *wholly cooperative* altruists. Such individuals would:

- be ideally motivated to supply all public goods with direct payoffs; and
- know exactly what everyone else is doing, and how best to contribute.

They would still provide vaccinations, build highways, and so on, i.e. all public goods with direct payoffs. But they would have no use for institutions. Institutions are valuable only because they solve problems of "direction" or "pace" in the conflicted and boundedly rational society.

Fourth, institutional strategies must structure individuals' payoff *expectations*. A parent who imposes a curfew on their teenage child has introduced a household

[108] Recalling that $r_s$ is an extrinsic form of commitment that works via winning some share of the material payoffs of others.
[109] The rewards or subsidies promised by institutions are backed by a system of third-party punishment that would penalise non-delivery of those rewards (i.e. a punishment for failure to transfer).
[110] The use of knowledge being one function of markets, as Darwin machines. Thus, rules are useful even where all actors are already ideally motivated to cooperate, so function A is unnecessary.

institution. A parent who sporadically punishes or rewards random behaviours may change their child's payoffs, but they have not established an institution; there is no way for the child to form expectations and strategically respond. Similarly, a one-time bribe to a political official in an otherwise healthy democracy does not constitute an "institution". By contrast, systematic exchanges of favours between market and regulatory actors are institutional even if they are informal. Formal codification of institutions is not strictly required, but it becomes necessary as systems become complex—if the relevant actors cannot discover information about how institutions operate, their expectations cannot become structured.

### 1.2 Hierarchical versus individual-level third-party punishment

The differences between institutions and individual-level third-party punishment are of great significance, but space constraints mean the detail must be shifted to A.9.2.

In brief, institutional hierarchies introduce a complex *division of detection and enforcement labour*. This helps to:

- solve problems of second and higher-order free riding, i.e. defectors refusing to contribute to third-party punishment. Institutional hierarchies assign clear domains of responsibility, resolving informational problems that hamper individual-level third-party punishment, and helping to suppress perverse counter-punishment behaviours.
- make detection and enforcement strategies cheaper and more effective, for the same reason Smith (1776) noted that divisions of labour increase the efficiency of pin-making (p. 1). The result is a panoply of detection and enforcement specialists: forensic dentists, sniffer dog handlers, forensic accountants, all the varieties of human resources and managerial staff, and so on.

"Vertical" accountability (e.g. from police minister, through police chief, middle managers, patrol officers, and finally to citizens) can be distinguished from "horizontal" accountability (e.g. media and the judiciary disciplining the legislative and executive branches). Liberal societies tend to have strong and independent horizontal accountability structures, while authoritarian societies tend to absorb them into the vertical accountability structure.

Finally, in Part 3, we will see that the institutional division of labour does not replace, but is in fact complementary to, individual-level "rules". That is, the *total regulatory structure* is necessarily composed of institutional strategies working in combination with individual-level (Chapter 7-style) strategies employed by market actors.

### 2.1 The institutional Red Queen's race

Institutions are strategies subject to counterstrategies, and their use creates an *institutional Red Queen's race*. The now familiar set of principles apply: there is an uncertain cooperator-defector bias in strategic change, there are no unbeatable strategies, and relative innovation capacity is a key long-run determinant of performance. To play the sleeping Hare and withdraw from the project of institutional adaptation—for example, in the belief that some set of institutions is unbeatable and self-regulating—is to invite the progressive decay of cooperativity (Box 9.2).

To run the institutional Red Queen's race, we need:

- *motivated cooperators* who will run on the side of a socially efficient institutional order. For all institutional orders, we need to solve the problems discussed in Chapter 8: What motivates individuals to repair and improve institutions?
- *competent* racers, who are sufficiently rational to block novel parasitic strategies, while permitting and supporting mutualistic innovation. How should institutions be designed in the world of the Red Queen?

**Box 9.2 Unbeatability and the end of history**

The idea that the liberal order is vulnerable to the winds of creative destruction may be more digestible today than it would have been during the special complacency of the late twentieth century. At that time it was easy to think that democratic capitalism had solved all major institutional problems, when it had in fact only solved the problem of rivalry with one autocratic superpower. The rapid return to Gilded Age conditions, and the erosion of democracy even in nations that have been its foremost champions, has been the rudest of awakenings. The belief that a given arrangement is secure in the face of strategic change reflects a lack of imagination; the mass of individuals exploring the design space will prove cleverer than any theorist or lawmaker who thinks they have arrived at the end of history.

## 2.2 The Red Queen's relay

It is possible to approach these problems more systematically by distinguishing three key innovation races in the institutional hierarchies of market societies. These three arenas of innovative conflict will be the subjects of Chapters 11, 12, and 14. Together they form what I call the *Red Queen's relay*, because any efficient and stable market society must set a good pace in each of them.

Before I define them, a little more theoretical scaffolding will be useful. The Red Queen's relay is a hierarchical, multi-level interaction. Recall that Chapter 6 introduced the "external" Red Queen's race between systems (from here, this would be the race between states), and Chapter 7 introduced the "internal" Red Queen's race between actors within those systems (from here, between defectors acting to exploit the total regulatory structure and cooperators innovating to improve it). Given the discussion so far, if a state is to succeed in the between-state innovation race, it must optimise mutualistic innovation while running internal races against parasitic innovations.

Here we will be dealing with a three-level race. A natural approach is to analyse it as a multi-layer principal-agent model, and our setup shares features with the three-tier hierarchical agency models of Peltzman (1976), Tirole (1986), Laffont & Tirole (1993), and Dal Bó (2006), which extend the principal-agent theory of Ross (1973) and Mitnick (1975) to address the problem of regulatory capture. They posit a principal (usually a regulator), which in turn must motivate a supervisory intermediary, which in turn regulates firms. They assume a benign regulator, or at least they do not make inquiries into the motives of the regulator. Two kinds of races emerge: first, firms may dupe the supervisory intermediary; and second, the supervisory intermediary may dupe the regulator. As in the Red Queen's relay, the actors at each successive level of the Russian-doll system (i.e. regulator, supervisor, and firm), must do their jobs well if the outcome is to be socially optimal.

The principal-agent model is the dominant game theoretic approach for modelling institutional health and dysfunction. All 115 studies examined in Ugur & Dasgupta's (2011) review of institutional corruption took some form of principal-agent approach. The approach is natural given the hierarchical nature of institutional

structures, which must transmit payoff transformation from the top of the hierarchy down to the bottom.

The three legs of the Red Queen's relay form a hierarchical agency model that is a little more complex than the aforementioned classics. It will have three distinct races, rather than two. Two of the races we discuss are distinct from those in typical principal-agent models, in the following two ways.

First, I do not assume the regulator is benign, and will be interested in *regulatory capture*—there is a meta-institutional contest over transforming the regulator's payoffs, and so influencing who the regulator serves. The approach follows the spirit of Dal Bó's (2006, p. 207) observation that standard three-tier models of regulatory capture (regulator/supervisor/firm[111]) could equally merge the supervisor into the regulator, and, in democracy, place citizens at the apex of the hierarchy (voter/regulator/firm). The regulator then becomes the agent of the voter. However, I will depart a little from this too, because the model (A) generalises beyond democracy and encompasses authoritarian systems, and (B) recognises that voters can behave parasitically in relation to their fellow citizens. Thus, instead of "voter/regulator/firm" we will have "meta-regulator/regulator/firm", where that meta-regulator represents the actor(s) whose interests are "supposed" to be transmitted through the institutional hierarchy, and will differ depending on the political system. This will be explained more in a moment.

Second, we are also interested in the innovation race over detection and enforcement that takes place *between market actors*. This race is normally concealed. In the usual three-tier models, market actors race against the regulator, as the imposer of rules, rather than against one another. Indeed, market actors innovatively racing against one another is a driver of efficiency, *if* competition is limited to its domesticated, optimising form (O-competition). Here, this O-competitive race is not part of the Red Queen's relay. The Red Queen's relay is *a race over detection and enforcement strategies*, such that each of its legs is concerned with problems of "feral" A-competition. I mentioned in Section 1.2 that the "total regulatory structure" that produces efficient markets is a *hybrid* of institutional and individual-level detection

[111] In my terms; Dal Bó actually refers to government/regulator/firm, with a narrower meaning of regulator than mine. The substantive point is that, by any name, the intermediary may be captured.

and enforcement powers. Thus, the race between market actors is a race to evade the "rules" (i.e. detection and enforcement strategies) of other market actors. This reflects that every actor who uses rules will provoke a Red Queen's race.

### 2.3 The three legs of the relay

With that in mind, we can describe the three legs of the Red Queen's relay.

The first leg, just discussed and analysed in Chapter 11, is the between-market actor Red Queen's race. This is a race between individual actors and/or firms engaged in voluntary transactions. With ideal market institutions, individuals voluntarily engaging in privately optimal strategies inadvertently optimise for the group. That is, individuals who can detect private returns $c$ will also optimise group returns $B_g$. The first leg is mainly a race over transactor capacities to detect $c$ on the one hand, or to disguise it on the other.

The second leg of the relay, analysed in Chapter 12, is between market actors (mostly firms) and the regulator. The benign regulator aims to detect the social consequences $B_g$ of firm strategies, and the consequences of possible regulatory responses. Firms may benefit by concealing or misrepresenting the effects of their strategies or of regulatory responses, or by evading enforcement. A key problem is how to manage weaknesses in the regulator's innovation capacity relative to firms.

The third leg of the relay, analysed in Chapter 14, is between all actors (i.e. individuals, firms, and organisations, including military, judicial, media, and political actors) for control of the institutional regulator. The assumption of a benign regulator that seeks to maximise $B_g$ is lifted. Instead, regulatory actors aim to maximise their private payoffs, and will serve broad or narrow interests depending on which actors most effectively transform those payoffs. The regulator is the apex through which transformation flows into all lower orders of the institutional hierarchy. At the extreme, if an actor can gain total control over that apex, they can create rules that align all lower-order actors with their own interests (e.g. the Olsonian dictator's interest in tax maximisation, or the ideal democracy's interest in broad welfare maximisation). Or regulatory capture may take more local, narrow forms such as particular firm rent-seeking. In either case, there is an innovation race for control of the regulator.

The third leg, being at the top of the hierarchy, raises the "infinite regress" problem of Chapter 8. We will see that both authoritarian and democratic social orders require commitments to be channelled from the top of the hierarchy and through the structure. If commitment is absent, we get anarchy.

To make things a little more concrete, consider a stylised picture of how the three legs of the relay might play out in an ideal liberal democratic society. Meta-institutions, including a state constitution and judicial powers, both empower and constrain voters. The constitutionally-constrained voting public applies detection and enforcement to regulatory actors, holding them to account for producing desired social and economic outcomes. Well-motivated regulatory actors therefore have incentives to devise and apply socially optimal detection and enforcement strategies to firms, so that firms behave in ways that are consistent with voters' desires—that is, engaging in domesticated O-competition and efficiently providing public goods.

**Box 9.3 The Red Queen's relay and the Global Financial Crisis**

Take the Global Financial Crisis as a case of the various ways that the Red Queen's relay can be lost. This catastrophic event was the consequence of parasitism—that is, privately beneficial, but group harmful, strategies—in each of the three arenas of conflict. First, in Chapter 11 we will see that it involved some market actors outrunning others, with financial firms developing innovative ways to deceive and exploit boundedly rational consumers, investors, and firms. As one example, predatory lending designed to exploit limits to consumers' capacities to identify $c$ are thought to have contributed to about a third of the subprime mortgage failures.

Second, it was partly a consequence of firms evading regulators. Novel derivatives were too complex for regulatory bodies to detect their social consequences ($B_g$), i.e. whether these privately lucrative strategies were productive or parasitic. Chapter 12 discusses how the dominant philosophy of regulatory design, especially prior to but also following the crisis, has tended to intensify the innovation race and greatly increase the odds of failure.

Finally, the crisis was partly a consequence of regulatory capture. Chapter 14 will refer to research suggesting that the deregulation of the financial sector was driven by coalitions of financial interests, who used lobbyists, campaign finance, and the "revolving door" to shape regulator payoffs and so influence the regulatory strategy selected.

Finally, market actors, working in a regulated system of voluntary exchange, are sufficiently rational to handle the tasks allocated to them. Mainly this is detecting privately favourable transactions (i.e. identifying *c*) and enforcing via the decision to engage or withdraw from transactions. If the regulatory apparatus is ideal, and the market actor can identify private returns, we get optimality.

This system can, of course, go wrong in many ways—a reality I explore in two major case studies (the Global Financial Crisis and Opioid Crisis)—and I show how these result from mismanagement of the Red Queen's race.

## 3 Simplifying assumptions and their relaxation: Towards a full account of social order

The remaining chapters of the thesis can be understood in terms of the assumptions used to simplify and focus the analysis. We begin with many assumptions in place, and progressively relax them. Each time an assumption is lifted, new dynamics arise and are studied. By the end of Part 4, we end up with a complete model of democratic capitalism. The remaining two parts and their respective chapters are sketched out below, and the step-by-step process of relaxing assumptions is summarised in Table 9.1.

**Part 3** presents the argument relating to the pure problem of competent regulatory design, in the Red Queen's world of unforecastable innovation and no unbeatable strategies. Ultimately this leads to conclusions about the appropriate roles of market and state. It begins with Chapter 10, which argues that identifying an efficient division of D&E labour is the most important practical task of economics, and free markets, central planning, and mixed economies each offer competing designs.

The first two legs of the Red Queen's relay are examined in Chapters 11 and 12:

- Chapter 11 isolates the first leg, the race between market actors, by assuming that the regulator is (A) perfectly benign, which eliminates the third leg, and (B) perfectly rational, which eliminates the second leg. The regulator can neither be captured nor outsmarted. The remaining question is whether market actors can play their allotted role in the total regulatory structure.
- Chapter 12 adds the second leg by relaxing the assumption of a perfectly rational regulator. The regulator is still benign, but can be outsmarted. Thus

we are purely concerned with regulator competency in running against the parasitic innovations of market actors.

Chapter 13 summarises the argument and closes Part 3. Its main contribution is to pair the Hayekian impossibility theorem for central planning with an *impossibility theorem for libertarianism and Hayek's own classical liberalism*. To use Hayek's (1988) terms, they are twin "fatal conceits". The former conceit led to the dismal performance of planned economies, while the latter is responsible for the declining performance of liberal market economies since the 1980s.

**Part 4** is concerned with the motivation of the regulator, and the conditions that sustain democracy. It covers the third leg of the Relay:

- Chapter 14 adds the third leg by relaxing the assumption of a benign regulator. The regulator can be captured, sparking an innovation race over its control. In turn, control of the regulator determines how it runs the first two legs, and so whether the institutional structure serves general or narrow interests. We arrive at the full Red Queen's relay.

The results of the analysis in Chapter 14 are consistent with the framework in Chapter 8. The source of institutional order is commitment ($r > 0$), and well-designed rules feed back to preserve that commitment. Commitment and rules are essential complements. Authoritarian systems are anchored by dictatorial payoff-dependence ($r_s > 0$), and rules (the institutional structure) in turn defend that commitment. Yet in the democracy of self-interested actors, there is no form of commitment ($r = 0$). This stands against Olson's flawed account of democracy. It is therefore impossible to maintain the democratic order, and Hobbes (1651) was right: the democracy of self-interested actors inevitably collapses back into authoritarianism.

Of course, democratic social orders exist, and they require a commitment-based explanation. The extrinsic mode of commitment, structural payoff dependence, is ruled out. The only remaining option is *intrinsic* or motivational commitment to others' or group payoffs ($r_c > 0$). Chapter 15 provides an interlude focused on the theoretical and empirical basis for other-regarding preferences. It largely echoes Adam Smith, whose liberal market order rested on "sympathy" and "public-

spiritedness". The process of relaxing assumptions continues into Chapters 16 and 17.

- Chapter 16 drops the assumption of self-interest, revisiting the Red Queen's relay with the addition of weak other-regarding preferences. Because the framework indicates that $r$ anchors long-run cooperativity, a key puzzle is *how democracies leverage substantial cooperation from relatively low values of* $r_c$.
- Chapter 17 drops the assumption that individuals' other-regarding preferences are non-parochial. It allows faction and fanaticism: individuals may have affiliations with, and antipathies toward, the different cultural sub-groups of society. Self-interest is a rare and trifling phenomenon; as Smith argues, *narrow other-regarding preferences are the main threat to the liberal order*.

Chapter 18 concludes the second core argument of the thesis. It extends the Part 3 argument—that efficient self-regulating markets are technically impossible—to its political implications. Incompetence in running the first two legs of the Red Queen's race, in our time due to the influence of libertarianism and classical liberalism, guarantees trust-eroding crises and the rise of faction and fanaticism. Policymaking in the libertarian or Hayekian mode will shift democracy towards authoritarianism.

Hayek's "road to serfdom" is real, but it points in two directions. After rejecting central planning, we have strolled dangerously close to its equally flawed inversion.

**Table 9.1** *Progressively lifting assumptions through the remaining chapters of the thesis. Chapter 15 is excluded, as a theoretical interlude on other-regarding preferences.*

| | | Perfectly rational market actors | Perfectly rational regulator | Perfectly transformed regulator | Self-interested agents | Homogenous rc (i.e. no factions) | Legs of the RQR |
|---|---|---|---|---|---|---|---|
| **PART 3** | **Chapter 11** | Relaxed | Imposed | Imposed | Imposed | N/A | **1** |
| | **Chapter 12** | Relaxed | Relaxed | Imposed | Imposed | N/A | **1,2** |
| | **Chapter 13** | **Concluding on the roles of market and state: Two fatal conceits.** | | | | | |
| **PART 4** | **Chapter 14** | Relaxed | Relaxed | Relaxed | Imposed | N/A | **1,2,3** |
| | **Chapter 16** | Relaxed | Relaxed | Relaxed | Relaxed | Imposed | **1,2,3** |
| | **Chapter 17** | Relaxed | Relaxed | Relaxed | Relaxed | Relaxed | **1,2,3** |
| | **Chapter 18** | **Concluding on the conditions that sustain democracy: Two roads to serfdom.** | | | | | |

# PART III

## On the technical problem of regulation:
## The complementarities of market and state

*Any approach, such as that of much of mathematical economics with its simultaneous equations, which in effect starts from the assumption that people's knowledge corresponds with the objective facts of the situation, systematically leaves out what is our main task to explain.*

Hayek (1945, p. 530)

*…human decisions affecting the future, whether personal or political or economic, cannot depend on strict mathematical expectation, since the basis for making such calculations does not exist.*

Keynes (1936, ch. 12, §VII)

# 10

# Free markets, central planning, and the mixed economy: Competing divisions of detection and enforcement labour

On to the applications. To reiterate, Part 3 takes individuals as self-interested, so that systematic cooperation is the product of rules. Further, the regulator is a benign technocrat. This allows a total focus on the purely technical problems of regulating economic activity in the world of the Red Queen, mainly in markets but with some reference to central planning.

Before analysing the first two legs of the Red Queen's relay, it will be helpful to make the idea of the "division of detection and enforcement labour" more concrete, and to show why it is so momentous. Indeed, it is reasonable to say that identifying the appropriate division of labour is the core practical task of economics. Accordingly, the great debates about economic regulation over the last century can be characterised as disputes over the appropriate division of labour. The great achievements and failures of economic regulation reflect success or failure in the design of this division of labour.

It is worth repeating the most relevant theoretical points developed up to this point, so that they are front of mind.

- Private and social returns must be systematically aligned to produce growth, and here they are aligned only by rules[112]—by the successful achievement of a multitude of detection and enforcement (D&E) tasks.
- The total regulatory structure always involves a division of D&E labour. In all societies this division allocates some tasks to *institutional* D&E (i.e. the regulator) and others to *individual-level* D&E (i.e. market actors). The total regulatory structure is always a hybrid.
- Rules always create innovation races, such that there is a race over market actor D&E (Chapter 11) and a racer over institutional D&E (Chapter 12). Efficiency requires that both races are run well.

This chapte discusses the major competing designs aimed at realising efficiency.

Section 1 articulates the divisions of D&E labour in free markets and centrally planned societies. It shows how these emerge from different *assumptions* about the *comparative advantages* of institutional regulators and market actors.

Section 2 examines the significant, but partial, progress that economics has made in replacing these assumptions with knowledge. We know enough about the comparative advantages of regulator and market actor to see that central planning is dysfunctional. The rest of the territory remains contested.

**Box 10.1 On actor capabilities and the limits to neoclassical theory**

In some respects the analysis in Part 3 is substantially Hayekian, although some of its conclusions differ starkly from his. Hayek pointed out that neoclassical theory points to the desirable optimum, but leaves wholly unanswered the all-important practical question of how to get there. As noted in Chapter 1, the neoclassical model is consistent with anything from anarcho-capitalism to central planning. This is because it is silent on actor capabilities, and therefore silent on the appropriate division of D&E labour between regulator and market actor. Per the epigraph of this chapter, it "systematically leaves out what is our main task to explain" (Hayek, 1945, p. 530), and the practical question is decided by assumption.

[112] We are dealing with mercenary actors. Assuming self-interest means $r_c = 0$. Absent a dictator, no market actor has a share in total output large enough to produce meaningful structural payoff-dependence, so $r_s = 0$.

## 1.1 The division of D&E labour in grand theories of economic management

The 20th century was the staging ground for a contest, in theory and in practice, between different divisions of detection and enforcement (D&E) labour—those different divisions that comprise self-regulating markets, central planning, and the various possible designs of mixed economies.[113]

At one pole on the spectrum is the minimally regulated market society. The regulator is allocated the narrow task of enforcing the conditions for voluntary exchange and free entry, and perhaps internalising a few externalities. Market actors are allocated the task of identifying the private payoff $c$ for myriad transactions, and so deciding whether to engage or withdraw. Suppose that market actors can detect this private payoff $c$. In this case, every *voluntary* transaction necessarily benefits both parties—otherwise they would not agree to transact. Voluntary transactions therefore always lie in the mutualism quadrant (i.e. they benefit the group, so $B_g > 0$), at least if there are no externalities. However, this would also be true of voluntary transactions undertaken in highly inefficient markets riven by monopoly and monopsony: rational voluntary transactions would still be welfare-increasing, but only a suboptimal set of transactions may be available. Thus, we can improve things further with the free entry of firms, which allows new entrants to engage in domesticated competition over participation in transactions—and so over the efficiency with which preferences are satisfied. These core insights, about the marriage of institutional and individual D&E into a working system, have been with us since Adam Smith. Later formal work showed that under extremely favourable assumptions about (A) market actors' capacities to identify $c$ and (B) market competitiveness, this combination leads to the ideal satisfaction of preferences (Arrow & Debreu, 1954).

On the other end of the spectrum is the centrally planned society. It allocates a much smaller role to individuals, who may no longer be called "market actors". They need only select among a limited range of goods produced by the regulator, or in a more comprehensively planned model, goods may simply be allocated on a per capita basis or by some centralised assessment of need. Now it is the regulator that must solve

[113] It may by now little surprise the reader that such divisions of detection and enforcement labour are also found throughout nature. See A.10.1 for an example that would have fascinated Becker, who thought entomology would be a fruitful source of economic analogies (Becker, 1976).

the problem of resource allocation, balancing supply and demand by means other than the market. It must employ a D&E structure to motivate actors to be productive, in the absence of competitive labour and capital markets. If we adopt favourable assumptions about the regulator's capacity to identify $B_g$ and solve these problems, e.g. via the calculation of shadow prices, the centrally planned society likewise achieves optimality. This was shown by the neoclassical theorists Barone (1908) and Pareto (1906/2014).

For the agnostic theorist approaching the problem of economic organisation for the first time, who lacks any firm evidence on actor capacities, both of these models are plausibly effective *and* plausibly catastrophic. The *individual* identification of $c$ in the market and the *institutional* identification of $B_g$, i.e. market and hierarchy, may be perfect substitutes, essential complements, or one or the other may be strictly superior. That is, maximisation of group payoffs might be achievable purely via a state regulator, purely via individual action, and/or by some particular combination. Every point on this spectrum has had its own clutch of advocates: central planners, libertarians and classical liberals, and supporters of a wide variety of possible mixed economies.

To begin to place constraints on what an efficient overall D&E structure looks like, it is necessary to learn more about the capabilities and comparative advantages of these various actors, and the kinds of D&E tasks they will be confronted with.

### 1.2 Assuming the conclusion

Despite the problem of market actors' and institutional regulators' capacities being so determinative, it received—as Hayek never tired of complaining—peculiarly little attention. One cause of this neglect is that the problem requires difficult empirical work and is challenging to reconcile with formal models.[114] Another is that the answer to this *scientific* question has *political* consequences, and infringes on cherished ideological and moral commitments. Friedman (1953) put it well:

[114] Behavioural models of market actors, for example, produce laundry lists of heuristics and biases that interact and manifest differently across contexts and points in time. There is no way to insert such complex agents into Arrow-Debreu and still come out with an exact answer with respect to the consequences of markets.

> *Laymen and experts alike are inevitably tempted to shape positive conclusions to fit strongly held normative preconceptions and to reject positive conclusions if their normative implications – or what are said to be their normative implications – are unpalatable.* (p. 146)

For much of the 20th century, assumptions and ideology played outsized roles. We assumed the conclusion.

What happens when we assume that the regulator is a superior substitute for market actors? This mechanically inflates its role in the division of D&E labour and presses us towards planning.

- As Hayek (1960/2020) observed in *The Constitution of Liberty*, if planners are taken as omniscient, this profoundly undermines the case for markets.[115] The regulator could perfectly identify $B_g$ and market actors would have no advantage in the use of distributed knowledge.
- If we assume wholly irrational market actors, or irrational consumers alone, we get to the same result. Transactions would not on average be utility-increasing,[116] and nor would they discipline producer efficiency. Perfect market institutions plus irrational consumers produce gross inefficiencies and crises.

Indeed, advocates of central planning have tended to be excessively optimistic about regulator D&E and pessimistic about that of market actors.

Or we may make the opposite assumption: that market actors are infinitely capable. This naturally expands their role and shrinks that of the regulator, pressing us towards libertarianism. Extreme assumptions about market actor capabilities have been standard fare in economic theory. As Krugman (1996) summarises,

[115] "If there were omniscient men, if we could know not only all that affects the attainment of our present wishes but also our future wants and desires, there would be little case for liberty." (Hayek, 1960/2020, p. 81)

[116] There have been some misleading results on this front. Gode & Sunder (1993) show that "zero intelligence" traders can produce efficient markets, but the main task of intelligence is snuck in the back door: their traders can perfectly compare quality and so only need to pick the lowest price.

> *At least since Paul Samuelson published* Foundations of Economic Analysis *in 1947, the overwhelming thrust of conventional theory has been to say that agents are not only intelligent, they maximize—that is, they choose the best of all feasible alternatives.* (§3)

It is well worth recalling that the maximisation assumption was born as a modelling expedient. Vilfredo Pareto (1906/2014) is the source of the ordinal welfare theory that Samuelson (1938) would later redescribe as "revealed preference".[117] Pareto pointed out that economists could sidestep extreme empirical challenges if they treated observed behaviour as always revealing actual preferences, and then used this to construct preference orderings (Pareto placed strong constraints on the domain over which this method is appropriate, a matter for Chapter 11). The rationality assumption simplified the analytic problem, but it has a side effect: *perfect rationality implies perfect market actor detection powers*. The power of voluntary transactions becomes enormous, there is no race between market actors, and the regulator need only enforce basic market institutions.

Thus, an expediency gradually came to "answer" one of the most important scientific questions of economics.[118] Deregulation naturally becomes the default regulatory strategy—a position consonant with the belief in benign innovation, and view of market failure as the exception rather than the rule, in Chapter 5.

It is helpful to note that, in the framework here, "deregulation" is a misnomer: it is not a *reduction* in regulation. It is a reallocation of D&E tasks from the regulator to individuals. Moreover, it may *increase* the complexity of the D&E tasks that must be accomplished. Can market actors achieve them? Or are we, like the central planners of yore, building a regulatory structure that doles out impossible D&E tasks?

### 2.1 Partial progress: Discrediting central planning

The rejection of one of these extreme divisions of D&E labour, central planning, has been the chief success of the last century. The likely difficulties facing central planners were anticipated by Barone (1908) from the outset. His paper providing a

[117] See Houthakker (1950) on the identity of these theories, which Samuelson accepted.
[118] It was paired with theories of regulatory incompetence, with Hayek's theory of the blind evolution of institutions being an extreme example. Following the Chapter 5 discussion of the limits to blind evolution, within a design space that is vast beyond our imagining, his theory is impossible.

mathematical basis for central planning warned that the data collection and coordination problems would be vast and possibly unsolvable. The decisive discrediting of central planning would, however, take most of the rest of the century. Such a tremendous feat required the marriage of theory with catastrophe:

- Barone's caveats being extended most notably by von Mises (1922) and Hayek (1935/1967), with Schumpeter (1934; 1942) offering some important contributions on the role of the entrepreneur; and
- the observed failures of the USSR and communist China.[119]

Either alone would be insufficient—a hypothesis without evidence, or evidence without a theory to interpret it.[120] As it turns out, the theories of distributed knowledge and of the entrepreneur are consistent with the observed failures of allocation and innovation in centrally planned states. They are also consistent with the USSR's recourse to quasi-markets to manage the allocation of labour and consumer goods, and some innovation processes (e.g. Katsenelinboigen, 1977; Gaddy et al., 1991; Harrison, 2006).

## 2.2 Stalled progress: The limits to libertarianism and Hayekian liberalism

Progress in articulating the technical limits of the other extreme division of D&E labour—of libertarianism and Hayek's classical liberalism—has been more limited. Another symmetry with central planning is the sounding of warnings from the outset about the limits to actors' powers of detection. In this case, limits not to Barone's planner, but to Pareto's market actors. Pareto (1906/2014) laid great emphasis on the significance of irrational forces for understanding human behaviour, both within and outside of markets. His rationality assumption was to be used for a "pure" economic science concerned with describing *logical action*. He thought it wholly inappropriate for "applied" economics, i.e. the study of *actual* actions and

[119] Just before the fall of the Berlin wall, Samuelson and Nordhaus (1989) could still write that the USSR showed that centrally planned societies could "thrive". After its fall—and much improved data on actual performance—this position was no longer tenable.

[120] Deductive theories are insufficient on their own. They allow for hypothesis formation, but must be judged by their fit with empirical evidence. On the other hand, empirical evidence is insufficient without a theory to interpret it. Particular failures could be attributed to *particular*, rather than *general*, causes: a theorist might reasonably wonder if it was not central planning at fault, but the particular methods of Stalin, Mao, and their followers. We make progress when theory predicts a pattern of failures, and they turn up in the data.

appropriate policy responses (pp. 72, 124, 184-187). I discuss Pareto's views further in Chapter 11.

Keynes' (1936) *General Theory* was the most obvious breakthrough. To continue the pattern, it was another marriage of theory with catastrophe. Gratefully, progress did not in this case require the death of democratic capitalism—although it did require a near-death experience. The Great Depression and other similar crises are, in Keynesian theory, intimately tied to complexity, uncertainty, and limits to market actors' capacities to identify $c$. Market behaviours are affected by irrational animal spirits and sticky expectations. In turn, the limits to market actors' capabilities expanded the role of the regulator, in this case as a provider of a public good: moderated demand. Lesser but ubiquitous breakthroughs can be found in the ever-growing regulatory apparatus used to simplify transactions in an ever more complex economy (see A.10.2 for more discussion).

The discrediting of central planning mercifully narrowed the territory of reasonable debate. If other gravely damaging proposals retain credibility, it would be worthwhile seeing if they can be understood and cleared away too. From here, the theoretical problem of regulating in the world of the Red Queen is married with the Global Financial Crisis, Opioid Crisis, and the general stagnation of contemporary capitalism. In short, I follow the pattern again: theory is combined with catastrophe, to see if it is possible to place further constraints around the appropriate roles of market and state.

*The market is a place set apart where men may deceive one another.*
Anacharsis of Scythia (c.600 BC)[121]

*…consumer's sovereignty can never be established as long as the initiative lies with the producer. For the general run of consumer goods, the buyer is necessarily an amateur while the seller is a professional.*
Joan Robinson (1969, p. xii)

# 11

# The first leg of the relay: The race between market actors

Now we turn to the first leg of the Red Queen's relay: the innovation race between market actors. This leg, like the other two, is an innovation race over the use and evasion of detection and enforcement (D&E) strategies. In terms that may be more familiar, one could characterise it as a race over rationality and access to information, or a race over market actors' capacities to impose, or evade, "market discipline". This race has been profoundly important in the history of capitalism, both distant and recent. Most notably, it is behind the most severe systemic crises of capitalism over the last century or so—the Great Depression and the Global Financial Crisis—as well as costly, if more contained, calamities in healthcare, insurance, education, and other complex markets.

Section 1 points to the developing body of research on the limits to individual rationality. Rationality is not an infinitely flexible multi-tool, but a set of concrete capacities to search the design space for privately rewarding strategies. In a Red Queen's race, of particular importance are the *changing asymmetries in market actors' innovation capacities*, and how this affects the division of D&E labour.

Section 2 focuses in particular on the changing innovation capacities of producers and consumers. Recent decades have increasingly favoured producers. The average consumer, who does not understand compound interest, is pitched against firms

[121] As quoted in Laërtius (3rd century AD/1925, p. 105).

armed with rational-scientific optimisation machines, composed of large teams of experts using big data, computer modelling, and increasingly AI, to model consumer behaviour.

Asymmetric optimisation capacity is only a problem *when detection tasks are complex*. If the adversarial contest is over the playing of tic-tac-toe, access to big data, artificial intelligence, and so on, provides no advantage—the contest remains symmetric. If the task is identifying or concealing the value of complex securities, medical treatments, or political information, asymmetries translate into opportunities for extensive parasitism. Complexity is the enemy of the consumer and friend of the modern firm, and firms, accordingly, invest in designing transactions that consumers cannot understand.

I conclude that market actors must be allocated detection tasks they can accomplish. Failing that, I argue that crisis is *guaranteed* in any market that is complex and welfare-critical, such as finance, insurance, healthcare, education, and political information markets. The Chapter 11 Empirical Supplements examine the Global Financial Crisis of 2008 and the US Opioid Crisis in more detail.

### 1.1 Normative versus positive accounts of rationality

Herbert Simon (1986) contrasted the rationality of economics with that of psychology. Economics posits that behaviour is "globally consistent" and "objectively rational in relation to its … present and future environment as the actor moves through time" (p. 279). Or as Krugman (1996) put it, it assumes people maximise, choosing "the best of all feasible alternatives" (§3). Psychology, in contrast, seeks to understand concrete cognitive procedures, how they affect behaviour, and how they shape our goals. As Simon (1986) observes, psychological research finds that only "selected aspects of reality are noticed", and "nonrational processes… influence the focus of attention and the definition of the situation that set the factual givens for the rational process". In turn, those inputs feed into "very limited information processing capacities" (Simon, 1986, pp. 279-280).

As is widely acknowledged, the rationality of economics is normative—a theory of how agents *ought* to make decisions (Briggs, 2019)—while the rationality of psychology is positive—a description of *actual* cognitive procedures and capacities. Again, Pareto (1906/2014) is useful as both the progenitor of the normative method,

and at the same time an advocate for the positive method in practice. He offers a simple model of the relation between the two: optimal choice (the "subjective fact") is united with actual choice (the "objective fact") when learning has taken place:

> *we will consider only repeated actions to be a basis for claiming that there is a logical connection uniting such actions.* (Ch 3, §1)

The use of the rationality assumption was, of course, extended far beyond this domain. Hicks' (1939) attitude is representative:

> *a great many such extensions appears at once when we consider how wide is the variety of human choices which can be fitted into the framework of the Paretian scale of preference.* (p. 24)

Much of the work on heuristics, by Herbert Simon and behavioural economists such as Tversky and Kahneman (1973; 1989), can be taken as further development of Pareto's reasoning: there are various reasons that the subjective and objective facts may not coincide, and they are worth understanding. Simon (1968/2019; 1972) discusses the process of learning in detail, and adds the problem of complexity—the reason that no human, even with infinitely repeated action, fully understand the game of chess. Complex problems and incomplete learning lead people to use heuristics, or detection strategies that are, to varying degrees, flawed.

Actors with advantages in innovation capacity, both in their capacity to handle complexity and their capacity to invest in improvement (from Chapter 5, matters of trial quality and quantity), will employ superior strategies. The reader of this thesis is quite possibly a better chess player than the author. Garry Kasparov would have trounced us both. Kasparov, in turn, was beaten by IBM's Deep Blue in 1997, Deep Blue is outclassed by modern desktop chess software, and none can compare to the AlphaZero learning algorithm. But each new version of the latter algorithm can generally beat past iterations. And so it goes. But no future AI, nor any entity that shall ever exist, will reach the sublime rationality of *Homo economicus*.

It is not our purpose to restate the particular quirks of human thinking catalogued by the behavioural economic and related literatures, though we will touch on some evidence in this chapter and the empirical supplement that follows. Our main goal is to explore the implications of asymmetric rationality within the Red Queen's race.

### 2.1 Growing asymmetries in producer versus consumer innovation capacity

The between-market-actor race occurs between transacting firms, and between firms and market actors. Here we will focus on the case where the balance between racers is changing most rapidly, and has the clearest welfare implications: that between the producer and consumer. As discussed in Chapters 6 and 7, the outcomes of the Red Queen's race depend on three factors:

(A) the stochastic cooperator-defector bias in strategic change in a given period;
(B) relative innovation capacities, which increase the likelihood of finding an adaptation in a given period; and
(C) the payoff-transforming environment, which in this case refers to the regulatory design of the market.

When factors A and B favour a balance between consumer and producer rationality—a balance between capacities to conceal versus detect the value $c$—then the regulator's strategy for the design of C can be deregulatory. Where A and/or B bring growing imbalance, e.g. favouring "mimics" (see Box 11.1) over productive transactors, then market discipline fails and efficiency declines. Any efficient regulator design C will attempt to compensate, a matter for Chapter 12.

Concerns about consumer manipulation are not new. A century ago, McCraw (1986, p. 107) writes, judge Louis Brandeis urged journalists to "teach the public" to "look with suspicion upon every advertised article." The factors A and B of the preceding paragraph were, however, not too unfavourable. The science of psychology, and so of cognitive manipulation, was nascent and crude. Data was scarce and expensive. The range of goods and services was smaller and simpler. Sometimes the advantages possessed by firms were softened by changes to the regulatory environment, such as penalties for misleading advertising or restrictions on the incentive structures that apply to trusted advisors such as financial planners and doctors. Sometimes they were not. One case that looms large is that of the manipulative financial strategies that aggravated the Great Depression, which I discuss a little later. Another is that of tobacco firms, which learned to employ doctors to promote their products, to advertise to children, to increase the addictiveness of their products, and to suppress

mounting information about the harms of cigarettes. As history tells us, consumers developed few effective counter-

**Box 11.1 Mimics, models, and their detection**

Any system that falls into the general theory developed in this thesis must grapple with bounded powers of detection. Sociobiology is no exception, and deceit plays a significant role in social interactions throughout the kingdom of life. To apply a normative approach to understanding such systems would obviously be a permanent obstruction to grasping the positive reality. Much would be inexplicable. Though I might have preferred to weave the sociobiological and human threads alongside one another throughout this thesis, space is lacking. See A.11.1 for a discussion of some fascinating cases, including of deception in biological "markets".

One bit of terminology I would like to borrow is that of "mimicry". Strategies that involve evasion of detection, especially in market-like sociobiological settings described in A.11.1 and A.4.4, can be described as an effort to *mimic the appearance of* cooperative exchange partners (the cooperative partner is called the "model", its features simulated by the mimic). In the anthropic case, market actors may pursue strategies that mimic the learned signals of transaction favourability.

The value of the term is that is suggestive of a strategy pursued with a certain degree of skill, and so immediately points to Red Queen dynamics of strategies and counterstrategies: a contest between mimicry and the ability to discern mimic from model. It helpfully removes some of the normative baggage in economic descriptions of bounded rationality. It avoids discussing market actors in terms of the loaded binaries of "rational" and "irrational", or "informed" and "uninformed"—the second term, in both cases, being pejorative. Rather, we have one optimiser racing against another. What matters is who has the greater resources for strategic innovation, and it is surely obvious that the availability of such resources may be asymmetric.

In sociobiological "markets" too, the presence of parasitic mimics harms "consumers" as well as productive "producers", and reduces total system efficiency. It triggers a Red Queen's race, and the outcomes of that race—and so system efficiency—are substantially a product of relative innovation capacities.

Strategies, and the regulator was slow to adapt. Annual tobacco-related health and productivity losses still reach extraordinary magnitudes in the twenty-first century, estimated at around 1.8 percent of global GDP (Goodchild et al., 2018).

Today we are entering a concerning stage of history for the producer-consumer race. The last few decades of technological change have made the storage, transmission, and utilisation of information progressively cheaper. Firms have access to increasingly detailed and individualised data on consumer behaviour,[122] means of cheaply conducting individually-targeted experiments, and growing capacities to construct complex behavioural models with techniques such as machine learning. Much of this knowledge is *cumulative* and *nonrival* in the senses described by Romer (1990); that is, there is a fixed cost for generating new strategies, and a low or sometimes effectively zero marginal cost for their application and distribution. Consumer rationality, meanwhile, remains comparatively fixed and constrained by behavioural biases. Growth in cognitive and mental storage capacity does not expand with Moore's law. Where products are complex, learning is frequently impossible; this is one reason why improved disclosure rules have vanishingly small impacts on many consumer behaviours (e.g. see Macro International's 2008 study for the US Federal Reserve). Where learning does occur and accumulates over the lifespan, it cannot be costlessly and instantly downloaded to others; rather, each new generation must relearn even the most basic strategies from scratch.

Today's market societies pitch consumers with relatively static capacities against rational-scientific innovation machines, composed of the collective efforts of many thousands of marketers, psychologists, and data analysts with a rapidly improving toolkit for mimicking value. Firms may acquire almost complete knowledge of individual customers' weaknesses. The race would be balanced if the reverse was also true: if consumers had rich data and models of each firm's strategies. As it is, consumers become more transparent while firms increase the complexity of their strategies and become more opaque—the topic of the next section.

[122] Retailers can pay to access "data enrichment services" that provide detailed information about individuals, such that a given email address may be associated with not only their gender, marital status, zip code, and whether they have children, but also their spending proclivities, patterns such as impulsivity and responsiveness to advertisements with particular framings, and so on.

### 2.2 Complexity and the race over innovation capacity

This asymmetry is not intrinsically dangerous. Two further conditions are required to introduce serious harm:

- complexity, or rationality and information-intensity, of the task of detecting *c*; and
- high costs of failure, where the choice is welfare-critical or has spillovers that make private optimisation a public good.

I discuss complexity here, and welfare-criticality in the next section.

Where detection tasks are simple, asymmetries in rationality are of little consequence. Take, for example, a team of virtuoso logicians armed with the world's premier supercomputer playing a game of tic-tac-toe against an ordinary teenager. The simplicity of the game means that, with a little learning, the performances of both teams will soon be symmetric. Or take a case where products are commodity-like, this is obvious to all transactors, and so consumers need only compare prices. In such cases, the limited and simple universe of strategies means there is no Red Queen's race. But there are very few, if any, real-world cases of such simplicity, or where the parameters of the game are so artificially fixed. The more product complexity grows, the more consumers must rely on cognitive shortcuts, emotional valence, and so on, to make choices. Where firms have the means to understand how the average consumer uses such shortcuts, this opens up a whole field of activity for distorting consumer perception of *c*.

Asymmetries in the Red Queen's race, then, motivate efforts to *increase and exploit problem complexity*. Firms invest in creating products that can be understood by in-house modelling, but are opaque to consumers (e.g. Gabaix and Labison, 2006, and Anagol & Kim, 2012, on firms' incentives to "shroud" product attributes). Consider the US healthcare insurance market, where consumers must choose between tens of thousands of products that they cannot understand. In one behavioural study, only two percent of participants could correctly identify the cost of a single hospital stay under a typical insurance plan, even with relatively straightforward figures and all necessary information provided (Loewenstein, et al. 2013). In a study by Handel (2013), the average employee could cut their health insurance fees by around half by

switching to better plans.[123] Sophisticated in-house models of product design ensure that this remains the case. A final and important contribution on the matter of healthcare comes from Abaluck & Gruber (2016, p. 1), who find that the welfare costs of poor choice in health insurance markets are *increasing* over time. Firms' parasitic strategies grow more sophisticated, while among consumers there is "little learning at either the individual or cohort level."

Finance is another industry with complex products. We will turn to the Global Financial Crisis shortly. Before that, consider some behavioural studies selected from a large literature: a minority of consumers have a sound understanding of compound interest, and just *two percent* can do a simple compounding calculation (Hastings & Mitchell, 2011).[124] That alone should cause alarm at the misuse of Pareto's "pure economics"; he might observe that there is little opportunity to learn, because in the short-run there is no noticeable difference between simple and compound interest. In the field, the average user of payday lending thinks annualised interest rates are around 10 percent when they are in fact 443 percent (Bertrand & Morse, 2011). The fees associated with effectively identical S&P500 index funds, which should be obviously commodity-like, vary by an order of magnitude with no sign of convergence over time because consumer choices are shaped by irrelevant information and emotional valence (Hortaçsu and Syverson, 2004; see also Choi, Laibson and Madrian, 2010, who show that even an Ivy League business school education provides zero to minimal improvements in choice).

Similar problems occur in a range of other sectors from energy, where producers rely on powerful status quo biases that leave consumers with overpriced plans, to gambling, where Bet365, for example, uses algorithms to identify high-rationality gamblers that are likely to win and restricts them from making bets (Cannane & Taylor, 2019), and gaming, where firms are learning to generate and exploit addiction and turn children into gamblers (e.g. Drummond & Sauer, 2018; Li et al. 2019; and Gambling Commission, 2022).[125]

[123] Similar results were found by Abaluck and Gruber (2009): only 12 percent of consumers optimised in selecting healthcare plans, and for the remaining 88 percent the average possible savings were around 35 percent.
[124] The two percent that could were markedly wealthier and more educated.
[125] The Gambling Commission (2022) found that 44 percent of children and adolescents had purchased "loot boxes" designed with gambling-like incentives.

If we place a D&E burden on consumers that they cannot carry, Friedman's *ex post* "natural selection" of firms over profitability means that the surviving firms will be the best predators (see Box 11.2).

**Box 11.2 Friedman's "as if" rationality as a driver of market inefficiency**

Again, following Alchian (1950), Friedman (1953, p. 158) argued that even if the rationality assumption is unrealistic, firms can be treated "as if" they are rational because of a process of "natural selection" over *ex post* profits. This is a reasonable theory as far as it goes, which is not very far (as discussed in the introduction to Chapter 2). Even if this theory holds perfectly, it is decidedly neutral with respect to the merits of the free market. It solves the problem of pace, not direction, and pace without direction can be catastrophic.

The issue is that firms' rational optimisation would equally be dedicated to the exploitation of consumers. Crucially, Friedman's selection process only works to enhance firm optimisation, not consumer D&E. Firm survival depends on profitability, while consumers that cannot identify *c* persist within the system. The more Friedman's "as if" rationality argument holds for firms, the more it exacerbates asymmetries between firms and consumers.

Herbert Spencer's (1851) solution was selection over consumers too—a *social* Darwinism, where society "excretes" its "imbecile members" (i.e. insufficiently rational consumers) in a "purifying process" leading to the "multiplication of the competent" (p. 324).[126] Such an attitude is implicit in Hayek and Friedman's shortage of sympathy for exploited consumers—Friedman's "fools" and "suckers".[127] Given the extraordinary rate of cultural evolution (for a striking example, advances in artificial intelligence) versus the plodding pace of biological evolution, we may all be excreted in the long-run.

[126] Hayek's views substantially overlap with Spencer's, though he remarked that Spencer "spoiled a good argument by the crude and insensitive way in which he applied it" (Hayek, 1958, pp.243–244; also see Hayek 1979, p.174 and Hayek, 1985). There are two chief distinctions between their approaches. First, Hayek (1988, p.25) focused on the selection of norms and practices, rather than of individuals. The difference may end up slight, as Hayek implicitly endorses the excretion of those born into uncompetitive cultural groups. Second, Hayek proposes basic income support for individuals incapable of work (e.g. the disabled), though no parameters are specified.

[127] Friedman's approach to the business-exploited consumer and voter is one of individual responsibility—of "caveat emptor" and of "confidence men" outwitting rubes (1978, 108:48), where we are "suckers" (1965, p. 15) and "fools" if we "let them exploit us" (Friedman, 1990, p. 5).

It would be remiss to ignore countervailing technical advances that may assist the consumer. Information technology has provided new tools for assessing prices (e.g. comparison-shopping websites) and product quality (e.g. consumer reviews). Yet the innovation race continues apace. For the innovative firm these present new opportunities for misrepresenting *c*, such that there is now "a large and fast-moving online market for fake reviews" (He et al., 2022, pp. 1-2). The German-based travel booking company Trivago, for example, advertises itself as an "impartial and objective" platform for price comparison, but the Australian Competition and Consumer Commission found that its rankings were misleading in 66 percent of cases, and "were largely based on which advertiser would pay [Trivago] the most money" (Khadem, 2020, para. 7). From an analysis of ReviewMeta data, the *Washington Post* reported that 61 percent of electronics product reviews on Amazon were fraudulent (Dwoskin & Timberg, 2018).[128] If the consumer is to be rescued from the race by new information services, they must first have the D&E capacity to tell "mimic" information services from the "model". Scant few consumers utilise sophisticated review-verifying tools such as ReviewMeta, and even these are vulnerable to mimicry. As a result, the new information services can be expected to accentuate the divergence of consumers into savvy and parasitised sub-groups.

As social scientists, we will usually find ourselves in the savvy group and may see little reason for concern—but we bear responsibility for the ways that D&E labour is allocated. Friedman, for example, hoped to eliminate the licensing of doctors.

> *No more licensing of doctors. No more regulation of drugs. Not of any kind. Period.* (see Pearson & Shaw, 1993, p.39)

He may personally have managed very well in such a world, being educated, wealthy, and well-connected. The masses of patients who cannot detect the payoffs of a medical procedure offered by a local charlatan, would, in his view, bear the blame for being "suckers". This is the economist playing disciplinarian: health, wealth, and safety are to be distributed only to deserving actors who have appropriate capabilities, while those lacking such capabilities are blameworthy. Welfare losses are morally justifiable. In the world of the Red Queen, such standards of

[128] Amazon, for its part, is engaged in its own Red Queen's race with fake reviewers. The company has established policies and procedures for weeding out fake and "incentivised" reviews, but ReviewMeta data suggests that average review trustworthiness has continued to plummet. See Woollacott (2017).

deservingness and blameworthiness are arbitrary and dynamic. The difficulty of keeping up with parasitic innovations changes over time: those who deserve success today may deserve blame tomorrow. If the strategies of mimicry improve and the difficulty of detection increases, then the list of "deserving" consumers shortens. It moves dangerously close to the logic of Spencer's social Darwinism (Box 11.2), where the cleverest exploiters deserve to eat their quarry.

If the goal is to maximise social welfare and achieve broad prosperity, then we will attribute blame and deservingness quite differently. If economists and policymakers set patients and consumers hyper-complex D&E tasks they cannot possibly accomplish, then we are to blame for the resulting private and systemic harms. We have failed the task set to us: to design a functional total regulatory structure.

### 2.3 Welfare-criticality: Turning complex markets into crises

Finally, asymmetries and market complexity matter most in *welfare-critical* markets. If bounded rationality mostly results in minor private welfare losses—the purchase of some overly-expensive luxury goods here, or some shoddy electronic equipment there—then the consequences of the rationality assumption are modest, though still of some interest. Empirical studies find that "mimics" on Amazon produce welfare losses for consumers, and may harm high-quality sellers by weakening the value of the review signal (He et al., 2022). These welfare losses may be significant; one field experiment found that fake reviews cost around 12 cents in every dollar spent (Akesson et al., 2022).

However, most of the cases considered in this chapter occur in markets with substantial welfare consequences—healthcare, finance, and insurance, among others. Here, failures in market actor D&E may

- produce unacceptably large private harms, which at scale may also produce systemic damage; and
- be associated with major positive or negative externalities.

In such markets, widespread private failures to identify *c* may produce systemic impacts that compromise the integrity of both the market and political system. I turn to two such cases in the empirical supplement to this chapter: the Global Financial Crisis and the Opioid Crisis.

# 11E1

# Empirical case 1:
# The Global Financial Crisis and the first leg

The discussions in these empirical supplements, here to Chapter 11 and later to Chapter 12, require some qualifications. They are necessarily brief and omit many important details. I cannot improve on the technical diagnoses offered by others, and indeed completely depend on pre-existing research. The purpose is to place this existing work within the framework of the thesis, and so within the Red Queen's race, to show how these crises illustrate more general phenomena—and the shortcomings in philosophies of regulatory design. The analysis may help us identify which of the available diagnoses, and proposed solutions, may address or aggravate the problems raised by the Red Queen's race.

## 1.1 Systemic damage: The between-market actor race and the Global Financial Crisis

The leading economic theories of finance—the capital asset pricing model (Sharpe, 1964; Lintner, 1965), the theorem of capital structure irrelevance (Modigliani & Miller, 1958), modern portfolio theory (Markowitz, 1952), and the efficient market hypothesis (Fama, 1970)—together suggest that risks in financial markets are correctly priced, that leverage is no threat, and that bubbles cannot form in the absence of regulatory distortions. Their influence cannot be understated; each of these ideas is associated with a Nobel prize. As Awrey (2012) writes, these theoretical propositions became

> *the central articles of faith of the ideology of modern finance: the foundations of a widely held belief in the self-correcting nature of markets and their consequent optimality as mechanisms for the allocation of society's resources.* (p. 237)

They suggest that minimal market institutions combined with rational market actors are sufficient to provide the twin "keys to paradise": direction and pace. This was to be a regulatory "end of history". The economist's role becomes that suggested by Hayek (1985): to defend the minimal regime against its detractors, especially against

the "rationalists" who would foolishly try to improve the unimprovable. The Red Queen dynamics discussed in this research suggest claims of a final and unbeatable regulatory solution will always prove short-lived. Nonetheless, this faith fuelled a widespread push for deregulation from the early 1980s onward.

Alongside basic financial services and intermediation, the key product that the financial sector sells is the amelioration of risk. This allows the expansion of credit. The greater the risk and the more it is reduced, the greater the returns (Haldane et al., 2010).[129] Soaring financial profits post-deregulation were interpreted as a "productivity miracle" (p. 87), with financial sector TFP growth in the UK some two to four times the rate of the general economy. In the US, finance grew from two to eight percent of GDP from 1950 to the present day. Apparent productivity growth reflected the apparent taming of riskier and riskier tranches of debt by bundling them into collateralised debt and loan obligations that had actuarially predictable returns. Mortgage-backed securities are the best-known example. These exotic derivatives were sold, repackaged, and on-sold again, through chains of buyers. They promised low-risk positive returns in ordinary states of the world, but the risk remained, "parked in the tail of the return distribution" (Haldane et al., 2010, p. 102). Insurers got in on the game, promising to insure clients against some of the tail-end risks. The resulting expansion of credit created further demand for these products, and the credit bubble expanded further still.

As Haldane et al. (2010, p. 87) write, the productivity "miracle" was a "mirage". In the terms of Box 11.1, we could term it the "mimicry" of productivity. What masqueraded as productivity growth was the parasitic transfer of wealth from society to the financial sector on an extraordinary scale. This was an inefficient transfer, to say the least. For every dollar the finance sector extracted before the crisis, the losses to society were orders of magnitude larger. While it is the bail-out packages that captured public attention, they amounted to a cost of around one percent of GDP. The main damage was the recession itself, with the Bank of England estimating long-term losses at "between $60 trillion and $200 trillion for the world economy", or

[129] If risky *future* returns can be misrepresented as safe, this drives *present* asset price inflation and raises apparent wealth. Apparent risk amelioration also generates apparent consumption smoothing, and so increases investors' appetite for risk. Together, these effects further increase the resources available for betting on uncertain future returns, creating more risk, and so creating further opportunities to mimic risk reduction.

between 90 and 350 percent of global GDP (Haldane, 2010, p. 4).[130] These estimates exclude the unquantifiable costs of political reactions in democratic polities, from trade wars to the many other inefficient populist policies, and the weakening of democratic capitalism in its contest with authoritarian capitalism.[131]

### 1.2 Failures to identify *c*

With a general poverty of information about these complex products sold through chains of buyers, their connection with underlying risk—and their actual private payoff *c*—became impossible for mere mortals to trace. Again, where *c* cannot be identified, even the most perfectly supplied market institutions become worthless or harmful.

Financial actors themselves were unable to gauge risk, as reflected in large losses and some bankruptcies. This is Keynes' story of animal spirits, reassured by the Great Moderation and roused by the unprecedented profits being made by their peers throughout the sector. Haldane et al. (2010) observe that the

> *[r]isk illusion is no accident; it is there by design. It is in bank managers' interest to make mirages seem like miracles.* (p. 106)

Privately, banking executives remained major net beneficiaries over the crisis period, and the trick of shifting risk into the tail-end of the distribution is worth repeating. We can expect continued innovation towards this end.

This sets up the Red Queen's race behind the crisis. Financial actors can only extract from debtors and creditors if they can misrepresent *c*. Clients, debtholders, and shareholders with rational expectations would possess unbiased evaluations of systemic risk, and exert appropriate market discipline. Together, this would align executive returns with social returns. This a gossamer-thin foundation for such a complex and welfare-critical area of economic activity, and if we bring in concrete market actor capacities, we can recognise it as a crisis-producing machine.

[130] The bulk of this cost comes from hysteresis effects, with the financial crisis impacting the growth pathway for years or decades.
[131] If a democratic future for the world translates into higher growth (as expected, e.g. Acemoglu et al., 2019), highly valued civil and political rights, and lower odds of destructive warfare (e.g. Ray, 1993), the losses from severe dysregulation are large indeed.

With respect to retail clients, again, only a modest proportion of ordinary individuals understand compounding interest. A much smaller fraction still understand even the meaning of "systemic risk", and very few of even the best-informed players *accurately* evaluated that risk. The standard heuristic is to evaluate assets on recent trends, which stimulates the formation of bubble-bust cycles (e.g. Clayton, 1997; Hommes et al., 2008; Greenwood & Nagel, 2009; and Eichholtz et al., 2014). Banking executives benefited from such bubbles and so from inflaming the weakest instincts of their clients. The most egregious subset of cases, involving "predatory lending" to vulnerable borrowers, is estimated to have raised default rates in the subprime class by at least a third (Agarwal et al., 2014, p. 29).

Even sophisticated clients may be outmatched in the innovation race. The other side of the predatory lending coin is the predatory sale of debt. In one of the most infamous cases, Goldman Sachs sold clients mortgage-backed securities designed by a hedge fund run by John Paulson, where Paulson then bet that the underlying mortgages would fail. This parasitic scheme *par excellence* transferred about US$1 billion from those clients to Paulson, with some share going to Goldman in fees (Freifeld, 2016).

Finally there are the shareholders and debtholders (or firm creditors holding bonds). As is widely understood, these stakeholders have distinct incentives, and in the case of shareholders in particular these incentives can be problematic. The matter of misaligned incentives goes beyond the problem of rationality asymmetries, and is a topic for Chapter 12 on the second leg of the Red Queen's relay. Such asymmetries, however, also played a significant role. With rational expectations, debtholders in particular would moderate bank risk-taking as they are most exposed to insolvency. Yet banks' balance sheets are unusually opaque, and both executives and shareholders may benefit by increasing bank complexity and so opacity; indeed in the pre-crisis period, Jones et al. (2012) find that banks innovated in the pursuit of opacity as a means of driving over-valuation (in keeping with theoretical propositions in Chapter 11, Section 2.1). The more opaque the bank, the more severe its price declines in the 2007-2008 crisis.

Each of these patterns has its parallels in the period leading up to the Wall Street Crash of 1929 and the Great Depression: extreme asymmetries in innovation

capacity, alongside opportunities to socialise risk, provided fertile ground for extractive banking innovations. Ferdinand Pecora, the chief legal counsel investigating the Wall Street Crash, exposed a smorgasbord of "predatory operations" used to extract wealth from the public—price manipulation, insider trading, aggressive marketing of high-risk securities, and active efforts to inflate profitable speculative bubbles (U.S. Senate, 1934, p. 68). History rhymes: two periods of "deregulated" markets allocated impossible tasks to ordinary market actors.

In the case of the GFC, the pivotal irrationality was that of the regulator, in its persistent assumption of perfect market actor rationality—in defiance not merely of common sense, but of centuries of financial history and contemporary empirical evidence. Brooksley Born, who became famous for having warned Greenspan, Summers, Rubin and Levitt of the risks of deregulating financial markets, recalled that Greenspan rejected her concerns like so:

> *He explained there wasn't a need for a law against fraud because if a floor broker was committing fraud, the customer would figure it out and stop doing business with him.* (see Roig-Franzia, 2009, para. 17)

If we design the total regulatory structure on the assumption that market actor D&E is perfect, then in *complex* markets, that structure *will* fail. I suggest the nature of the Red Queen's race means that this can be taken as an economic law.

# 11E2

## Empirical case 2:
## The Opioid Crisis and the first leg

The North American Opioid Crisis is another case in which the total regulatory structure allocated market actors impossible detection tasks. If market discipline is perfect, then existing healthcare institutions would be largely sufficient, and indeed "deregulation" could extend much further, as Friedman preferred. Rational patients would possess unbiased estimates of the payoffs of medical treatments and there would be no need to interfere—"the customer would figure it out". Rational doctors—if we can take them as well-motivated—would on average possess accurate priors about the merits of a drug, and the marketing agents of pharmaceutical firms could not systematically manipulate those priors. Even if doctors were self-interested and influenced by payments, this would matter little for patients with rational expectations; on average, they would accurately adjust for the probability that a prescribing doctor had been co-opted. Prices for doctor visits would in turn reflect the quality of the information they provide, and act as efficient signals of that quality.

This alien universe of rational agents bears no relation to empirical reality. Patients had very little, and easily biased, information about the net utility consequences of ingesting 6-deoxy-7,8-dihydro-14-hydroxy-3-O-methyl-6-oxomorphine, marketed as "OxyContin". They could not possibly identify $c$, but depended absolutely on trust in the expert judgement of doctors.

As the gatekeepers to prescription drugs, doctors were the natural targets for pharmaceutical firm strategies. Firm strategies have been two-pronged:

- first, firms co-opted doctors by means both subtle and overt; and
- second, firms created an information environment to mislead well-motivated doctors about the payoffs for their patients.

The first strategy exploited weakly regulated channels for directing payments to doctors who prescribed the drug, contributed to favourable research, or acted as spokespeople. In the US, medical payments from pharmaceutical and medical device firms are systemic; a majority of doctors, some 67 percent, receive such payments

(Inoue et al., 2019). In 2019, some 624,000 doctors received US$3.7 billion in "general payments" from pharmaceutical and medical device firms (Centers for Medicare & Medicaid Services, 2021). One in ten doctors receives more than 20 percent of their income from general payments alone (Gill et al., 2020). General payments are only one third of the total paid out, alongside another US$5.8 billion for research contributions and US$1.1 billion from shareholder returns distributed in 2019. Firms wine and dine doctors, pay for sham consulting, gift paid trips to Hawaii or the Olympics, and invite them to bolster their publication portfolio by adding their names to firms' research articles, to which those doctors have made no contribution (US Senate, 2007, p.172).

Such payments have been shown to reduce the general quality of prescribing, and increase the prescription of relevant drugs (Brax et al., 2017). A review of studies finds consistent evidence of temporal associations between payments and prescriptions, as well as a dose response such that more payments lead to more prescribing (Mitchell et al., 2020). Doctors report believing that their prescription practices are unaffected by payments, but the data shows that even negligible payments and gifts-in-kind affect behaviour—a common pattern thought to arise from innate human propensities for reciprocity (Elliott, 2014).

Incredibly, pharmaceutical firms are able to purchase data on physician prescribing behaviour, sold by firms such as IMS Health, which allow them to track particular doctors' responsiveness to payments (Yeh et al., 2016). As discussed in Chapter 11, these new "big data" strategies allow firms to fine-tune models of actor behaviour and allocate money accordingly. Neither doctors nor patients have fine-tuned models of pharmaceutical firm behaviour. Economists falling prey to the peculiar modern assumption that all innovation is benign have tended to assume that increased data availability and computing power will translate into greater efficiency.[132] Within a Red Queen's race, asymmetric advances in data and computing power will frequently *reduce* social efficiency.

[132] Mihet & Philippon (2019) make similar observations: *"the common belief is that the use of new technologies, such as AI, machine learning, and Big Data should have raised productivity over the last decade. In spite of this micro-level evidence, we do not see the impact of Big Data on aggregate productivity."* (p. 34) One explanation offered is greater market concentration. The Opioid Crisis is suggestive of increased asymmetries in rationality and information, which harms aggregate productivity.

Firms' second strategy was to exacerbate this asymmetry by reducing the quality of information available to doctors and patients. Firms concealed negative information about drug impacts, and used front organisations such as the American Pain Society and International Association for the Study of Pain to influence drafting of World Health Organization (WHO) guidelines. In 2020 the WHO guidelines were retracted, with a remarkable admission of undue influence (Dyer, 2020). Doctors in receipt were paid to assuage the concerns of their peers and present favourable case studies. "Independent" panels were set up to provide prescription guidelines, and "independent" patient groups paid to extol the virtues of the drug.[133] This is a coordinated strategy of what Sismondo (2021, p. 1) terms "epistemic corruption".

Doctors too are boundedly rational—or, in the terms of Chapter 11, they employ concrete detection strategies and have limited resources to invest in improving those strategies. They are time poor, and must keep on top of a changing landscape of thousands of drugs and changing practices. Just as patients have no choice but to trust in doctors, so doctors have no choice but to trust in these channels of information about drug impacts—research data, peer communications, patient reports, and so on—for their mutualistic contributions to patient health. These channels are frequently reliable and welfare-enhancing. The challenge for Purdue Pharma in promoting Oxycontin was to create a "mimic" that looked like the "model", and they rose to it.

The total economic cost of the Opioid Crisis is estimated at more than US$500 billion *annually* (Maclean et al., 2020), or around 3 percent of US GDP. For Purdue, Oxycontin generated an estimated $31 billion in revenue to 2017; its owners, the Sackler family, are one of the richest families in the United States. Clearly, with the Global Financial Crisis, Purdue's set of strategies lies to the extreme lefthand side of the social strategy space.

It is instructive that Purdue Pharma, like financial innovators, took such a strong interest in the concrete D&E capabilities of the relevant actors throughout the total regulatory structure. They did not assume that patients or doctors "would figure it

[133] Also see Elliott (2014) on sophisticated behavioural strategies employed to encourage over-prescribing of Neurotin "for everything".

out". While conventional economic assumptions suggest that there is no Red Queen's race between market actors, firms will busy themselves with winning it.

**Box 11E2.1 The rationality assumption: Allocating impossible tasks to patients**

As discussed in Chapter 10, Section 1.3, unshakeable commitment to the rational actor model produces predictable responses. Once a component of the D&E machine is assumed perfect, its role must always be maximised. The Opioid Crisis must, in such a world, result from "too many restrictions on prescribing, not too few". So argues the Cato Institute, in an article endorsing "outright legalization" of opioids (Miron et al., 2019, para. 3, §9). Given asymmetries in actor rationality, this would be a tantalising prospect for pharmaceutical companies and their investors, and a devastating one for consumers and the market society. Every incremental consumer added to the ranks of the addicted, whether by captured doctors or misinformation, becomes, like the smoker, a steady source of income.

Cato makes its argument by conflating two distinct problems: one of access to opioids for the non-addicted, and the other of access to opioids for those already addicted. The first problem requires reform and tightening of the institutional D&E structure that domesticates healthcare markets. The second requires expanding a public good, namely methadone clinics. Neither calls for further increasing the burdens on market actor D&E, which would be no less than the sacrifice of people and society at the altar of what was, for Pareto, an obviously erroneous expediency.

*To widen the market and to narrow the competition, is always the interest of the dealers…The proposal of any new law or regulation of commerce which comes from this order… comes from an order of men whose interest is never exactly the same with that of the public, who have generally an interest to deceive and even to oppress the public, and who accordingly have, upon many occasions, both deceived and oppressed it.*

Adam Smith (1776, p. 287)

*those exertions of the natural liberty of a few individuals, which might endanger the security of the whole of society, are, and ought to be, restrained by the laws of all governments… The obligation of building party walls in order to prevent the communication of fire, is a violation of natural liberty, exactly of the same kind with the regulations of the banking trade which are here proposed.*

Adam Smith (1776, p. 353)

# 12

# The second leg of the relay: The regulator/market actor Red Queen's race

Next we turn to the regulator's capabilities in the second leg of the relay: the innovation race between the regulator and market actors. So far I have examined the consequences of allocating detection tasks to market actors who cannot achieve them. Now the picture becomes more complex: we have an institutional regulator who may step in where market actors fail, but who also faces constraints to its own detection and enforcement (D&E) powers. The optimal division of D&E labour is co-determined by the absolute and comparative advantages of regulator and market actors—the capabilities of each must be combined into a working system.

The regulator needs to choose a total regulatory structure—an assignation of detection and enforcement tasks to institutional actors *and* market actors—that controls parasitism and allows mutualisms. Market actors, though they play a

regulatory role,[134] cannot choose this total regulatory structure. Running the second leg of the Red Queen's relay entails observing how this structure is performing, and adapting it as required.

Section 1 begins by introducing the second leg of the Red Queen's relay. There are two broad problems the regulator needs to solve:

1. first leg problems, where incentives are sound, but markets are too complex and capabilities too asymmetric for market actors to play their role; and
2. pure incentive failures, such as externalities and market power, where inefficiencies will arise even if actors are perfectly rational.

I examine some standard regulatory responses, all of which invite counterstrategies from firms.

Section 2 considers how the choice of regulatory strategy affects the intensity of the Red Queen's race. I introduce a spectrum of regulatory strategies, with "marginal" regulation at one end and "structural" at the other. Marginal regulation can be undertaken by market actors or the regulator, while only the regulator can structure markets.

- A dominantly "marginal" regulatory structure is designed to precisely carve out parasitisms while leaving mutualisms untouched. It maximises theoretical allocative efficiency, but its price is high complexity, which increases the returns to innovation capacity and intensifies the Red Queen's race.
- A dominantly "structural" regulatory approach employs broad rules and bright lines to rule out classes of transaction that are broadly parasitic. It sacrifices mutualisms and departs from theoretical allocative efficiency, but its advantage is that it simplifies regulatory problems and eases the Red Queen's race.

[134] Without commitment, it is rules—a system of detection and enforcement—that explain the systematic alignment of private and group payoffs, i.e. cooperation. Again, market actors are an element of the hybrid total regulatory structure. The structure relies on their use of detection (and to a lesser degree, enforcement) to identify privately optimal transactions, which under voluntary choice will optimise for the group.

At the extreme, both are destructive: excessive marginal regulation is too complex and guarantees loss of the Red Queen's race, while excessive structural constraints may win the race but unnecessarily sacrifice mutualisms. Efficiency lies in between.

Section 3 closes by observing that structural and marginal regulation are essential complements. The task for the institutional regulator is to simplify markets structurally, up to the point that marginal regulation (by both market actors and institutional actors) can do its work.

The argument is then applied in the empirical supplements that follow.

### 1.1 Two regulatory problems to solve

To keep up in the second leg of the Red Queen's relay, the regulator must solve two kinds of problems.

The first kind emerge from the first leg of the relay: costs that arise where incentives may be perfectly aligned (i.e. there are no externalities, market power, etc.), but market actors fail to optimise. Consumers have incentives to avoid the devastating costs associated with, for example, inappropriate drug sales or predatory lending, but lack the capacity to identify those costs. As noted in Chapter 11, the consequences may be private, economically systemic, and/or political.

The second type is a pure incentive problem, where there need not be any optimisation failures. More rationality does not help; the problem is that private and group payoffs are misaligned. Take, for example, incomplete property rights (i.e. the costs of theft), transactions that affect third parties (e.g. systemic risk, health effects, pollution), and strategies that obstruct market competition (e.g. market power, network effects) or exploit regulatory arbitrage (e.g. cross-border tax evasion).

### 1.2 Some regulatory responses to the between-market-actor race

There are two obvious institutional responses to "first leg" problems. The first is to reduce the size and impact of asymmetries in innovation capacity, so that market actors can accomplish the detection tasks allocated to them. Strategies include enforced simplification (e.g. restrictions on contract terms), supporting transactor rationality (e.g. education or mandatory information disclosures), or curbing

transactor advantages (e.g. restricting pharmaceutical firm access to information on doctor prescribing behaviour).

The second is to modify firms' incentives, in ways that block parasitic strategies or ensure that they bear some of the costs of those strategies. The regulator might restrict interactions between certain classes of actor (e.g. prevent pharmaceutical firms from paying doctors or funding "independent" third parties), price externalities created by the advantaged party (e.g. ensure the costs of systemic risk fall mostly on financial actors that engage in predatory lending), or provide the good or service publicly (i.e. eliminate the profit motive behind parasitism at the cost of reducing that for mutualisms).

Such institutional reforms provoke counterstrategies. Where the regulator attempts to address "first leg" problems, they become engaged in a Red Queen's race against the *advantaged* transactor, who gains from the status quo of exploitation (e.g. from predatory lending or excess opioid sales) and innovates to defend and extend this advantage. Firms will search for, and often find, new weaknesses in reformed policies (e.g. Duarte & Hastings, 2012). They have incentives to mislead the regulator about the private and group impacts of parasitic strategies—concealing harms, exaggerating benefits, arguing that market actors are rational and acting as moral champions of "choice", and so on. The regulator faces a perpetual detection task: discriminating genuine information about the public good from its mimics.

### 1.3 Some regulatory responses to incentive misalignments

In the case of pure incentive problems, the regulator's task is to modify incentives. Examples include Pigouvian taxation and subsidy (e.g. as reflected in carbon tax or trading policies), prohibition of classes of activity (e.g. involuntary transactions, as in the case of property rights, or asbestos mining in most advanced economies), or other constraints on economic strategies (e.g. regulation of the total content of financial firms' portfolios). Responses to market power typically include information-intensive strategies that seek evidence of anti-competitive strategies (e.g. investigations of Apple, Google, Intel, and other tech companies colluding on wages), approaches that sharpen competition and make parasitic coordination more difficult (e.g. restrictions on mergers and firm size), or public provision (e.g. of road, electricity, water, and gas supply). Responses to tax arbitrage may include investing

in a complex investigative apparatus, removing such opportunities altogether via tax uniformity (e.g. the same tax rates across jurisdictions), or simplifying detection of tax evasion (e.g. via cash-flow taxation, see Garnaut et al., 2020).

These strategies also provoke Red Queen's races. In the case of pure incentive problems, the introduction of a Pigouvian tax or subsidy triggers a contest over the identification of cases to which it applies. For example, many "artificial markets" for private delivery of public goods are exploited by gaming the rules that trigger subsidy payments (e.g. Thomas, 2007, on the Australia unemployment services market). Efforts to police collusion may trigger more subtle modes of collaboration that are difficult to prove (e.g. reductions in competition between firms owned by institutional investors). Closing arbitrage loopholes may shift activity towards new ones. Regulations intended to prevent another GFC, for example, may be out-smarted by clever new accounting tricks, partnerships, or by burying activities under overwhelming complexity. As Haldane et al. (2010) observe, many of the complex products linked to the GFC can be understood as innovations in regulatory evasion.

## 2 "Marginal" versus "structural" regulatory approaches

In the long-run, regulation must always adapt—the long-run replaces horses and carts with A380s, open outcry trading floors with electronic markets dominated by machine learning algorithms, and the quill and inkwell with Twitter bots programmed by foreign state actors. In the "medium-run", however, the choice of regulatory approach significantly determines the likelihood of keeping up in the Red Queen's race.

The task of the total regulatory structure—regulator and market actors—is to weed out parasitic strategies, while retaining the mutualistic. The question is how to identify them. As Bebchuk and Spamann (2009) put it:

> *regulation needs to rule out socially inefficient choices, but should not restrain socially efficient ones. Discriminating between the two is hard.* (p. 39)

The difficulty with discriminating between them is that all transactions are *sui generis* to some degree. They can be placed in categories of varying specificity, for the purpose of giving different types of transactions different treatment, ranging

from varying rules slightly from one sub-category to another, to allowing one and outright prohibiting the other. However, such categories are always incompletely descriptive and incompletely predictive of transaction payoffs.

The regulatory question, then, is how fine-grained or rough-grained the separation of mutualistic and parasitic strategies ought to be (or indeed, can be). The finer the grain, the more theoretically efficient, but the harder and more complex the detection tasks—and the greater opportunity for firms and other actors to exploit that complexity. Note that unlike in Bebchuk and Spamann (2009), here I recognise that the task of discriminating harmful and beneficial transactions is not just one for the institutional regulator, but it may also be given to market actors. But more on this in a moment.

To generalise, there are fine-grained regulatory approaches that are *information and rationality-intensive*, and there are rough-grained approaches that are *information and rationality-saving*. I will call these "marginal" and "structural" strategies respectively, for reasons that will become clear shortly. They are poles on a spectrum, with many shades of grey between.

### 2.1 Marginal regulation: Theoretical efficiency and intense innovation races

"Marginal" regulation aims to restrain parasitism via precise, context-sensitive evaluations of the social costs/benefits of particular market actor strategies. Its driving ideal is allocative efficiency: at the extreme, it is a regulatory scalpel that perfectly carves out all welfare-reducing transactions ($B_g < 0$) while leaving the welfare-enhancing ($B_g > 0$) untouched. I call it "marginal" because, at this extreme, it requires some means of identifying the effects of an additional unit of economic activity at the margin.

There are two sources of marginal regulation:

- *market actors*,[135] who evaluate the payoff $c$ associated with available transactions. If perfectly rational and transacting in ideal markets, they behave as a $B_g$-optimising scalpel; and

[135] As noted at the start of this chapter, market actors also play a regulatory role.

- the *state regulator*, who may evaluate $B_g$ directly and determine whether transactions should proceed or be blocked. The ideal central planner similarly wields the scalpel with perfection, as a perfect substitute for market actors. Or a less interventionist state regulator may only step in where market actors fail.

If either market actor or regulator capabilities are infinite, there is no need for broad rules that block certain classes of economic activity—e.g. certain product, service, or contract designs—as this risks needlessly blocking some subset that is mutualistic. The regulatory approach can be *maximally permissive*. If capabilities are insufficient, then permissiveness allows firms to profit from a wide variety of parasitic products that mimic value; firms have incentives to create complexity that consumers cannot handle, and design Gordian knots that the regulator cannot untie. In short, marginal regulation increases the returns to innovation capacity, and so intensifies the Red Queen's race.[136]

The empirical supplements following this chapter will examine the state regulator's turn towards marginal regulatory strategies of both kinds in financial and healthcare markets, which presents all members of the total regulatory structure with difficult races to run.

### 2.1.1 Bloodhounds chasing greyhounds

Firms not only have advantages in innovation capacity over consumers, but also, in aggregate, over the state regulator. This is a consequence of the same forces that make markets desirable as drivers of mutualistic innovation: markets are massively-parallel innovation systems that utilise distributed knowledge and strong incentives for profit-maximisation (see Chapter 6, Section 2). Firms' innovation advantage is a double-edged sword: it generates both mutualistic *and* parasitic innovations that the regulator, and armchair theorist, cannot foresee. As a consequence, the regulator, theorist, and indeed consumer frequently learn about parasitism only after the fact, once crisis—personal or systemic—has arrived, and progress often rests on cycles of crisis and (hopefully rational) reform.

[136] Another drawback, specific to the regulator's use of marginal regulation, is that it tends to rely greatly on discretion. This is a matter for the third leg.

In the arena of financial regulation, this is the problem that John Heimann (2000, p. 332), Comptroller of the Currency under President Carter, called "bloodhounds chasing greyhounds": fast-running firms must be sniffed out by a slower regulator. Hu (1993, p. 1463) writes that "[o]bservers agree that regulators know less than the bankers, and that they know too little", and Fisher (2013) that regulators are always

> *at least one step behind the actions taken by market participants. The more complex the rules, the more difficult it is to bridge the gap.* (p. 8)

### 2.2 Structural regulation: Pragmatic efficiency via a slowed innovation race

Regulation toward the "structural" pole controls parasitism by using broad and bright lines to limit the acceptable categories of economic activity. Payoff evaluation is rough-grained: the goal is to identify and close off classes of economic activity where the risks of parasitism outweigh the benefits of mutualisms. Structural rules may control contract terms, product types, market actor relationships, rates of return, information availability, maximum market shares, and so on. Structural regulation is generally the domain of the institutional regulator, given market actors cannot coordinate to impose structure on markets.

By making imprecise cuts, some subset of mutualistic transactions is lost and theoretical efficiency is sacrificed. The purpose of structural regulation is to simplify the regulatory problem, which in complex markets makes regulation—and so actual efficiency—possible.

Extreme structural regulation is no less absurd than extreme marginal regulation. One could structurally prohibit all economic activity, winning a victory over parasitism at the cost of every possible mutualism. More reasonably, prohibiting certain classes of financial trades (e.g. re-securitisation) may remove a mutualistic subset that is worth millions or billions in social returns, while reliably removing a parasitic subset worth trillions in social losses from systemic financial risk. Padoa-Shioppa (2004) examines this problem in finance specifically, where he notes a trade-off between the maintenance of free contracting and the management of systemic risk.

### 2.2.1 Boxing in the greyhounds

As discussed in Chapter 11, in a *simple* and constrained race (as in tic-tac-toe), investments in asymmetric innovation capacity proffer no advantage. To extend Heimann's metaphor, the institutional bloodhounds have the power to choose the topography of the landscape in which they pursue the greyhounds. If the race takes place in a structureless landscape, a multi-dimensional open field, then the bloodhounds are doomed to lose the scent. If the bloodhounds can control the structure of the landscape, they may take on the simpler task of a watchdog, guarding the narrow passes between acceptable and unacceptable activities.

## 2.3 Some examples

Again, with zero-cost and infinite regulatory capacity, structurelessness and permissivity maximises allocative efficiency. In practice, especially in contemporary economies, it is grossly inefficient.

Take minerals extraction for example. The standard, structural approach is for governments to allocate monopoly rights over the resource. It aims to avoid the inefficiency of the open-slather "goldrush" model, where many small actors race to extract the resource; these small-scale actors have lower incentives for long-term investment in efficient extraction, and this promotes wastage (Garnaut & Clunies-Ross, 1983). The regulator could aim for a more theoretically efficient "marginal" approach that is permissive, allowing the goldrush, but taxing or subsidising each actor's idiosyncratic extraction strategy according to its specific effects on the common resource. In practice, this would require such a complex regulatory apparatus that its costs would outweigh any efficiency gains, and many inefficient strategies would slip through the cracks; it would be worse than open-slather, and far worse than the regulated monopoly.

Intellectual property regulation similarly also involves the structural imposition of a temporary monopoly, which aims to endow the innovator with a stream of quasi-rents.[137] The period of protection is standardised to some number of years for a particular class of patent. Sometimes that period is excessive and the innovator

[137] Quasi-rents are returns excess to the marginal costs of production that are needed to justify an initial investment. They are not true rents, and are both equitable and efficient.

extracts genuine, and inefficient, rents. Sometimes the period is too short, providing quasi-rents that are insufficient to promote socially productive innovation. Vastly more efficient schemes would be possible with perfect information: the marginal regulator could ensure that each innovation was rewarded with precisely the quasi-rent necessary to justify the initial investment, and could eliminate all up-front capital costs that hinder smaller scale innovators. In practice, this would require a D&E structure that was vastly more sophisticated, with greater knowledge of the future, than the R&D structure it would be observing.

Even standard property rights are a form of structural simplification. As Posner and Weyl (2018) observe, ownership is a form of monopoly and can introduce some of the usual kinds of theoretical inefficiencies. They endorse reforms that would shift property rights towards the marginal pole of regulation. In doing so, they make all markets hyper-complex, greatly increasing the returns to asymmetric advantages in innovation capacity, and producing obvious opportunities for parasitism (see A.12.1 a more detailed discussion). Their proposal is the *reductio ad absurdum* pursuit of allocative efficiency, under the theory of infinite market actor rationality. It is rather like proposing, in an already inefficient planned economy, to increase efficiency by shifting towards more planning of everything—the mathematics may work on paper, but its results depend on whether the necessary capacities exist. Notably, the term "rationality" does not appear in the paper. To borrow Hayek's (1945, p. 530) terms again, it "systematically leaves out what is our main task to explain".

## 3 Concluding: Structural and marginal regulation are essential complements

To summarise, there is a trade-off between theoretical allocative efficiency and regulatory complexity. The extremes of structural and marginal regulation are destructive in opposite ways: the former blocks too many mutualisms, and the latter unleashes parasitisms. Efficient regulation lies in the middle ground.

More precisely, structural and marginal regulation are *essential complements*. We can deepen the discussion at the outset of Part 3 about the problem of designing an effective division of D&E labour.

- The comparative advantage of the state regulator is in structuring markets.

- The comparative advantage of market actors is in the use of distributed knowledge in marginal regulation, so long as they can handle the detection task (i.e. there are no problematic asymmetries in complex markets).

Optimal regulation integrates these capacities into a working system: it *structures markets so that detection tasks are simple enough for marginal regulation to succeed*.[138] In markets for private goods, voluntary exchange can then work its magic. In cases of public goods and externalities, effective state intervention becomes feasible. This reflects the broader observation that neoclassical theory and theories of actor capabilities are essential complements. Models of efficiency are only useful if integrated with realistic models of how we can get there.

Today, democratic capitalism suffers from a syndrome of economic mismanagement that has, as its base, emphases on allocative efficiency, benign innovation, and extreme market actor powers.[139] Each of these presses us towards an excessively permissive and marginal regulatory regime, unwittingly creating complex markets and intense Red Queen's races. Such races are guaranteed to be lost. Parasitism flourishes. The empirical supplements that follow return to the GFC and Opioid Crisis, to explore how the outrunning of consumers, and of regulators, emerges as a consequence of policy choices—of the failure to run the "second leg". At its core is a newfound reluctance to exercise the comparative advantage of the regulator.

[138] At the optimum, the additional mutualistic trades allowed by more "marginal" regulation must be sizeable enough to justify the additional detection costs, and the risks of parasitism where detection falls short.

[139] This is related to Ferri & Minsky's (1992) discussion of trade-offs in pursuing the allocative ideal:

> *Economists are given to talking about efficiency, and in the models of the invisible hand tradition, efficiency means allocative efficiency. But in a dynamic view of the economy a variety of efficiencies can be defined. Improvement in one 'efficiency' can lead to a deterioration in another. All too often the 'room for improvement' will be along 'one' of the efficiency dimensions…* (p. 23)

# 12E1

# Empirical case 1:
# The Global Financial Crisis and the second leg

The GFC is a story of unmanaged complexity. In conventional terms, it followed a period of financial "deregulation", a reduction in regulation. In fact, the regulatory burden was shifted to market actors, and this entailed an *increase* in the complexity of the detection tasks handled by the total regulatory structure. In an Arrow-Debreu world, and the world described by the various Nobel prize winners relating to the efficient markets hypothesis, these changes would improve allocative efficiency. In practice it exploded the returns to firms that could outsmart their transaction partners by investing in superior innovation capacity.

## 1.1 Financial regulation and complexity rents

The contours of this story, and its longer history, can be glimpsed in a fascinating study by Philippon & Reshef (2012). They examine the close relationship between the intensity of financial regulation, the skill-intensity of banking operations, and wage premia in the financial sector from 1909 to 2006 (Figure 12E1.1). In "deregulated" periods, i.e. periods dominated by marginal regulation by market actors, demand for skill sharply increased, these clever new hires devised novel strategies for generating *apparent* risk reduction, and the sector won lucrative returns. In the structurally-constrained Glass-Steagall period, wage premia evaporated and banking became relatively low-skill. As Krugman (2009, title) puts it, banking became "boring", and boring banking meant stability and growth.

Apparent reductions in risk generated not only high financial wage premia, but also large increases in the overall size and takings of the sector. Financial sector profits in the nineteenth century were on the order of 1.5 percent of GDP, before peaking at 6 percent of GDP before the Great Depression. Post-war profits fell to around 2 percent of GDP, before quadrupling to 8 percent before the GFC (Philippon, 2015).[140] In the UK, the balance sheet of the banking system relative to GDP increased 15-fold from

[140] This is also reflected in the share of all profits. In the post-war decades, the US financial sector reaped between 10 and 15 percent of all corporate profits. Prior to the GFC, the share in profits reached as high as 40 percent.

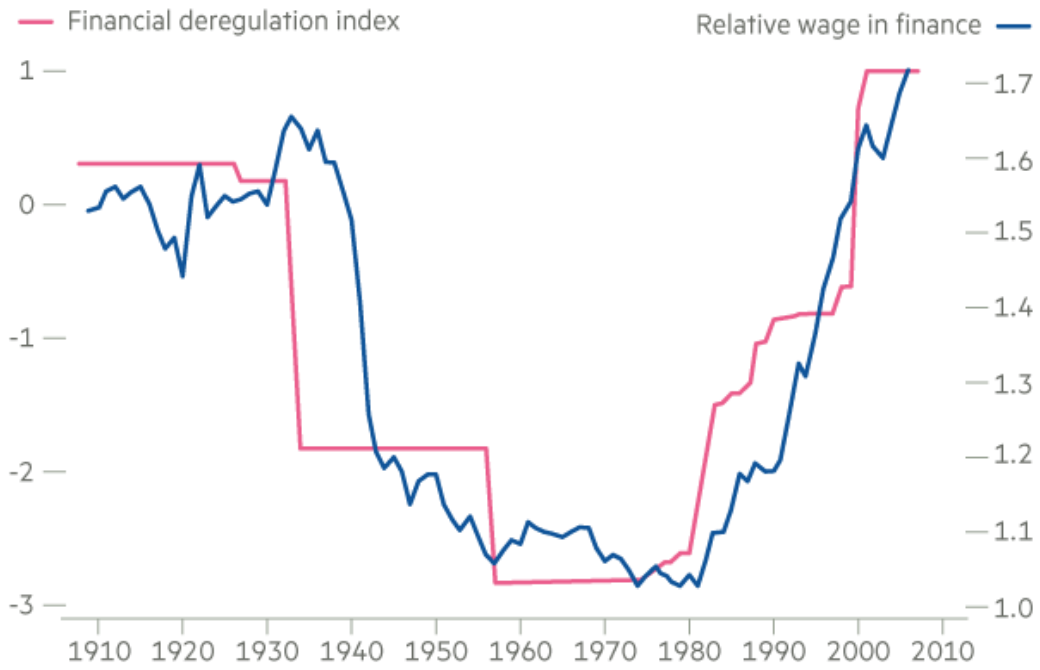



**Figure 12E1.1** Financial "deregulation", i.e. the shift of regulatory tasks to market actors, alongside relative wages in the financial sector.

1964 to 2007, from 34 to 500 percent of GDP (Turner, 2010). Similar patterns can be observed in all advanced economies.

Prior to the GFC, the dominant view was that financial innovations were driving a productivity explosion, and that growth in complexity was social beneficial.[141] In a field dominated by a model of innovation as benign—in Chapter 5, the "economics of the mutualism quadrant"—this was the obvious interpretation. "Deregulation" only allowed more mutualisms. Clever people profited by serving others and society in new, creative ways.[142] Growth in wage premia and sector returns is, of course, ambiguous, and Baumol's (1990) problem applies: the entrepreneur may allocate their efforts to productive or parasitic activities. Paul Volcker famously stated in 2009 that the only financial innovation of clear social value in the twenty years prior was the automatic teller machine. His broader comments after the GFC are worth revisiting:

[141] Turner (2010, footnote 20): "In the pre-crisis years", more leverage, or "more lending on any given level of bank capital", was "perceived as not only a rational private objective for individual banks, but as a valuable social objective."
[142] E.g. for Friedman (1953), speculation, as an activity of rational agents in competitive markets, is interpreted as a stabilising force that pushes prices back to fundamentals.

> *I wish that somebody would give me some shred of neutral evidence about the relationship between financial innovation recently and the growth of the economy, just one shred of information… A few years ago… I found myself sitting next to one of the inventors of financial engineering… he had won a Nobel Prize, and I nudged him and asked what all the financial engineering does for the economy… Much to my surprise he leaned over and whispered in my ear that it does nothing. I asked him what it did do, and he said that it moves around the rents in the financial system and besides that it was a lot of intellectual fun.* (in Murray, 2009, para. 11)

Figure 12E1.1, from the work of Phillipon & Reshef (2012), can be read as a history of this "intellectual fun".

Micro and macroeconomic patterns point clearly to parasitism. We discussed the evidence for "predatory operations" uncovered by Ferdinand Pecora in case of the Great Depression, and the same in the case of the GFC, in the Chapter 11 supplement. On the macro level, the period of low-skill, "boring" banking was characterised by exceptional GDP growth, strong investment, and financial stability (e.g. Philippon, 2008). The two periods of high-skill, "fun" banking were characterised by tepid GDP growth and weak investment, and both culminated in a severe, global financial crisis. An emerging literature of cross-national studies firms up this story. Schularick & Taylor's (2012) magisterial overview of the financial sector from 1870 to 2008 finds little evidence for benefits from the previous three or four decades of financial deepening. Going further, a suite of studies find an inverted U-shaped relationship between financial development and growth. Cecchetti & Kharroubi (2012) find that financial development is "good only up to a point, after which it becomes a drag", and "[t]he faster the financial sector grows, the slower the economy as a whole grows" (p. 12). Law & Singh (2014), Samargandi et al. (2015), Ductor & Grechnya (2015), and Purewal & Haini (2022) find similar results.[143] Arcand et al. (2015) find that the financial sector becomes a net drag when credit to the private sector exceeds 100

[143] See also Deidda & Fatouh (2002), Rioja & Valev (2004), and Aghion et al. (2005) for earlier work on non-linearities and diminishing returns to financial development. Demetriades & Rousseau (2016) find that the sector's contribution to growth is determined not by financial depth, but by the quality of the regulatory regime.

percent of GDP (also see Demetriades et al. 2016). The US reached around 220 percent in 2022 (OECD, 2023).

See A.12E.1 for evidence that traditional intermediation activities remain productive, although a small minority of current banking activity.

### 1.2 Moral hazard: Managerial and shareholder incentives over risk

A subset of these parasitic innovations entail the mimicry of risk reduction. There are other important modes of parasitism, but given space limitations and their particular relevance to the GFC, these are our focus. The strategy of pushing risk into the tail-end of the distribution depends on asymmetries in the first leg of the Red Queen's relay, and misaligned incentives with respect to risk. There are (A) sophisticated controlling parties, including shareholders but especially managers, that are exposed to the upside of risks more than the downside, and (B) unsophisticated clients (and taxpayers) who carry the downside.

Moral hazards—incentives to take risks known to be excessive—in banking are well understood. The downside risks of managers and shareholders are constrained in part by some policy choices. One is explicit (deposit insurance) or implicit ("too-big-to-fail") guarantees, which arise from state regulations that are in turn designed to prevent other crises (Box 12E1.1).

**Box 12E1.1 State guarantees: A necessity and a risk amplifier**

Moral hazards in banking partly result from explicit and implicit state guarantees that are essential for preventing bank runs, the latter having intolerable effects on financial stability and the democratic legitimacy of the market system. Bank runs only exist because the market as a whole can be profoundly "surprised" by emerging information about underlying asset values—because in complex markets, actors cannot identify $c$. While state guarantees address the problems created by irrational expectations and herding behaviour, they also protect people from downside risk. Accordingly, risk appetite increases. A complete regulatory system must employ further mechanisms to reduce risk, and this is the subject of much of the rest of this empirical supplement.

Another is limited liability, which gives shareholders full exposure to the upside but limited exposure to the downside (they may lose their original equity but no more). Debtholders, meanwhile, have no upside gains but share in downside losses. The consequences are illustrated in Figure 12E1.1, modified from Goodhart & Lastra (2020) and Miller (2021), in a graph of the same form as the Social Strategy Space of Chapter 3. A bank portfolio has expected social returns of C. Expected shareholder returns are C*. That is, the shareholder (and manager) can benefit from parasitic investment strategies.

The tension between shareholder and debtholder incentives was first laid out in Jensen and Meckling (1976), and has been explored in a large theoretical and empirical literature (e.g. Landier et al. 2011; Becker & Ivashina 2015, Valckx et al., 2016, Drechsler et al. 2016, Plosser & Santos 2014; Hackbarth et al. 2021). While shareholders are commonly thought to discipline managerial risk-taking, the empirical literature suggests that greater shareholder influence leads to greater risk-

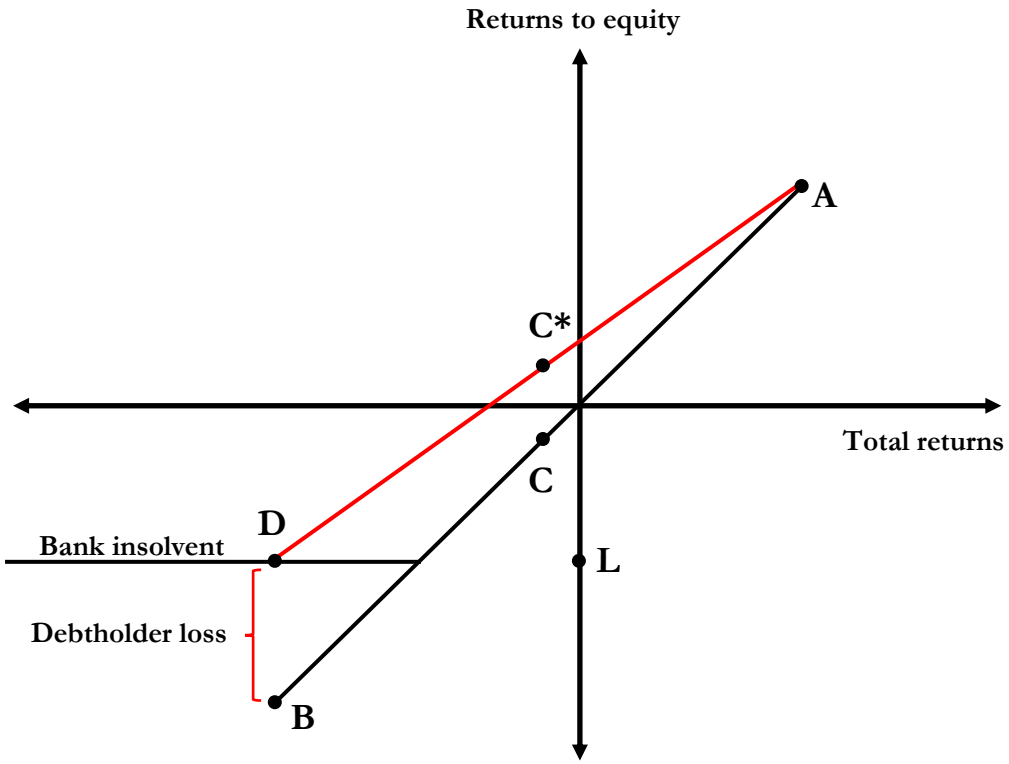



**Figure 12E1.2** A portfolio is equally likely to deliver either returns A or B, so that C is the expected return. Because C is negative, this portfolio is socially harmful. However, limited liability (L) and the sharing of losses with debtholders means that shareholder returns lie on the red line, with the unlucky state of the world represented in D. The expected return to shareholders, C*, is positive.

seeking and generated worse results during the crisis (e.g. see Saunders, Strock & Travlos, 1990; Bolton et al., 2006; Chen et al. 2006; Laeven & Levine, 2009; Beltratti et al., 2011; Fahlenbrach & Stultz, 2011; and Falato & Scharfstein, 2016).

Managerial incentives make things worse still. Payment of managers in stock options gives them a highly leveraged share in the upside, exaggerating these incentives (Bebchuk & Spamann, 2009, Garnaut, 2009). Further, managerial job security, and the tuning of pay-for-performance remuneration schemes, creates a high-stakes competition to keep up with the top-performing—and riskiest—firms (e.g. Hubbard & Palia, 1995, Crawford et al. 1995, Beck et al., 2013, and Aikman, Nelson & Tanaka, 2015). A reliable way to keep up in the competition is to herd around the same assets as peers, but this increases the odds of synchronised bank failures (Lakonishok et al., 1992; Devenow & Welch, 1996; Avery & Zemsky, 1998; Bikhchandani & Sharma, 2000).

**2 The structural turn: Glass-Steagall**

The strategy of relying on "marginal" regulation of banks by clients, shareholders, and debtholders—of relying on market discipline—has been a failure. The best-informed actors have incentives to pursue a parasitic overall portfolio; they simply need outsmart debtholders. Similar patterns prior to the Great Depression gave rise to potent structural regulations.

An important question, as Robert Lucas Jr. (2013) observed, is why Glass-Steagall was so successful in preventing systemic banking crises from 1933 to 1999, despite the presence of deposit insurance and limited liability, all with relatively little regulatory effort. Like Lucas, I do not suggest that Glass-Steagall is an "unbeatable" regulatory strategy. History does not end, and the post-1933 regulatory structure would need some reform after nearly a century of strategic change.[144] Nonetheless, the success of Glass-Steagall rests on underlying principles discussed in this thesis that are of enduring significance.

Glass-Steagall is best-known for introducing a legal distinction between commercial and investment banking. This can be understood as the recognition of banking as

[144] New realities include greater international competition, changed ownership structures, and the rise of shadow banking, fintech, and digital currencies.

providing distinct public and private goods. The function of commercial banking was to provide a set of public goods: broadly available financial services, trust in deposit security to prevent bank runs, and returns to deposits at around prevailing interest rates. As in Hardin's classic tragedy, and as discussed in previous chapters,[145] competition can drive the overexploitation of a common resource. In this case, competing banks race to attract depositors by offering higher interest returns at the cost of increased systemic risk. Depositors eagerly seek those returns—being insulated from downside risks by government insurance and in any case unable to interrogate banks' financial positions. To constrain bank competition, Regulation Q—Section 11 of Glass-Steagall—limited returns to deposits roughly in line with prevailing interest rates (Gilbert, 1986). This meant that commercial banks could only compete on factors such as the convenience of their services. Together, these restrictions made commercial banks *specialists in the provision of the public good of safe and stable returns, and the private good of convenient banking, for depositors*, and cut off their innovative options for doing otherwise.[146]

To use the metaphor adopted in Chapter 12, structural regulation simplified the topography of the landscape in which the bloodhounds chased the greyhounds. The bloodhounds only had to guard the boundary between commercial and investment banking, and watch over the interest rates offered to depositors. We can begin to see why this system stayed effective for so long: it presented the regulator, and market actors, with Red Queen's races that were simple to run. More precisely, it gated a complex race, over the policing of systemically-risky strategies, behind a much simpler race.

The function of investment banks was to fund riskier activity, ideally on the basis of particular, distributed knowledge that justified those risks. Importantly, prior to 1980, investment banks were primarily run as long-term *partnerships* where managers had considerable skin in the game (Hill & Painter, 2010; also see Etsy, 1998, Grossman, 2001, and Koudijs et al. 2018 on the historical effects of expanded

[145] Chapter 4 described examples in sociobiology (e.g. competition between symbionts increases virulence), firms (e.g. between-employee competition can produce group-harmful strategies, like undermining of fellow employees), and political systems (e.g. competition between would-be Olsonian stationary bandits can lead to destructive warfare, a problem Hobbes wanted to solve with the Leviathan), and Chapter 12 discussed it in relation to minerals extraction.
[146] Smith (1776/1976) writes favourably of the Bank of Amsterdam, for which "public utility… and not revenue, was the original object", to "relieve merchants" of "inconvenience" (p. 488).

liability). The absence of safety net for investment banks provided a spur to prudence, and the McFadden Act sharpened this spur by restricting bank size so that none could be "too-big-to-fail". Together, this translated into *structural commitment* ($r_s \gg 0$) for managers in investment banking. Here, the bloodhounds were little needed, as such commitments are self-policing.

This was a governable financial system. Regulatory errors, and the continuation of the Red Queen's race at its edges, led to its failure.

1. A change in the implementation of Regulation Q from 1966 meant that deposit rates were no longer indexed with, and fell far behind, real interest rates (Gilbert, 1986; Lucas, 2013). As the gap between deposit returns and interest rates grew, it generated enormous appetite for alternatives to commercial banks. It caused a rush of depositors into money market funds, including savings and loan (S&L) associations, in the 1970s. This was the beginnings of *de facto* liberalisation. Predictably, intensive competition between S&Ls to attract depositors led to increased risk-seeking, and laid the foundations for the S&L crisis in 1986.[147]
2. The decline of the private partnership model of investment banking, and the shift towards managers operating as agents of shareholders. This substantially increased managers' risk appetite. It can partly be understood as the disappearance of $r_s$-induced lockstep payoff transformation, and a turn towards control of managers via exploitable $c$-modifying rules: wages, bonuses, and options. The latter were subject to counter-innovations, in managers' myriad strategies for reaping returns while making unsound investments.
3. Finally, the merger of commercial and investment banking allowed moral hazard and competitive pressures to merge into what Martin Wolf (2010) described as a risk-generating "doomsday machine". Banks competed to attract depositors with higher interest rates, depositor money could be used for high-risk investments to generate the necessary yield, and managers controlled by exploitable rules rather than $r_s > 0$ could reap vast returns over the years that systemic risk accumulated.

[147] A third of S&Ls failed by 1995.

The bloodhounds were sent to their kennels, and the greyhounds were set loose in the belief that almost all innovation was good innovation. Market actors were locked in a complex Red Queen's innovation race, and allocated detection tasks they could not possibly achieve. This permissive, intensely marginal regulatory regime was tuned for extreme parasitism, and we know how the rest of the story goes.

## 3 The marginal-institutional turn: Dodd-Frank and Basel III

What of post-GFC regulation in Dodd-Frank and Basel III? Undoubtedly they bring improvements. However, from the post-Glass Steagall era, they retain a primary emphasis on permissive marginal regulation, and so the acceptance of great complexity. The nature of that marginal regime has been tweaked, with a greater role for institutional regulation—but the Red Queen's race remains intense. It is a gamble that the regulatory apparatus will be able, and willing, to keep up for decades to come, even as the memories of the crisis fade.

Consider Dodd-Frank's "Volcker Rule", which prohibits proprietary trading[148] and has been likened to an updated form of Glass-Steagall. If so, it is its highly "marginal" and unenforceable cousin. The definition of proprietary trading is subject to a variety of exceptions that can allow almost any trade, if the *motivation* of that trade is acceptable.[149] Even with access to deep distributed knowledge, motives are nearly impossible for the bloodhounds to sniff out. The only penalty for breach of the Volcker Rule, by Deutsche Bank in 2017, was self-reported (McLannahan & Dye, 2017),[150] and the rule is now considered more or less defunct.

The heavy-lifting of Basel III is done by tighter capital requirements and new asset risk-weightings. These are potentially protective, if potent and simple enough to enforce. There is cause for concern on both counts. As Wolf (2010) observes, to

[148] Proprietary trading occurs where banks use their own balance sheets to trade financial assets, rather than depositors' funds. It can be used to extract rents by front-running depositors and other clients, and has been blamed for increasing risk prior to the GFC.
[149] Rather like the religious inquisitions of history, it lends itself to gross injustice, or to impotence, depending on whether enforcers respond aggressively or conservatively to the impossibility of directly observing the thoughts and motivations of others.
[150] Meanwhile, large and risky trades persist: in 2016, Credit Suisse reported $1 billion loss on securitised pools of risky loans (Voegeli, 2016), and in 2018 Deustche Bank reported single-day losses on its trading account some 12 times larger than the bank's calculations of what it could lose in a single day (Nasiripour et al., 2018).

> *triple the size of capital reserves… sounds tough, but only if one fails to realise that tripling almost nothing does not give one very much.* (para. 2)

New capital requirements tighten the leverage ratio from 1 percent up to 3 percent.[151] This remains far below levels favoured by economic research, which has generally pointed towards minimum values of 15-20 percent. The gap is supposed to be filled by substitutes for equity—but these substitutes are complex, untested, and barely understood by the market actors that purchase them. Efforts to calculate risk exposure, meanwhile, have become increasingly complex over time. The number of risk categories increased from seven in Basel I to over 200,000 in Basel II, and Basel III is more complex still. It "provides near-limitless scope for arbitrage." (Haldane & Madouros, 2012, p. 8). See A.12E.2 for further comments and research on these and related matters.

Finally, the "marginal" nature of this regime is partly indicated by the complexity of the regulatory scalpel. Glass-Steagall and McFadden were 37 and 11 pages long respectively. The Dodd-Frank Act is some 2,300 pages with more than 22,000 pages of detailed rule releases, and Basel III has 616 pages just on how banks must calculate their capital requirements. Enforcement will necessarily be piecemeal and discretionary. Barth et al. (2007) noted with respect to the much simpler Basel II, "[m]ost supervisory agencies will never have sufficient human capital or budgets to implement Basel II successfully" (p. 227). Romer (2012) writes that these complex regimes of rules for the financial sector

> *will never keep up. The technology is evolving too quickly. The scale of the markets is enormous and continues to grow.* (p. 123)

The answer suggested by the Chapter 12 analysis is stronger structural regulation—to structure these markets until the regulatory tasks are manageable. Haldane and Madouros' (2012) closing remarks are pertinent:

[151] This is the ratio to Tier 1 capital at least. Even this raises problems, as Admati (2014) observes: Tier 1 capital includes sovereign debt, and allowed risk-seeking banks to turn to Greek bonds while complying with these restrictions.

> *you do not fight complexity with complexity. Because complexity generates uncertainty, not risk, it requires a regulatory response grounded in simplicity*. (p. 19)

For similar arguments, see Rajan (2006), Geithner (2017), Mirowski (2013), Aikman et al. (2021), Wolf (2014), Moschella & Tsingou (2013), and King (2009).[152]

### 3.1 Strange bedfellows: The permissive financial regulator and central planner

The nature of this regulatory regime is paradoxical. On the one hand, it aims to remain as pro-market and permissive as possible. As Palley (2008) remarks, financial markets are still "assessed in terms of the neo-classical allocative efficiency paradigm", where there is a persistent fear of "artificially" interfering with "natural" market activity (p. 5).[153] Yet regulators also recognise that they must prevent crisis. Having ruled out major structural reforms and committed to permissivity, the only remaining option is a hyper-complex marginal apparatus. Though such a regulator may conceive of itself as pro-market, *it edges close towards the behaviour of a central planner*. It must centralise and process incredible amounts of distributed knowledge, in an attempt to act as a substitute for market actors that cannot detect *c*. The task is in some ways harder, given the firms it seeks to control have enormous incentives, and capacities, to innovate ways of escaping it.

The stakes are high. It is unclear whether democratic capitalism can, at this stage in history, survive another major crisis. It seems prudent to avoid creating one of history's most complex Red Queen's races, where we race against among history's most sophisticated firms, and where the members of those firms are rewarded handsomely on every occasion they outsmart us. Glass-Steagall slowed the race to a

[152] King (2009, p. 7) warns of the "sheer creative imagination of the financial sector in dreaming up new ways of taking risks"… to belief that permissive regulation "can ensure that speculative activities do not result in failures is a delusion."
[153] Romer (2012) observes that in aviation, the burden of proof is on innovators; in finance, the burden of proof remains squarely on the regulator. Bell & Hindmoor (2014) write that policymakers have "persisted with a mindset that still values the economic centrality of a large and complex banking sector." And, "in accepting a large, complex and constantly evolving financial system with high levels of systemic risk, they have unwittingly placed themselves at a continuing disadvantage in the regulatory arena" (p. 342).

crawl by structuring incentives, and it served democratic capitalism well for nearly 70 years. It might, therefore, be taken as a good example of pro-market regulation.

## 4 Raising managerial $r_s$: Innovation-resistant financial regulation?

As a final aside, the framework of this thesis indeed suggests a kind of regulatory strategy that structures incentives: one that raises managerial $r_s$, or skin-in-the-game, so that managers have a lockstep, largely innovation-proof incentive to maximise social returns.[154] It is discussed in brief here, and in more detail in A.12E1.3. The logic of the approach emerged as I developed the framework, though I later found it much more richly and capably developed by Bebchuk & Spamann (2009). Their work is also an extension of the literature examining the merits of paying managers in inside debt and via long-term pensions (e.g. Sundaram & Yermack, 2007 and Bennett et al., 2015).

Bebchuk & Spamann (2009) argue that managerial remuneration should be tied to a "broad basket of securities" so that managers are sensitive to the returns to *all* stakeholders—shareholders, debtholders, and taxpayers. Ideally this would comprise

> *a given percentage of the aggregate value of the bank's common shares, preferred shares, and bonds … minus any payments made by the government to the bank's depositors, as well as other payments made by the government in support of the bank, during the period ending at the specified time.* (p. 43)

That is, remuneration should be tied to particular classes of equity and debt in proportion to their value, minus the value of rescue packages.[155] If remuneration is paid out over a sufficiently long period of time—at least a decade, as is the case for many existing corporate pensions—then managerial return becomes a given percentage in the returns to society as a whole.[156]

---

[154] This is related to Romer's (2012) call for a regulatory approach that emphasises "responsibility" rather than "process". The "legalistic, process-oriented approach" must "evolve to keep up with a rapidly changing world." (p. 111) The approach of responsibility gives individuals a stake in actual outcomes, regardless of process.

[155] This is akin to Edmans & Liu's (2011) suggestion that the mix of equity and inside debt should reflect a firm's degree of leverage, with the addition in Bebchuk & Spamann (2009) of mechanisms addressing the unique moral hazard problems for systemically important financial firms.

[156] Without options over shares or bonds, which would reintroduce one-sided risk.

This is the regulatory realisation of high $r_s$—of lockstep structural payoff dependence. It is worth revisiting the ubiquity of $r_s \approx 1$ in biological cases of productive and stable mutualisms. High degrees of "skin in the game" is the main method for sustaining cooperation in complex biological systems, where complexity means that the cooperative system is vulnerable to many avenues for parasitic attack. It is employed precisely because the alternative, rules, are necessarily incomplete and fragile amid such complexity. The bloodhounds need only watch over the means of generating high $r_s$—a relatively simple task.

This arrangement has some useful properties. The greater the leverage, the more managers become concerned about the value of inside debt and risk of government aid, and vice versa. This provides an incentive to choose levels of risk and leverage that maximise expected social value. If remuneration is paid out over a long period, then this incentive extends to ex-managers who hold pensions—producing a group of payoff-dependent actors with sectoral knowledge, insider knowledge, and an interest in both idiosyncratic and systemic solvency. The greyhounds may run on the side of the bloodhounds.

This would not address all the important sources of financial parasitism, only those associated with bank risk-taking. Nor does it address the problem of animal spirits—many executives fuelling the GFC were "true believers" in the new normal of financial stability. Nonetheless, it is worth considering a tool that harnesses lockstep payoff transformation in the world of the Red Queen.

# 12E2

## Empirical case 2:
## The Opioid Crisis and the second leg

Functional healthcare systems rely on a specific division of labour. Pharmaceutical firms innovate in the creation of novel drugs. Doctors and other medico-scientific actors impose market discipline on pharmaceutical firms by determining the value of those treatments for patients. Third party groups further discipline pharmaceutical firms by acting as independent channels of expert knowledge and patient experience, which in turn shape healthcare policy and doctor practice. Insofar as doctors and third parties carry out these roles, they act as the agents of patients.

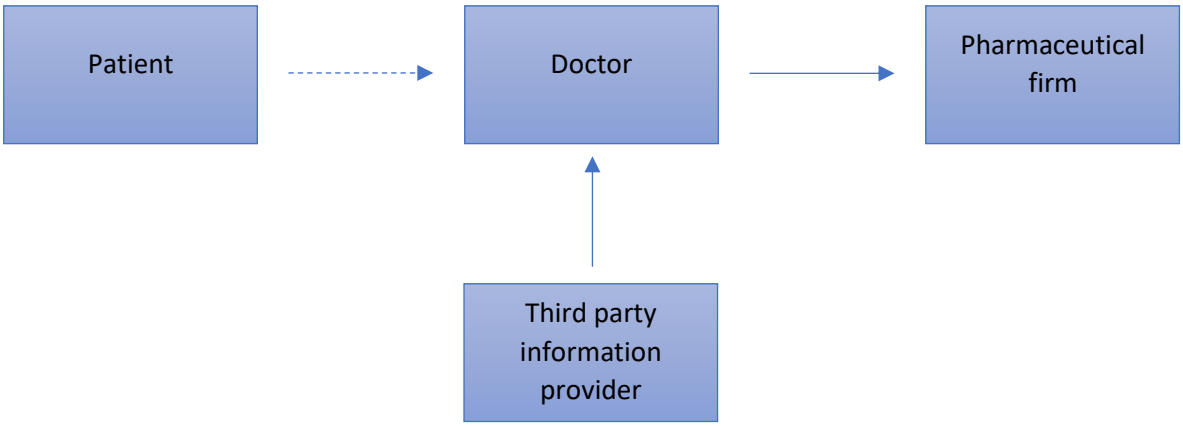



**Figure 12E2.1** A simplified schematic of the flow of "market discipline" to firms. Patient needs and information from the medical community flows through doctors, who in turn make decisions that exert discipline on pharmaceutical firms—i.e. determine firms' profitability and innovation incentives.

If the regulatory structure permits pharmaceutical firms to direct a share of profits to prescribing doctors and third-party advocacy and advisory groups, these relationships reverse: these actors become the agents of pharmaceutical firms. Doctors make an attractive addition to the pharmaceutical sales team: they have the legal power to prescribe, and they hold unique relationships of trust with patients and peer physicians. Advocacy and advisory groups offer the trust fostered by their apparent independence and long history of productive contributions. Pharmaceutical companies naturally lack these kinds of trust, given their motive of profit-

maximisation. This creates an opportunity for a new, parasitic division of labour, in which trust is exploited to generate excess sales. This can be framed as a novel tragedy of the commons: generalised trust in doctors and independent advisory groups is a public good that facilitates socially beneficial flows of information. Pharmaceutical firms innovate in means of extracting from this common resource of trust, without any private incentives to maintain it.

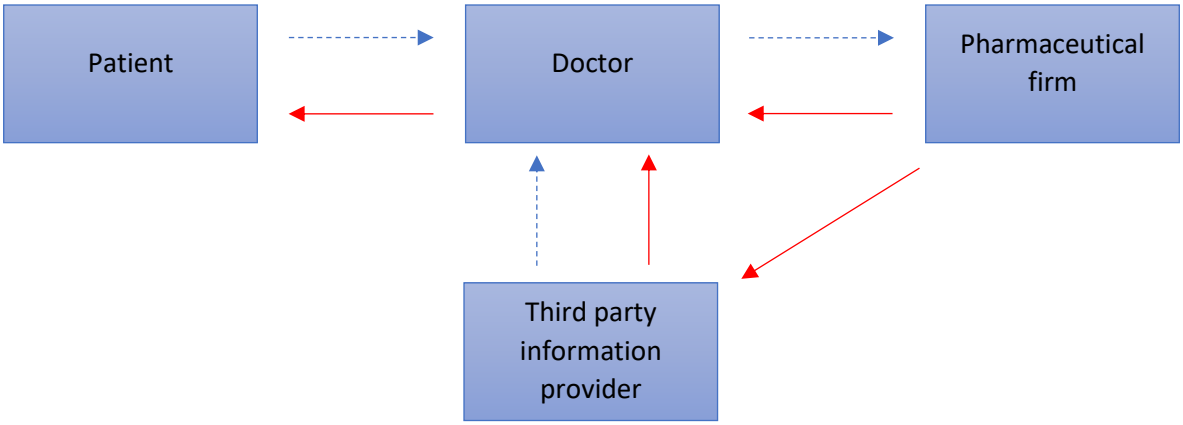



**Figure 12E2.2** A simplified schematic of a perverse division of labour, which serves pharmaceutical firm needs (i.e. profitability). Firms that can purchase favourable information, and pay doctors for prescribing, capture the system. Patient behaviour begins to reflect firm needs rather than patient needs.

Current regulatory responses are excessively "marginal" in character. Malpractice investigations rely on the laborious centralisation of distributed knowledge in particular cases. Again, there are over one million physicians in the US, and a majority are in receipt of general payments. Because their decision-making is complex and discretionary, and every healthcare practice has a unique context of client needs and preferences, there are many potential alibis for unscrupulous behaviour. Only the most obvious cases of malpractice are detectable—and at great expense. As of 2019, just 250 US doctors have been prosecuted for opioid-related malpractice (Berman & Li, 2020). Notably, these prosecutions only occurred after the crisis had already materialised and attracted a regulatory response. Meanwhile, in the absence of galvanising crises, myriad other pharmaceuticals are still overprescribed (Sacarny et al., 2016). In aggregate, the costs of this behaviour are likely enormous.

The 2010 Physician Payments Sunshine Act is another classical marginal intervention, relying on the theory that improved information will strengthen market discipline. It requires that all payments to physicians are reported and the data is made publicly available. The extent of weaknesses in patient detection—for one, that only 12 percent of US individuals know the data is publicly available (Pham-Kanter et al., 2017)—means the Act imposes little discipline, and consequently physician payments have not declined (Open Payments, 2023).[157]

[157] Another "light touch" intervention involved sending letters informing over-prescribing doctors that their treatment patterns were unusual. Effects range between zero (Sacarny et al., 2016) to modest (10 percent reduction in Sacarny et al., 2018).

*Economists of a libertarian persuasion have for some time labored under the delusion that there is something called laissez faire and that once there are in place "efficient" property rights and the rule of law the economy will perform well without further adjustment… not only must factor and product markets be structured at a moment of time to get the players to compete via price and quality (rather than by killing each other or engaging in other kinds of anti-social activities) but the conditions for maintaining market efficiency will vary over time with changes in technology, human capital, market conditions, and information costs.*

Douglass North (2005, p. 122)

*There is more than a little anarchism (usually phrased as libertarianism) in the current American credo…*

Herbert Simon (1968/2019, p. 155)

*Nature, to be commanded, must be obeyed.*

Francis Bacon (1620)[158]

# 13

# Concluding Part 3: Central planning and libertarianism as two "fatal conceits"

Here I conclude Part 3, on the technical problem of designing an adaptive total regulatory structure, and explain what has gone wrong—both in the two crises discussed, as well as the broader growth in economic rents described in the introductory chapter. The goal of this chapter is to set out the main implications for the consequences of grand theories of economic management: central planning, libertarianism and classical liberalism, and the mixed economies in between. Each of these can be interpreted as a theory about the optimal division of detection and enforcement (D&E) labour between individuals and institutions.

[158] Translated from Novum Organum, see Womald (1993, p. 168).

I argue that the two extremities, central planning on the one hand, and libertarianism and classical liberalism (see Box 13.1) on the other, are untenable and inevitably maladaptive. They commit to almost pure institutional D&E or almost pure individual-level D&E respectively, when the two are essential complements. They pledge themselves to fixed overall strategies in a dynamic world, when the comparative advantages of market and state regulatory actors are positive facts to be discovered, and moreover will vary over time and across different markets. What Hayek wrote of central planning in fact applies to both extremes: they are "fatal conceits".

Section 1 summarises the main implications of the Red Queen's race for central planning, and for libertarianism and classical liberalism. Their failures are symmetric. Both advocate divisions of D&E labour that rest on vast overestimations of certain actors' capabilities. Both fail to utilise the natural complementarities between the institutional structuring of markets on the one hand, and the individual use of distributed knowledge on the other.

Section 2 points to the "Libertarian's Trilemma", a paradox faced by scholars and policymakers committed to "deregulated" markets. They may have any two of the following, but having all three is impossible:

A. economic efficiency and stability;
B. freedom of contract (i.e. no structural regulation); and
C. small government (i.e. minimal institutional marginal regulation).

To choose B and C is to sacrifice A. It is to reject any adaptation of the total regulatory structure, and so play the sleeping Hare. The simple designs of the libertarian and classical liberal are progressively exploited by parasitic counterstrategies, generating soaring economic rents, slowing growth, and a parade of crises in complex markets.

Section 3 turns to revisit three key elements of Hayekian theory, which he wrongly believed pointed to the desirability of minimally regulated markets: first, distributed knowledge; second, the price signal; and third, the epistemological constraints on the theorist and policymaker. I show how Hayek's most important concepts can be used to dismantle central planning *and* his own classical liberalism.

The conceit of central planning is the idea that it may act as a substitute for market actors; its pretence of godlike interventionism requires something approaching omniscience if it is to be efficient. The conceit of libertarianism and classical liberalism is to play the deist god, who sets initial market rules (their "big bang") and argues that their rules will remain "unbeatable", and work as intended, in perpetuity. The very point of markets, their blessing and their curse, is that individuals will innovate in myriad ways we cannot foresee. Our unceasing task is to observe, experiment, and adapt.

**Box 13.1. Hayek's classical liberalism**

In this chapter, Hayek will be the representative for modern "classical liberalism", as its most influential advocate. To critique Hayekian theory, or even to attempt to summarise it, is to run into a minefield. Hayekian hermeneutics are acknowledged to be challenging by all his deep readers, critical and sympathetic alike. He was, in his own words, a "muddler" (Hayek, 1975), and claims about his arguments and beliefs can usually be refuted by an instance where he says the opposite.[159] It is worth briefly taking stock of the depth and breadth of this challenge, in part to illustrate why I cannot claim—and no one can claim—to have set out "the" coherent Hayekian position.

So then: which Hayek? The principled Hayek of zero compromise lest a single concession destroy the market, or the pragmatic Hayek making trade-offs and piecemeal concessions sufficient to build a welfare state?[160] The fierce advocate of "spiritual" and "sacred" liberty, or the evolutionary agnostic for whom liberty is only instrumental?[161] The Kantian, even Rawlsian,[162] or the utilitarian who argues

[159] Paul Krugman (2013), for example, critiqued Hayek's full-throated support for liquidationism as a cure for depressions. The libertarian Larry White (2013) agrees that Hayek advocates liquidationism, but considers it a "libel" to critique him on this basis, since on other occasions Hayek advocates the opposite and sounds Keynesian.

[160] Hayek (1960) writes that liberty must be "stubbornly adhered to as an ultimate ideal about which there must be no compromise for the sake of material advantages" (p. 130), while later supporting a broad range of compromises, including taxation, limited redistribution, various regulations, and the provision of many government services. Shearmur (1996), a supporter of Hayek, observes that he repudiates "piecemeal" intervention based on cost/benefit analysis, yet at the same time argues for piecemeal interventions that would require such analysis. "Hayek himself seems to need to be able to make quantitative assessments of the costs and benefits of various proposed policies if he is to be able to engage in argument about them" (p. 200).

[161] Within Hayek's (1960) *Constitution of Liberty*, for example, compare his remarks in the first chapter, on the sacredness of liberty (p. 52), against those in the second chapter, where the case for liberty rests wholly on its practical utility in a society with limited knowledge (pp. 80-81).

[162] Hayek (1976) writes that with respect to Rawls "the differences between us seemed more verbal than substantial", and includes a lengthy footnote (ibid, no.25, p. 188-89) indicating his own use of thinking "behind the veil" during the German bombing of London in 1940.

we must specifically maximise productivity because of the value of "movement for movement's sake"?[163] The acidic sceptic of the idea of a rational lawgiver who can consciously construct society, or the extraordinarily self-confident and radical utopian[164] who advocates for overturning regulated markets, purging primitive altruism,[165] eliminating national currencies,[166] and even offers his own eccentric constitution,[167] all presented with a "considerable degree of finality"?[168] The anti-interventionist, anti-constructivist Burkean with a belief in the wisdom of accumulated culture, or the impassioned supporter of Pinochet's radical reconstruction of Chilean society,[169] who hoped to "sweep away restrictions on free markets that have been in place for generations"?[170] The economist of imperfect, complex, and non-equilibrating markets,[171] or the economist who believes that government intervention is the only cause of market disequilibrium and expectations adjust perfectly to counteract Keynesian stimulus?[172] The legal scholar who argues that "true law" cannot "single out any specific persons or group

[163] Hayek promises an "ultimate justification" for liberty (1960, p. 49), and this appears to be its utility for using knowledge and consequent effects on productivity, which in turn are valued as "movement for movements sake" (1960, p. 95). Gray (1998, p.154) observes this is a "candidly nihilistic avowal".
[164] What we need, argues Hayek (1967, p. 194), is "a liberal utopia, a programme which seems neither a mere defence of things as they are nor a diluted kind of socialism, but a truly liberal radicalism which does not spare the susceptibilities of the mighty".
[165] See Chapter 1 of Hayek (1988) *The Fatal Conceit* and his discussion of the three levels of moral traditions in his interview with O'Sullivan (see Hayek, 1985). As Miller (2010) approvingly summarises, this primitive impulse is "an enemy of the extended, rule-governed, open society" (p. 71).
[166] See Hayek (1978) *The Denationalization of Money*. Friedman (1986) notes the paradox inherent in Hayek's averred traditionalism, coupled with this radical experimentalism. Gray (2013) is more acerbic: "Let us imagine that the European Union or the United States would actually dare to conduct such an experiment. What would be the consequence? The Chinese would be laughing up their sleeves – the renminbi would become the world currency. The world is not simply a textbook" (para. 18).
[167] Including such oddities as a government stocked by 45 year-olds voted in by their peers and ruling for 15 years. As Gray (2015) writes, "his scheme for an ultra-liberal constitution was a prototypical version of the philosophy he had attacked" (para. 15).
[168] As put by Eugene Miller (2010, p. 67).
[169] Hayek acknowledged that in his support for Chile's radical economic and constitutional changes, he would "lay myself open to the… objection of being inconsistent… I have been arguing that constitutions in the old Whig tradition ought to grow and not be made; and to suggest any completely new constitutional system is somewhat absurd." (Hayek, 1960b, p. 819)
[170] A quote from Gray (1998, p. 153). Gray continues: "Hayek is torn between his Burkean evolutionary theory and Enlightenment rationalism—arguing that traditions and institutions that have persisted over generations embody accumulated wisdom *beyond that which a rational mind could assess*, and yet that *his own rational analysis* and the yardstick of productivity (and, even more strangely, supported population) unquestionably show that free markets are truly wise and contrary institutions are erroneous."
[171] Hayek (1945) emphasises the "unavoidable imperfection of man's knowledge" (p. 530), and elsewhere that rationality is not a given but is *selected for* within market institutions (Hayek, 1960). For such positions, Hayek is even claimed by some to be a progenitor of behavioural economics, e.g. Frantz & Leeson (2013) and Rizzo (2016).
[172] As the then-enthusiastic Gray (1984, p. 88) summarised: "he believes economic discoordination results always from institutional factors [i.e. government], so that at any rate large-scale disequilibrium would be impossible in a catallaxy of wholly unhampered markets." When there is discoordination, which must result from government error, then the appropriate response is "simply to allow the spontaneous cleansing process of recession to take its course."

of persons",[173] or who argues in the same breath that there must be laws "that apply to different classes of people" with "properties that only some people possess"?[174]

Here we will focus on Hayek at his most memorable and assertive, and so at his most influential. This is the Hayek who argues that it is

> *part of the liberal attitude to assume that, especially in the economic field, the self-regulating forces of the market will somehow bring about the required adjustments to new conditions.* (Hayek, 1960/2020, p. 521)

Gray (1984), who Hayek dubbed the only surveyor of his work who "fully understands" it, summarised the Hayekian position as preferring that

> *the self-regulating tendencies of the process be accorded unhampered freedom and that governmental intervention be recognized as the major disruptive factor in the market process.* (p. 90)

Later, Gray's "full" understanding led him to become a major critic of Hayek's work (Gray, 1998, postscript). In any case, this is a clear enough position on the division of D&E labour: Hayek did not endorse *laissez faire*, but he was much closer to that position than almost any figures in the mainstream that he criticised.

## 1 Central planning and libertarianism as defective divisions of D&E labour

The last couple of chapters concluded that marginal and structural regulatory approaches are complements. Market actors have a comparative advantage in assessing marginal payoffs, so long as they can handle the complexity of the detection task (and there are no externalities). The state regulator structures the problem, so that market actors can solve it.

Central planners, and libertarians and classical liberals, make the same kind of error: they hope to rely mostly or wholly on either individuals or state institutions, while

[173] Williams (2006, para. 9) remarks that "…during dinner with the late Nobel laureate Friedrich Hayek, I asked him if he had the power to write one law that would get government out of our lives, what would that law be? Hayek replied he'd write a law that read: Whatever Congress does for one American it must do for all Americans."
[174] These statements are found in Hayek (1960, p. 222, pp. 317-318).

rejecting the other. Such approaches are guaranteed to fail. Their main failures are as follows:

- Central planners suppress and lose the benefits of *domesticated competition*. Extreme structural constraints prevent market actors from supplying various valuable *mutualisms*; they are blocked from using distributed knowledge to allocate resources, and from supplying massively-parallel search of the design space.
- Libertarians and classical liberals permit and invite the costs of *antagonistic competition*. Structurelessness allows firms to supply a broad variety of *parasitisms*, using distributed knowledge to disrupt other market actors' use of marginal regulation.

For both, complexity is the critical driver of failure.

- Central planners believe market actor marginal regulation can be replaced with a marginal institutional apparatus. They may be able to muddle through in a simple economy, and indeed the USSR excelled at capital-intensive growth of heavy industry. As economic complexity inevitably grows, however, the replacement of market actors becomes less and less feasible.
- Libertarians and classical liberals place the burden of D&E on market actor marginal regulation. Market actors may be able to muddle through in a world of wool, wheat, and wine, and where one can look a horse in the mouth. But as complexity grows—bringing a world of 350,000 chemicals, devices with myriad unknown points of failure, and complex derivatives—they confront problems that grossly exceed their capacities.

The framework of this thesis predicts the kinds of failures we see in both systems. Central planning suffers from broad inefficiencies: widespread misallocations and widespread failures to motivate worker productivity.[175] It specifically fails to drive innovation in diverse and high-quality consumer goods and services, but may perform reasonably well in areas where knowledge is less dynamic and more accumulative. Libertarian / classical liberal failures are concentrated in complex and welfare-critical markets, including finance, healthcare, education, and information

[175] Shirking being a simple form of D&E evasion.

markets, where complexity supports profitable investment in market actor exploitation.

**Box 13.2 On market failure**

A theorist might respond, "there are indeed market failures". I want to suggest, however, that this is rather like a central planner admitting that there will be planning failures. Both are true. The problem is that beginning from the default of planning, or of minimally regulated markets, leads to tinkering at the edges of a starting point that may be grossly maladaptive.[176]

Extensive market failures cannot be resolved by tinkering with "marginal" regulation. I have suggested the marginal approach to Basel III, for example, reflects the attempt to stabilise maximally free markets via intensive tinkering (but also see, for example, the U.S. healthcare system). Because this tinkering raises complexity, it has the second-order effect of intensifying the Red Queen's race. We can predict, therefore, a parade of *new* market failures, and both market actors and institutional actors remain saddled with tasks they cannot complete—inviting still more tinkering and more complexity.

Thus, in the world of the Red Queen, market failure is not a static fact, but a *dynamic process* affected by the choice of regulatory strategy. Where the market default is grossly inadequate, as in the case of complex public goods such as financial stability, this process can be managed via structuring markets. This may entail control over incentives, as in the cases of Glass-Steagall or doctors being insulated from pharmaceutical payments. It may be outright prohibition, such as bans on resecuritisation. I add some further comments on the middle ground between markets and planning at the end of this thesis.

## 2 The Libertarian's Trilemma

The libertarian and classical liberal result can be stated in terms of a *trilemma*, which poses a paradox for theorists and policymakers with normative commitments to minimally constrained markets and small government. Let us frame the ardent "deregulator" as having normative commitments to:

[176] The worst example being "nudges", which are—except in a few endlessly trumpeted cases, such as organ donation—generally as insignificant in their effects as the name suggests.

1. economic efficiency and stability (i.e. a society that can compete, prosper, and last, and so necessarily has mechanisms for containing parasitism);
2. freedom of contract (i.e. permissiveness and structurelessness, and therefore heavy dependence on marginal regulation provided by market actors); and
3. a small and simple state regulatory apparatus (i.e. minimal state investments in D&E, especially costly marginal D&E).

Such theorists are fearful of state *marginal regulation,* because it implies large government, and fearful of *structural regulation*, because it limits choice and theoretical allocative efficiency.

In *complex* markets, we may satisfy any two of these but not all three.[177] This is the "Libertarian's Trilemma". As is usual in trilemmas, we can point to three possible systems that each satisfy only two of the libertarians' goals:

First and most important, the committed libertarian who chooses freedom of contract and a small regulatory apparatus (#2 and #3) sacrifices economic efficiency and stability (#1). To reject structural and marginal regulation is to reject institutional adaptation altogether. Since market actor performance is not in our control except by use of institutions, this is to *reject adaptation of the total regulatory structure*. The regulator withdraws from the Red Queen's race and sleeps at the side of the track, in the conviction that they have found an eternally "unbeatable" regulatory combination: basic structural regulations such as property rights, plus market actor rationality. I return to this epistemological claim towards the end of this chapter. Its consequences are the GFC and Opioid Crisis, and the still worse economic and political calamities that would visit the genuinely libertarian society.

Second, if they are committed to efficiency and stability and cherish total freedom of contract (#1 and #2), they must concede the need for a complex marginal apparatus and big government (sacrifice #3). In this case they at least have some means of adaptation, and some modicum of efficiency and stability may be maintained with a

[177] Note that if any contemporary markets appear "simple", this is usually the product of regulation, and under deregulation most markets would range from complex to hyper-complex. For example, deregulated markets for food, clothes, cosmetics, and so on, would bring risks of exposure to an ever-growing variety of chemicals that may—alone or in nearly infinite possible combinations—cause long-term health problems, but that may be useful for cheaply producing apparently desirable products. Individuals cannot police this alone, short of making the choice of home production.

hefty marginal institutional apparatus like Basel III, if much larger still. This approach demands very large investments in regulatory capacity; without structural regulation, the bloodhounds must chase the greyhounds over limitless open plains, and have their work cut out for them.

Third, if they are committed to efficiency and stability and prioritise a small and simple regulatory apparatus (#1 and #3), they must sacrifice freedom of contract (#2). Again, they at least have some means of institutional adaptation. All remotely marginal institutional strategies are off the table because they entail larger, more complex government. To preserve some modicum of efficiency, they must use simple and blunt structural tools to rule out major parasitisms—and sacrifice large parts of the mutualistic choice set.

To return to the case of the GFC, "deregulation" from the early 1980s onward was a move towards the first option. Post-GFC, regulators have largely selected the second: an excessively permissive, and hence hyper-complex, marginal regulatory apparatus that creates difficult Red Queen's races. To the extent that the regulator cannot handle these tasks and/or does not invest sufficiently in the adaptation of this marginal regulatory system, this becomes the *de facto* continuance of the first option.

Finally, note that the theorist who is committed only to efficiency and stability is agnostic about the choice of regulatory instrument, so is not bound by the trilemma.[178] Structural and marginal approaches may be integrated in whatever way best uses the comparative advantages of market actors and the state.

## 3 Turning Hayekian theory against Hayekian classical liberalism

To complete the argument, it is helpful to revisit Hayek's key concepts and arguments in support of free markets: distributed knowledge, the price signal, the entrepreneur, and the problem of epistemological constraints on the regulator and theorist. These concepts are essential, but properly understood they are all double-edged, cutting into central planning but also libertarianism and Hayekian liberalism. The nub is that these ideas undermine *any* theorist who commits to applying any

[178] Of course, they may also be sensitive to citizens' preferences about the management of trade-offs between freedom of contract and efficiency.

particular division of D&E labour across a range of markets, in advance of the laborious work of studying actor capabilities and comparative advantages.

### 3.1 The two edges of distributed knowledge and the price signal

Hayek's (1945) notion of distributed knowledge is widely considered a critical justification for the market system, and the framework in this thesis concurs. It is the *unforeseeability* of the strategies that market actors will discover and enact within market rules, and the unforeseeability of the adjustments they will make to changing local conditions, that makes well-regulated markets superior to central planning.

Unforeseeability, however, inflicts similarly grave wounds on the case for the free market: knowledge about available parasitisms is no less distributed, and knowledge about future parasitisms no less unattainable. Neither the regulator nor the armchair theorist can foresee the parasitic strategies that massively-parallel search will uncover, from mortgage-backed securities to opioid marketing strategies. To confidently predict that the market will prove "self-regulating" in the next period depends on such impossible centralisation and foreknowledge. All we can foresee is that individuals will devise and pursue innovations in the top two quadrants of the social strategy space.

It is logically untenable to claim that we *cannot* foresee innovative mutualisms, and so we need markets, but that we *can* foresee innovative parasitisms, and so know in advance that a specific set of market rules will contain them. In short, Hayek's theory is one of *distributed knowledge in the mutualism quadrant*, consistent with his broader economics of the mutualism quadrant.[179] A complete theory of distributed knowledge gives us no guidance whatsoever about the desirability of free markets, or any specific set of market rules. It supports the case for *domesticated* markets, but it leaves the question of what D&E structure domesticates markets at any given point in history completely open.[180]

[179] Hayek's ideological neighbour, Buchanan (1977), observes that Hayek makes the error of reading the "invisible hand" and "spontaneous order" as necessarily benign. The "littered beach", Buchanan writes, is equally a spontaneous order, that no mind intended, created by the invisible hand of individual optimisation. Hayek leads himself "into what we must classify finally as a logically inconsistent position." (p. 37)

[180] Another metaphor Hayek (1964, p. 5) employs, likening the market to the "emergent" or "spontaneous" order of the crystal, is equally ambiguous. Both ordered *and* disordered molecular structures emerge spontaneously, and the result is one that no individual could arrange by hand. To get order, the technician needs to use the right mix of *elements* in the right *conditions* (i.e. about

The same ambiguity forces a reinterpretation of Hayek's price signal. His critique of central planning is well-made. Freely moving prices allow a shock to production or preferences in some remote district to be reported in prices throughout the whole system, so that all actors may adjust their plans accordingly. Prices may guide the overall allocation of labour, capital, energy, and other resources to their most valuable uses. They guide entrepreneurial effort, by providing information on expected input costs and the sale price for outputs. As Hayek emphasised, their total effect is to produce a mutual adjustment of plans along a development pathway that no central planner could possibly foresee.

Yet the price signal can be no less dysfunctional under libertarianism and Hayek's classical liberalism. A given price system may perfectly match supply and demand, and yet be consistent with devastating systemic collapse. Prices for mortgage-backed securities balanced supply and demand, but they were mispriced due to failures of the total regulatory structure—the allocation of impossible tasks to market actors. The evidence for excessive financialisation surveyed in Chapter 12ES1 suggests that financial activity is still shockingly mispriced. On the matter of entrepreneurial effort, prices do not selectively reveal information to Hayek's productive entrepreneur and conceal it from Baumol's extractive entrepreneur; they equally guide investments in misrepresenting *c*, in collusion, and in regulatory capture. Regardless of whether the cause is *individual-level* or *institutional*, mispricing can flow throughout the entire economic system and produce overall misallocations.[181]

Hayek is both brilliantly correct and devastatingly wrong. The marvellous integrative power of prices may chart an emergent and spontaneous pathway to economic prosperity *or* to economic collapse, that no theorist could foresee. Which we get depends on the effectiveness of the total regulatory structure (as observed by North, 2005, in the epigraph to this chapter). For prices to work their magic, markets must be structured so that the Red Queen's races provoked by marginal regulation are tolerably symmetric.

---

1200°C and 700,000 psi to make a carbon diamond). Those elements and conditions are not a matter of ideology, but of discovery. In human orders, we have much more to understand about the relevant "elements" and "conditions", and they are ever-changing.

[181] Both the GFC and Opioid Crisis can be framed as gross mispricing, with major distortions flowing throughout the whole economy.

### 3.2 Two varieties of conceit: Libertarianism as "economic deism"

With a nod to Hayek's observation that the efficient central planner would require omniscience, here I paint the twin errors of libertarianism and central planning with reference to the inhuman powers they require. We have, for more than a century, been trapped between two extreme strategies of economic management that suppose godlike knowledge: the economic *interventionists* and economic *deists*.

The special conceit of the central planner is that of someone who imagines they may behave something like an interventionist god—that they may, as Smith (1759/1976) put it, "arrange the different members of a great society with as much ease as the hand arranges the different pieces upon a chess-board" (part VI, ch. 2, para. 17). Given the Chapter 6 discussion of Darwin machines, and comments on the limits to regulator capacity in Chapter 12, the framework of this thesis supports Hayek.

The special conceit of the libertarian is subtler but even greater: it is that of someone who imagines they may ape the deist god. In the deist religious tradition, it is commonly believed that a Creator set the initial conditions and laws of the universe, and thereafter does not intervene. Whatever initial settings they choose, they can foresee its full implications. The economic deist similarly advocates a specific division of D&E labour, and claims that they can foresee its happy consequences. For the libertarian or classical liberal, the creation moment is the formation of basic market institutions and a tightly constrained state (e.g. Pinochet's radical, and failed, economic "big bang" guided by Hayek, Friedman, and Chicago alumni). Their simple system, they believe, is unbeatable and self-regulating, so that our main task is to leave it alone.

Deists face extreme epistemological challenges. In the theological case, it takes an omniscient deity to foresee that the initial design is sound. Economic deists must claim advance knowledge of how these initial settings will interact with present and future strategies, which would require centralising distributed knowledge and seeing the future. They must know that the system can reproduce its political foundations—that it will not generate second-order economic, cultural and social effects, such as increases in inequality, poverty, or reductions in social trust and cultural tendencies for cooperation, that may lead to economic and political decline. They must know that no critical public goods will be underprovided, when there is much to learn

about the public goods of the present, let alone the future. This knowledge must be rich and detailed across a range of markets, enough that they can foresee how the multitude of Red Queen's races will play out. Greenspan knew that in the case of financial fraud, the "customer would figure it out"; Hayek knew that we need not worry about large aggregations of capital,[182] because he knew that they would never develop strategies for colluding or otherwise suppressing competition; Friedman knew that doctor deregulation would lead them to efficiently compete on price and quality; and so on.

Here libertarians find themselves in an untenable epistemological bind. They must claim an extreme degree of *a priori* knowledge with respect to the future effects of their preferred economic management strategy, while at the same time sustaining the argument that we cannot use concrete time, location and problem-specific *a posteriori* knowledge, as we acquire it, to justify any other strategy.[183]

## 4 Abandoning conceit: Toward observation, experimentation, and adaptation

Francis Bacon (1620) put the matter pithily: "[n]ature, to be commanded, must be obeyed."[184] This is obvious enough for the physical sciences, but it applies no less to the social. We cannot command economic efficiency, nor realise any other shared value, without working within observed constraints. Bacon's approach to science begins with humility before a complex world. The two arrogant gods of these systems—the omniscient interventionist and omniscient deist—pretend to have conquered this complexity. They are facades, and behind them lurk limited mortals.

Mortals must adopt a middle strategy. Because we cannot centralise enough knowledge to intervene at every step, we must build a structure that harnesses others' distributed knowledge. Because we cannot know enough to settle on one such knowledge-harnessing structure, we must regularly intervene by understanding its failures. Between the two fatal conceits lies a regulatory structure that gets the most

[182] In Hayek (1979), "[s]ize has thus become the most effective antidote to the power of size" (p. 79). Compare to Smith (1776), who remarks "[b]ut whoever imagines… that masters rarely combine, is as ignorant of the world as of the subject." (p. 65)
[183] Specialised, *a posteriori* knowledge has its uses: more Brooksley Borns (Chapter 11, Section 1.2) would have led to a different start to the 21st century.
[184] Translated from Novum Organum, see Womald (1993, p. 168).

out of individuals, but that forever adapts, because the future will humble our designs.

Adam Smith and John Maynard Keynes are figures of the middle road, both non-ideological pragmatists who were open to any regulatory strategy that appeared to serve human happiness. Both advocated markets or intervention wherever doing so best served society.[185] Both valued freedom of choice, freedom from deprivation, and the stability of the state, and recognised there were trade-offs to make.[186] It is fitting to conclude with Keynes' remarks in 1936, as he anticipated the competition between Western capitalism, fascism, and the emerging communist societies of the time—his hope that

> *we in this country may discover how to combine an unlimited readiness to experiment with changes in political and economic methods and institutions, whilst preserving traditionalism and a sort of careful conservatism, thrifty of everything which has human experience behind it, in every branch of feeling and of action.* (Keynes 1982, pp. 333-334)

[185] Smith (1776), for example, advocated for powerful interventions where they made markets serve society (e.g. the case of financial externalities, p. 353), and argued for public provision of any public goods that were found "advantageous to a great society" (Bk. 5, ch. 1).
[186] Keynes (1944, quoted in Keynes, 1980, p. 385), for his part, expressed "deeply moved agreement" with Hayek on the significance of economic freedom—if, of course, disagreeing with Hayek's "economic dicta".

# PART IV

## Democratic and authoritarian equilibria in the world of the Red Queen

*If a covenant be made… in the condition of mere nature, which is a condition of war of every man against every man, upon any reasonable suspicion, it is void; but if there be a common power set over them both, with right and force sufficient to compel performance, it is not void. For he that performeth first, has no assurance the other will perform after… without the fear of some coercive power… And therefore he which performeth first, does but betray himself to his enemy…*

Hobbes, *Leviathan*, 1651, ch. 14, para. 18

# 14

# The third leg of the relay: The meta-institutional race

One of Schumpeter's insights is that innovation displaces incumbents. Let us remember, then, that institutional forms such as democratic capitalism and competitive markets are incumbents too—and if they are displaced, it will be via a process of human innovation. How do incumbent democracies and authoritarian systems survive in the world of the Red Queen?

Each of the three legs of the Red Queen's relay is a race over the use and evasion of rules. Having dropped the assumption of a benign regulator, we arrive at the third leg and a race over the use of *meta-rules*: rules that shape regulatory actors' payoffs, and so determine in whose interests the regulator runs the first two legs of the relay. Meta-rules may take the form of formal institutions—*meta-institutions*, with constitutions the archetypal case—but also include informal strategies such as protest, quid-pro-quo payments, threats of violence, and so on. All these *de jure* and *de facto* means of shifting regulatory actors' payoffs are, once again, strategies that are vulnerable to counterstrategies.

Because there is no longer a benign third-party enforcer, we arrive at the Chapter 8 problem: a purely rules-based order is impossible, and we need a source of commitment that can anchor cooperation. An assumption retained in this chapter is that of self-interest ($r_c = 0$). Under self-interest, only commitments in the form of structural payoff dependence ($r_s > 0$) remain. A key goal of this chapter is to see how

far the self-interested model can go in explaining the survival of democratic and authoritarian social orders. I seek to answer two questions:

(A) Can we construct a stable regulator in a world of pure self-interest? Or is the system anchorless, dissolving into an anarchy of roving banditry?
(B) Supposing the self-interested regulator is stable, will it serve, to a tolerable degree, the general welfare? What determines the level of extraction?

Section 1 turns to briefly examine authoritarian systems in a world of self-interest. Consistent with Chapter 8, rules and commitment act as essential complements. Authoritarian orders are anchored by Olson's (and Hobbes') encompassing interest ($r_s > 0$), and in turn, the authoritarian uses rules to defend that commitment mechanism.

Olson's dictator has an (imperfect) interest in economic efficiency, given the bonds of taxation, but no interest in citizen welfare. Further, this interest in efficiency is in tension with the primary purpose of rules, which is to maintain $r_s > 0$ (i.e. the power to tax). Holding power is *complex*, and amid a Red Queen's race with would-be usurpers, authoritarians frequently favour the intensive use of crude "structural" strategies to simplify the contest. These are frequently incompatible with economic efficiency, entailing departures from the Olsonian optimum. Finally, even with such strategies, authoritarian order is precarious. Note that because I will not be discussing authoritarian systems after this chapter, I briefly note how motivational commitments ($r_c > 0$) modify these outcomes, both in theory and empirically.

Section 2 turns to the main case, democracy. In this chapter, democracies have no viable source of commitment. Self-interest means that $r_c = 0$. No actor has a large share in economic output, so $r_s = 0$. McGuire and Olson's (1996, pp. 83-84) democratic "superencompassing interest" is a phantom; its existence depends on citizens providing key democratic public goods, and self-interested citizens will not do so. Even if we assume that the vote is cost-free, self-interested citizens have almost zero incentive to run the Red Queen's race against regulatory capture. Nor can rules control judicial and military actors, who are opportunistic, rent-maximising mercenaries. Hobbes is correct, therefore, that the self-interested democracy is an anarchy of roving banditry. In turn, roving bandits race to become stationary bandits, and the system returns to authoritarian rule grounded in $r_s > 0$.

The Chapter 8 analysis is supported. All complex social orders are founded on commitment, and commitment and rules act as essential complements. Following Hobbes' (1651) primordial analysis of the world of self-interested actors, democracy would be an interregnum—anarchy before the return of the Leviathan.

## 1 The full race: The case of dictatorship

Authoritarianism is one option for constructing social order in our world of open-ended innovation. It is the most common in human history, and an option considered favourably or at least respectfully by many of history's greatest minds. The possibility of authoritarian order that serves the general interest can be found in the Chinese "Mandate of Heaven", with its origins in the ancient Zhou dynasty, in Plato's guardians or philosopher-kings, through Hobbes' Leviathan, and McGuire and Olson's (1996, p. 73) "public-good-providing king", among many others—the former two being primarily based in the virtue of the ruler, and the latter two being models that emphasise the self-interest of the ruler.

The picture of authoritarian order in this chapter is incomplete, with rulers, elites, and citizens being limited to the motive of material self-interest. This is by design, as part of this chapter's purpose is to explore the possible varieties of social order that can survive within a world of pure self-interest.

### 1.1 Authoritarian commitment

Can we understand authoritarianism in a world of pure self-interest? Clearly we have a ready-made answer to the Chapter 8 problem. The dictator's structural payoff-dependence on society's economic output provides the source of commitment we need. In, Hobbes (1651) this is expressed as follows:

> *no King can be rich, nor glorious, nor secure; whose Subjects are either poore, or contemptible, or too weak through want, or dissention, to maintain a war against their enemies…* (ch. 21, para. 3)

In Olson (1993), this is formalised as the ruler's "encompassing interest" that I introduced in Chapter 4. Again, this is Olson's $tb > c$, where the tax rate ($t$) acts as a form of structural payoff dependence ($r_s$). This results in the dictatorial payoff function:

$$B_i = r_s b - c$$

This gives the dictator an innovation-proof motive to repair and improve the institutional structure—to introduce property rights, defend the rule of law, and run the first two legs of the Red Queen's relay. An "invisible hand" gives the ruler "an incentive to make himself a public-good-providing king" (McGuire & Olson, 1996, p. 73).

### 1.2 Rules sustaining commitment

But commitment alone is not sufficient for social order; it must be supported by a system of rules that defend that commitment. Olson's model is fine enough for describing the case of an unassailable ruler—perhaps one empowered by the gods, as if the myths propagated by the old kings and pharaohs were true—who need not bother with rules.[187] I call this the "uber-dictator". Such a ruler would establish an institutional structure purely designed to maximise efficiency according to Olson's formula $tb - c$.

In practice, the uber-dictator's Olsonian optimum is unachievable. Following the discussion in Chapter 8, commitment mechanisms are vulnerable to free-rider-like behaviour. Competition for the seat of power causes a tragedy of the commons: while the ruler and every competitor for the throne would be better off maximising efficiency and sharing the fruits in a Coasian manner, there is no third-party enforcer, and every actor is better off plundering in almost any way necessary if it allows them to win the leadership contest. As discussed in Chapter 4 (Box 4.3), this mirrors the sociobiological case where intense competition between parasites reduces the value of $r_s$ and makes them more virulent, destroying the public good of host productivity and survival.

Chapter 8 argued that committed actors devise rules to preserve the source of commitment—in this case, to defend the power to tax. Gerschewski (2013) identifies three pillars in the dictator's strategy for maintaining power: repression, co-optation, and legitimation. Repression and co-optation are *c*-modifying rules, and are the forces that have been emphasised in most of the literature since the 1970s. They will

[187] This would be, in Neundorf et al.'s (2020) terms, a maximally exclusionary regime where the dictator is a "winning coalition" of one, and needs no supporters.

be our focus here. Legitimation is excluded here, as it relies on building genuine commitment to the regime among citizens and the elite ($r_c > 0$), though I add a few comments on its significance towards the end of this section.

Who must be co-opted and repressed? As North (1979) points out, in a world of material self-interest, free rider problems prevent collective action from the citizenry, such as mass protest.[188] All threats to the dictator will take the form of "palace coups" (p. 258). The ruler's problem is how to control elites. If elites are to continue to support the ruler, and opponents are to be thwarted, the ruler must in turn apply repression and co-optation to elites—and mainly to elites with significant *de jure* and/or *de facto* power. The ruler will weight their payoffs, and so design an institutional structure that serves their interests, according to their degree of power and whether he/she decides to co-opt or repress them. Military elites are the foremost targets of co-optation, given their comparative advantage in the exercise of violence, and so enforcement power. If they can be co-opted, these same capacities make them among the most useful agents of repression. To co-opt military elites, however, the dictator must be able to credibly promise to distribute rents—and so he/she must also have a credible program for economic extraction, and for the co-optation of business elites who may, in modern dictatorships, be provoked into capital flight (Gerschewski, 2013).

The winning coalition—the group of actors that are necessary to secure the ruler's power, and so benefit from the rules he/she chooses—may therefore be quite large, although some elites are weighted much more than others.

All this logic equally applies to would-be usurpers: they must evade detection while assembling a parallel co-optation and repression structure that can credibly reward allies and punish enemies, and so build a new winning coalition (de Mesquita et al., 2003). Regime change comes if emerging competitor co-optation and repression structures offer larger credible rewards and threats than the presiding.

[188] The private risks of protest are high, and the benefits uncertain and diffuse.

## 1.3 Intensive structural regulation: The costs of slowing a complex third leg

There is a natural tension between the Olsonian optimum created by structural commitment, and the imperatives of maintaining that commitment via rules. That is, tax maximisation and the ruler's power-retention strategy come into conflict.[189]

The reason is that the dictator's co-optation and repression apparatus will usually involve "structural" regulatory strategies that are, to some degree, crude and inefficient. This is because of the same factors that in Part 3 called for the use of structural strategies:

1. *stakes of uncontrolled parasitism are high*, or one may say in the language of Chapter 11 that retaining power is "welfare-critical". Deposition, and possible imprisonment or execution, is unpleasant to say the least. The ruler can be taken as having a low marginal utility of income, above that base level of resources that is necessary to sustain power;
2. *the political problem is complex*. Pressing towards allocative efficiency, by opening up markets and eliminating rents, increases tax takings, but has second-order effects on the distribution of political power, and on the alignment of elite interests. Knowledge about the consequences of these changes is distributed and belongs to the future; it is difficult to centralise and subject to profound uncertainty. It may be difficult to compensate elites who lose access to rents;[190] and
3. *innovation powers may be asymmetric*, at least in a rapidly-changing politico-economic landscape, where the ruler is confronted by the parallel innovations of many emerging economic and political players.

De Mesquita et al. (2003) write that rulers

[189] In their examination of archival documentation from the Stalinist state, Gregory & Harrison (2005) note that, in contrast with the expectations of Western theories of socialism, "[t]his regime was indifferent to calculation, preoccupied by the need to punish and deter its enemies, and bent on implementing its decisions through a complex administrative hierarchy of agents motivated by threats and promises." (p. 724)

[190] Among the many things the ruler does not know is how much compensation powerful actors require for their continued support (Abdukadirov, 2010).

> *are not likely to survive the elimination of patronage or the corrupt benefits of cronyism. For autocrats what appears to be bad policy often is good politics.* (p. 19)

In other words, rulers frequently sacrifice social efficiency *b* to prevent the tax rate *t* from falling to zero (i.e. via their overthrow).

North (1979) observes that a common strategy is to provide powerful elites with large and inefficient protections for their industries and firms, and to extract from closely held monopolies. Compensating Coasian transfers would be an efficient substitute, but they are in practice a poor one: elites' *de facto* power is tied to their special positions and networks in the economy. With the introduction of competitive markets, elites would reasonably fear shrinking as economic forces. The dictator cannot credibly commit to sharing the extractive surplus with elites that have lost influence. As Acemoglu (2003) argues, political Coase theorems fail primarily because of the absence of credible third-party enforcers.[191]

Dictators sometimes appear to deliberately suppress growth as a means for keeping alternative coalitions in the selectorate weak, or to exclude actors from the selectorate altogether. Acemolgu, Verdier & Robinson's (2004) fascinating study of Mobuto Sese Seko and Rafael Trujillo explores how some dictators sustain their rule by keeping elites, and especially military elites, divided, disoriented, and ineffective (also see Pilster & Böhmelt 2012). Debs (2007a and 2007b) argues that dictators may be more likely to retain power under low-growth conditions, and have incentives to encourage subordinates to waste resources so that they cannot develop a reputation for aptitude (see also Svolik, 2008).

For similar reasons, Egorov & Sonin (2011) posit that dictators may favour incompetent, non-threatening subordinates. Others such as Caselli & Cunningham (2009) and Robinson et al. (2006) argue that poor, resource-rich states in particular are likely to suppress growth, because growth poses the risk of political change while adding relatively little to the flow of rents from international revenue.

Box 14.1 discusses how changes in relative innovation capacity affect such patterns.

[191] I.e. there is no outside agent that can ensure the dictator honours his agreements.

**Box 14.1 The efficient surveillance state? Changing relative innovation capacities**

Evolution evolves, and humans too devise innovations that enhance innovation capacity. The tension discussed in this section may be relieved if the ruler's access to information, and capacity to process it, grow. Today's nascent authoritarian surveillance states, with ICT infrastructure increasingly paired with AI, may increase the effectiveness of the co-optation and repression apparatus. Automated tracking of individuals' movements, associations, and communications may allow coordination to be thwarted in its infancy. The digitalisation of all transactions allows most economic activity to be traced and allows punishments and rewards to be more precisely targeted—allowing a shift from more structural towards more marginal power-maintaining strategies.[192] This appears especially likely to suppress mass collective action (beyond the discussion so far, though mentioned in the next section), at least of the pre-planned variety. How it will impact elite competition remains less clear. The race will continue.

This may have some paradoxical effects on economic performance. If the information dictatorship makes the ruler more secure, it may make him/her more able to pursue Olson's efficiency optimum. At the same time, it may allow extraction to be targeted more precisely according to observed elasticities between tax rates and effort (see A.14.1 for an example from Stalin's USSR). The information-rich stationary bandit may prove an especially dangerous adversary for competitor states, if it can combine heavy extraction with an efficient economy, increasing the resources available to invest in between-state conflicts.

## 1.4 Authoritarian regimes as precarious

Do these investments in co-optation and repression make authoritarian order stable? This depends on the relative competitiveness of the ruler and his/her opposition in assembling "winning coalitions". It is not a question that can be answered in theory, and I turn to some empirical evidence. In doing so, I must be clear that this empirical evidence reflects patterns in *actual* authoritarian societies, in a world where mass collective action exists and contributes to authoritarian instability. It does not tell us

[192] China's Social Credit System is already used to apply harsh penalties (e.g. limits to mobility, financial access, and so on) only to actors evaluated as uncooperative with the state—a vastly more efficient mode of securing power than applying such constraints broadly.

what would happen in a world of pure self-interest, although we will see that the observed patterns are suggestive.

Interestingly, Papaioannou and Van Zanden (2015) find that over the period 1960-2009, the probability of ruler deposition is log-linear on the length of tenure, remaining fixed at around 0.11 each year. This is consistent with Red Queen dynamics, where the relative competitiveness of the incumbent and his/her opposition changes stochastically. Tenure length does not predict survival in the next period. If there is such a thing as talented, rather than lucky, rulers, perhaps the talented take greater risks, or face more talented usurpers.

Levitsky and Way (2022) find a similar pattern of decay for regimes rather than rulers (Figure 14.1). Their study also provides a chance to identify the effects of the third of Gerschewski's (2013) three pillars, legitimation (in this thesis, $r_c > 0$). Given that the remainder of this thesis is concerned with democratic capitalism, I will add a few brief comments on the effect of $r_c > 0$ in authoritarian systems here. Gerschewski (2013) notes the common claim

> *that legitimation simply does not matter for the stability of autocracies, as such regimes do not need to rely on people's support*. (p. 18)

One of his goals is the reintroduction of legitimation, once "at the core of classic studies" (e.g. in Weber, 1922/2009, termed a legitimacy belief). Levitsky and Way's (2022) research supports this effort. They find that over the period 1900 to 2015, regimes built on social revolutions (e.g. China, Cuba, Iran, Mexico, Vietnam, and the USSR) were significantly more stable, and they explain this by more profound legitimation processes at work—particularly by a loyal elite and coercive apparatus. Similar results are found in Kailitz's (2013) study of regimes from 1946-2010: ideologically-based regimes, particularly monarchies and communist regimes, survive 2-3 times as long as non-ideological regimes. We can add the multi-century survival of religio-cultural monarchic dynasties throughout history.

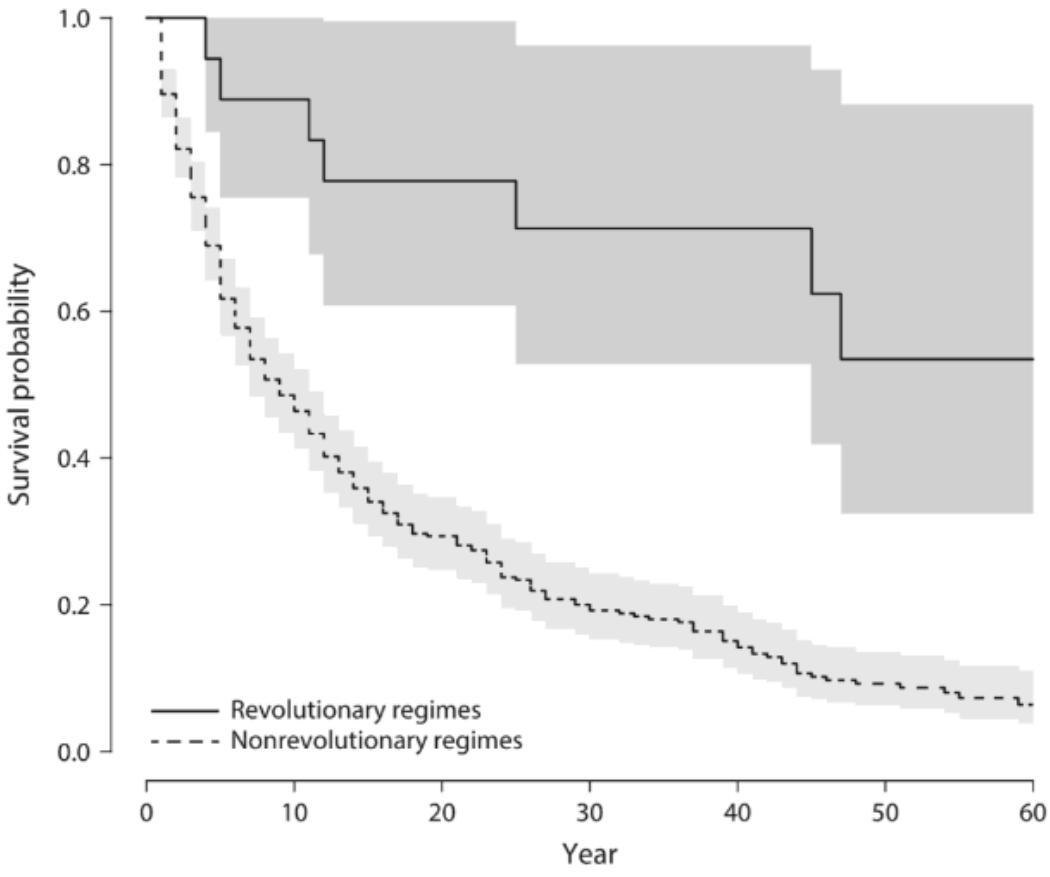



**Figure 14.1** Rate of regime survival by years in power. From Levitsky and Way (2022).

Within the framework of this thesis, shared commitments lighten the burden placed on rules. Concretely, citizens and elites who have some genuine commitment to the ruler and his/her goals, will place less value on parasitic strategies, and are more likely to pay costs to contribute to the D&E apparatus—alerting the ruler to suspicious activities, for example. The resistance of such commitments to innovation allows them to slow the Red Queen's race. Note that the more a state depends on motivational commitments, the more it must devise rules that maintain that commitment—via information control, but also delivery of favourable outcomes.

Regimes that lack the force of legitimation are, empirically, less stable. However, in a world of pure self-interest, only non-legitimised states would exist, and we could expect a higher average rate of ruler and regime turnover than that observed in our world.

## 1.5 The fickle and fleeting Leviathan

The simple discussion above is enough to answer question A from the introduction: it is possible to put together a coherent story of authoritarian order in a world of pure self-interest. The source of commitment is the ruler's dependence on the state;[193] this commitment is both secured by, and channelled to repair, an institutional structure of co-optation and repression.

On question B, whether authoritarianism under pure self-interest will tolerably serve the general welfare, the answer more complex and mixed. The ruler's structural payoff dependence—with the tax rate $t$ playing the role of $r_s$ in $tb > c$—gives them incentives to make investments that will increase societal productivity. However, there some important caveats:

1. actual optimisation over $tb > c$ is reserved for the special and non-existent case of the uber-dictator. There are tensions between the pursuit of efficiency and the ruler's strategy for maintaining power.
2. structural payoff-dependence gives the stationary bandit a pure preference for *productivity* and none for the *standard of living*. All human values unrelated to productivity may be sacrificed. High-information dictatorships may be efficient, but highly extractive.
3. rates of leader and/or regime failure remain steady and uncomfortably high. Such coups are frequently bloody and costly, and they shift rulers from behaving as the long-term parasites of Chapter 4 (Box 43) to the short-term.

There is an interesting contradiction in the work of Thomas Hobbes. He argued that the democracy of self-interested actors is impossible, and the general welfare is optimised by broad acquiescence to an authoritarian leader—a Leviathan. We will find support for his arguments about democracy in the next section. Yet his optimism about the Leviathan is wishful thinking. On the one hand, Hobbes is correct that it would be beneficial for the elite to surrender to the Leviathan: this would create an Olsonian uber-dictator, who can relax his/her investments in co-optation and repression and optimise purely over $tb > c$. On the other hand, Hobbes is asking *self-*

[193] Empirically this is made most obvious by contemporary and historical cases where authoritarian rule is imposed by occupying powers—repression and co-optation alone are sufficient to generate social order at least for a short time.

*interested* elites to *commit* to foregoing opportunities for private gain. This requires collective action guided by social norms, i.e. the potent operation of $r_c$. This is inconsistent with the self-interested model, and such commitments can never be credible.

If authoritarianism reigns, the costly Red Queen's race between rulers and their competitors will carry on as it always has, heedless of Hobbes' pleas. Neither the ruler nor the ruled can sleep easily.[194]

The picture becomes more complex when we introduce motivational commitments. The framework suggests that a benevolent ruler, with a significantly public-spirited elite, may be stable and serve the citizenry well. Still, they must look to the future with trepidation. Palace games tend to select for the most cunning and ruthless. No system of rules can prevent the crown from passing, sooner or later, into hands that care little for the public good.

**2 The full relay: The case of democracy**

Democracies, like dictatorships and all other institutional orders, have a top layer of actors with rule-making power that is channelled down through a hierarchical structure of rules. No longer is this topmost layer composed of a narrow elite with mainly *de facto* D&E power; it is instead supposed to be composed of voters granted the *de jure* power to reward or punish regulatory actors, and courts that enforce constitutional constraints on both regulator and voter choices. For democracy to be prosperous, we need voters, and other key actors, to be able to create some approximation of the benign regulator of Part 3, who would run the first two legs of the Red Queen's relay in the public interest, and be motivated to prevent and resolve problems such as the Global Financial Crisis and Opioid Crisis. Like the authoritarian leader, voters need to act as a *universal repair mechanism,* perpetually drawing the regulator towards a favourable social order.

Recall that our overall interest is to see how far a model of pure self-interest, where order emerges from *c*-modifying institutions and $r_s > 0$, can go in constructing a broadly prosperous society. I have already foreshadowed the answer in the case of democracy: it cannot take us any distance at all. The self-interested model will

[194] As Shakespeare has King Henry IV put it, "Uneasy lies the head that wears a crown."

nonetheless allow us to study some important Red Queen dynamics in democratic systems, so is not without interest.

### 2.1.1 Anchorless democracy

In a democracy of self-interested agents, we arrive at a structure that is unsupported by any form of commitment. Self-interest means that $r_c = 0$. As Olson (2000) observes, in the absence of a dictator, no actor has more than a tiny share of total economic output. For the average voter, that share is effectively zero, thus $r_s = 0$ too. Lacking both forms of commitment, voters' payoffs are all defined by:

$$B_i = -c$$

This is a problem because basic acts of participation in democracy such as voting or gathering political information, as well as collective action to defend democratic institutions, are all public goods: privately costly with diffuse public returns. Voters without commitment may prefer they were supplied, but will not contribute to them in the absence of rewards. Following Chapter 8, this problem cannot be solved by rules (Box 14.2).

The problem of the absence of private incentives to vote goes back at least to Condorcet, who observed in 1793 that:

> *where there are a great many voters, each voter's influence is very small. It is therefore possible that the citizens will not be sufficiently interested.* (quoted in Mclean & Hewitt, 1994, p. 245)

The implications of the self-interested model were explored most famously by Downs (1957), producing the Downs paradox. He observes that for voting to be privately beneficial, we need the following to be the case:

$$PB > C$$

Here $P$ is the probability of being the pivotal voter (i.e. of one's vote actually changing the electoral outcome), $B$ is the expected private benefit derived from one's preferred candidate winning, and $C$ is the cost of voting. With $P$ vanishingly small in all major elections, there is no likely scenario where the benefits of voting outweigh its costs. This problem extends to every aspect of democracy that depends on

collective action—a broader "paradox of participation". There is also no incentive to acquire information on candidates, and no incentive to protest or otherwise contribute to the suppression of rent-seeking, the origination of democratic institutions (hence the focus on palace coups in the discussion of authoritarianism), or the defence of existing democratic meta-institutions from capture. The private benefits of democracy are simply too diffuse, and private effort too impotent at the margin. In short, *the self-interested citizenry does not value the vote or democracy*. Voters are themselves Olsonian "roving bandits" .

The Downsian problem is "troubling for economists" and "goes to the heart of any analysis of political economy" (Romer, 1996, pp. 196-197). Fiorina (1990, p. 334) calls turnout is "the paradox that ate rational choice theory". The dominant response in the literature has been, as McLean (2002, p. 540) puts it, "leaving the loose ends of Downsian theory untied". With a wave of the assumption wand, and a generous sprinkling of fairy dust, the problem disappears. As the ancient Greek comic Antiphanes wrote of the use of *deus ex machina* in the theatre,

> *when they don't know what to say*
> *and have completely given up on the play*
> *just like a finger they lift the machine*
> *and the spectators are satisfied.*

Social scientists should not be satisfied with the theoretical equivalent of the arrival of Zeus on high, granting a happy ending. In democracies, there are few more pressing problems to solve than why people do, or do not, contribute to reproducing the political system.

There have, at the periphery, been some theoretical efforts to resolve Downs' paradox while retaining the assumption of self-interest. Each is inconsistent with the empirical evidence. Scholars have suggested that voters are irrational, that voting is zero-cost, that voters participate to signal cooperativity, or that voting is merely "expressive". The latter is the most influential, but it cannot, for example, explain why voters respond strategically to the stakes and closeness of elections. These explanations are rebuffed in more detail in A.14.2. I consider, and rebuff, one apparent solution suggested by Olson in the next section. My intent thereafter is to show that even if these explanations of self-interested voting hold, they are not

enough to make democracy work—the Red Queen's race gives us bigger problems to solve.

**Box 14.2 The impossibility of rule-driven participation**

One may suppose a system of rules, rather like the "repair layer" of Chapter 8, that rewards individuals for an informed vote, for protest in defense of democracy, and so on, thus solving Downs' problem. For reasons discussed at length in Chapter 8, such rules are impossible. Such rules would themselves be within the Red Queen's race, and when they are attacked, no mercenary citizen has the least interest in paying costs to defend and adapt them. They also face codification problems, in that the defense of democracy by its citizenry requires the provision of complex, context-sensitive public goods; we cannot develop a schedule of Pigouvian rewards that reliably makes mercenaries emulate genuinely committed citizens.

### 2.1.2 A useful dead end: McGuire and Olson's "superencompassing interest"

Before we get to the voter's Red Queen's race, let us briefly turn to McGuire & Olson's (1996) analysis of democracy ). Olson (1993) was useful for explaining the "cooperative" behaviour of the materially self-interested dictator via $r_s > 0$, and McGuire and Olson's (1996) extension to democracy seems to provide a way out of Downs' conundrum: the voter majority too, they suggest, is shaped by what we would term $r_s > 0$. Like the dictator, when a rational majority extracts from the minority, it will have at minimum an encompassing interest in the minority—it will limit its rate of extraction, $t$, to that which optimises revenue, and will provide public goods so long as the increase in the amount extracted from the minority exceeds the costs of providing those public goods. Compared to the dictator, however, the majority has even more favourable incentives with respect to economic efficiency. First, the majority's treatment of its own members will be even more efficient. Second, the majority coalition "not only controls the fisc but also earns market income" (McGuire & Olson, 1996, p. 73). When it taxes the minority, it is also taxing its customers and employees. Thus, the majority must wear the deadweight loss from taxation *and* from reduced market income. When it provides public goods, it reaps gains through higher tax revenues *and* through increased market income. The democratic majority

therefore has a *superencompassing interest* in economic efficiency (though as noted in Box 14.3, this result has no special connection with democracy).

The fatal difficulty, and indeed the *deus ex machina* that rescues democracy from Downs' paradox in this case, is McGuire and Olson's decision to treat "the majority or ruling interest" as "an optimizing monolith" (p. 84), or a single interest. They do not long dwell on the significance of this move, or the conditions required to support it. The problem is that the voter monolith can only be spun into existence as the product of *individuals' private decisions* to pay the costs to vote (and, we can add, to become informed, defend democratic institutions, and so on). The monolith is a product of costly collective action. Because self-interested voters as individuals have no incentive to turn out, there is no monolith, and there is no superencompassing interest.

**Box 14.3 Olson's slippery majority**

There is some sleight of hand in McGuire and Olson's (1996) analysis. A "majority" is not necessary for a superencompassing interest, only some aggregate with *dependence on market income*. It is therefore not so much an argument for democracy as an argument for rule by a group with a large share of market income. They acknowledge that the same logic applies to a potentially authoritarian oligarchy that earns considerable market income, and is left in the uncomfortable position of treating oligarchic rulers as the potentially superencompassing "majority".[195] Regardless, McGuire and Olson's logic is sound to the extent that any "majority", if treated as an aggregate, possesses an interest in the economic performance of the minority and the state as a whole that is superior to that of the dictator.

[195] McGuire and Olson (1996, p. 84) write that "[t]here is normally a minority of the society (or, in the case of oligarchic democracies with restricted franchises and 'minority governments,' more than a minority) that is not part of the government." This is a rather obfuscatory way of saying that Olson's economic majority need not be a political majority.

## 2.2 Assuming a zero-cost vote: A deeper grave for the democracy of the self-interested

The above analysis, combined with the observed reality of the active, public-good providing citizenry, is sufficient in my view to reject the self-interested model. But this is no reason to cut the discussion short—there is still more that can usefully be said about how self-interested actors innovate and counter-innovate within a model of democracy that has, at minimum, willing voters.

Let us analyse a secondary case with a *deus ex* machina: the assumption that for most people, the vote has *zero opportunity cost*. Voting is therefore a *private good* with a positive expected payoff. They will turn out, and so they will actually form Olson's majority monolith. Our supernatural sprinklings are not all-powerful, however, and individuals still have zero commitment and will not provide public goods—i.e. the assumption of self-interest will still rule out protest and all other costly modes of collective action. With that, we can press on.

### 2.2.1 The zero-cost voter as sleeping Hare

If they are to domesticate the regulator, voters must run a Red Queen's race against various actors who are innovating to capture it. They must run against roving bandits, i.e. special interests, who seek to gain from rent extraction, from capturing the regulator, and from capturing democratic institutions. They run against regulatory actors themselves, who may have incentives to innovate in the extension of roving banditry. They also run against aspiring stationary bandits, i.e. those in a position to erect a coordinated system of co-optation and repression.

Constraints on space prevent the inclusion of the full and lengthy analysis here. That analysis is an extension of Peltzman's (1976) prototypical model of regulatory capture to meta-institutions, and the interested reader can find it in A.14.3. Here it is enough to point to the main consequences of introducing zero-cost voting; the explanations rely on concepts that may, by now, be familiar and make the argument easy to follow.

First, a crucial observation is that while the vote has positive value, that value is close to zero. Taking a typical large-scale election, Edlin et al. (2007) observe that even if the voter was to reap a $10,000 windfall should their preferred candidate win, the

expected value of the vote is still less than 10 cents. In more typical elections with lower stakes, it will be worth a fraction of a cent. The material cost of being uninformed and voting erroneously is, on average, comparable to that of misplacing a sheet of toilet paper.

Second, in any given electoral contest, the citizen will *vote*, but they will not in practice run the Red Queen's race against special interests. Special interests have large incentives to invest and innovate in strategies of regulatory capture—in buying floor votes, influencing congressional committees, spinning the revolving door, sharing inside information, wielding "structural" power,[196] and much besides. And then they have incentives to invest in campaigns and advertising for captured candidates, to turn them into highly attractive "mimics" of public-spirited candidates. The returns they win from favourable policy are concentrated. The returns to individuals from curbing a particular instance of rent-seeking are so diffuse that they are effectively zero. This is a result familiar to any readers of Stigler (1971) or Olson (1965). The voter, then, will not invest in running the race to distinguish "mimic" from "model".

Third, the "market" for political candidates is a highly complex one. Even if voters are well-motivated, or running the race against special interests was zero-cost, innovation capacities are starkly asymmetric. This is a manifestation of the first-leg (Chapter 11) problem, where even ideally motivated consumers (transacting private goods with large welfare consequences) were unable to detect payoffs. In complex policy areas especially, even well-motivated voters can be expected to lose the race against sophisticated parasitisms. As in Chapter 11, special interests have incentives to magnify apparent complexity (climate disinformation being an obvious case).

Fourth, the parties in power have an incentive to modify meta-institutions in ways that increase the odds of electoral success in the next round. This includes engaging in Schedler's (2002) seven strategies for democratic capture (also in A.14.3). They may disenfranchise opposition voters and gerrymander, increase the power of allied special interests and weaken opposing special interests (e.g. firms versus unions), disempower or even delicense opposing media, and so on. There is a special case in

[196] E.g. large firms, as major employers, wielding the threat to leave, or promise to build in, a jurisdiction to win policy favours.

which this may work in voters' favour: if a party has a comparative advantage in serving voters, then it has incentives to weaken special interest influence.

Fifth, if we modify Peltzman's (1976) model and allow special interests to direct material payments to political actors (an extension he suggested; empirical evidence for such payments is discussed in the next section), then this favourable special case disappears. Political actors have clear incentives to join special interests in racing against the voter. This includes engaging in various means of meta-institutional capture that increase special interest influence. For any political actor, *once elected*—and it is incumbents who shape policy—the optimum is a completely uninformed polity, as this maximises the special interest payments that are consistent with continued political success (again, see A.14.3 for deeper analysis). Political actors then run with special interests against the voter.

Sixth, the zero-cost vote does not translate into zero-cost protesting, or any other costlier political activities that affect incumbents. If incumbents successfully innovate in institutional capture, voters have no material incentive to discipline them. As a final extension of this point, for materially self-interested voters, the non-zero value of the vote carries no implication that voters will value democratic principles such as the universal franchise.[197] Those only emerge from motivational commitments. For the self-interested voter, the *de jure* power granted by democracy has no special moral status; it is like any other resource that individuals will prefer lies in their hands and those of their perceived allies. Under a wide variety of circumstances, the voter may willingly collaborate in regulatory capture—this may be the most privately valuable (if still miniscule) use of the vote. This point is also covered in the A.14.3 extension of Peltzman's model.

To summarise, voters end up in a complex Red Queen's race against the combined forces of political actors and special interests, where there are asymmetric incentives to invest in running the race, and asymmetric access to information and capacities to process it. That race also occurs over meta-institutions, but voters have no meaningful interest in defending those institutions, and indeed may sometimes prefer that they are captured. As democracy is captured, the polity emits a collective

[197] This arises from the domain-incomplete nature of the $r_s > 0$ incentive. Like the dictator, Olson's self-interested but willing voters care only about efficiency, and will engage in any form of extraction (e.g. slavery) that is privately beneficial.

yawn and carries on with its business—or perhaps some may grumble, but only as one grumbles about the weather. The voter is a sleeping Hare, and the sleeping Hare always loses the race.

### 2.2.2 Some empirical observations

Before continuing, let us briefly consider the empirical literature on special interest strategies of influence in democracies. As noted already, these self-interested models cannot possibly explain all empirical patterns (just as they cannot explain basic patterns of voter behaviour). Nonetheless, wherever private returns are large, and self-interest operates without any countervailing other-regarding force, empirical observations may resemble and sometimes match the predictions of such models. In A.14.4, I present a detailed overview of the literature, particularly in the U.S. but including some international evidence. I also examine these matters with respect to influence in the case of the policy changes that led to the GFC in particular. The key papers discussed are listed in Table 14.1.

In some of the matters covered in Table 14.1, early research failed to detect effects—most notably, in the cases of campaign spending on vote share and detection of floor vote-buying. In the case of campaign spending this was primarily a consequence of omitted variable bias, and in the case of vote-buying this was a failure to identify special interests' strategies (e.g. paying to influence less visible congressional committee voting). This is discussed in more detail in A.14.4. With improved specification, the detected effects are frequently large in magnitude. Some key results include:

- First, carefully-specified models find that campaign spending increases vote share. Observed effect sizes are significant: a 15 percent increase in television advertising boosted incumbents' vote share by 1.2 percentage points (Stratmann, 2004), a 10 percent increase in mailed political pamphlets increased vote share by 1.5 to 3.5 percentage points (Gerber, Kessler & Meredith, 2011), and a 1 percentage point increase in a party's share of contributions generates a half percentage point increase in its share of the legislature (Hall, 2016).
- Second, successful electoral candidates win significant private spoils in excess to their salary. For example, the wealth of those who win Senate races

**Table 14.1.** Significant empirical papers on firms' and representatives' rent-seeking strategies in democracy.

| | **Significant papers** |
|---|---|
| **Special interest investment in influence varies strategically** | Hart (2003), Drutman (2010), Hill et al. (2013), and Brulle (2018). |
| **Campaign spending increases vote share** | Green & Krasno (1988), Abramowitz (1988), Grier (1989), Snyder (1990), Gerber (1998), Erikson & Palfrey (2000), Stratmann (2004), Stratmann (2005), Gerber, Kessler & Meredith (2011), and Hall (2016). |
| **Representatives win private returns from serving firms** | Diermeier, Keane, & Merlo (2005), Eggers & Hainmueller (2009), Fisman, Schulz & Vig (2014), Palmer & Schneer (2015), Huang & Xuan (2016), and Fafchamps & Labonne (2017). |
| **Firms win excess returns, evade policing, or win other advantages by investing in the political process** | Jayachandran (2006), Khwaha & Mian (2005), Faccio (2006), Goldman, So & Rocholl (2006; 2009), Faccio et al. (2007), Claessens et al. (2008), Ferguson & Voth (2008), Kim (2008), Faccio & Parsley (2009), Krosner & Strahan (1999), Richter et al. (2009), Cooper et al. (2010), Yu & Yu (2011), Igan, Mishra & Tressel (2012), Duchin & Sosyura (2012), Blau, Brough & Thomas (2013), Goldman, Rocholl & So (2013), Hill et al. (2013), Bertrand, Bombardini & Trebbi (2014), Chen, Parsley & Yang (2014), Akey (2015), and Acemoglu et al. (2016). |
| **Firm strategies are designed to evade voter detection, e.g. vote-buying versus participation-buying** | Hall & Wayman (1990), Wright (1990), Grier & Munger (1991), Snyder (1992), Romer & Snyder (1994), Milyo (1997), Krosner & Stratmann (1998), Stratmann (1998), Engel & Jackson (1998), Stratmann (2002), Tripathi, Ansolabehere & Snyder (2002), Fellowes & Wolf (2004), Krosner & Stratmann (2005), Stratmann (2005), Figueiredo & Silverman (2006), Witko (2006), Mian, Sufi & Trebbi (2010), Suarez & Kolodny (2011), Hill et al. (2013), Igan & Mishra (2014), Krosner & Strahan (2014), Powell & Grimmer (2016), and Akey (2015). |

increases by US$1.7 million relative to narrow losers (Diermeier et al., 2005), ex-senators gain US$125,000 per year on average from serving on firm Boards (Palmer & Schneer, 2015), and in the UK the wealth of Conservative MPs doubles after winning office compared to narrow losers (Eggers & Hainmueller, 2009).

- Third, these investments are rational for firms. For example, US$1 spent on lobbying leads on average to a US$6-20 reduction in tax burden the next year (Richter et al. 2009), lobbying firms are 38 percent less likely to be detected committing fraud, fraudulent firms spend nearly twice as much on lobbying (Yu & Yu, 2011), and various studies find large abnormal shareholder returns relating to political investments.
- Finally, patterns of political influence reflect a contest over D&E and its evasion. Firm payments are timed to punish or reward representatives (e.g. Engel & Jackson, 1998 and Stratmann, 2002). Firms buy representatives' floor votes where issues are not salient to voters, but on salient issues prefer to buy less-visible votes in the congressional committees that shape the legislation brought to a floor vote (e.g. Snyder, 1992, Fellowes & Wolf, 2004, Witko, 2006). This fits well with our extension of Peltzman (1976), given representatives face a softer trade-off if special interest influence can be concealed.

Paired with the case studies of Part 3, and the summary of evidence for growing economic rents as a share of GDP in the introduction of Chapter 13, such evidence suggests that U.S. voters are losing the Red Queen's race over regulatory capture. There is no sign of efficient Coasian bargaining. Rather, it is consistent with an economic system moving towards an uncoordinated anarchy of roving banditry. This creates opportunities for new coordinating powers to rise in their place, a matter I discuss in more detail towards the end of the thesis.

### 2.3 Generals and judges in the democratic selectorate

Now I turn briefly to the other major meta-institutional actors, and possible bulwarks, in democracy. Elected representatives' consolidation of power, and both regime change and decay, may be blocked by constitutional constraints—if those constraints are:

- enforced by the judiciary; and
- secured by military actors that hold a monopoly on violence.

Let us consider the efficacy of these constraints under the assumption of self-interest.

### 2.3.1 The self-interested military

The military of self-interested actors is a hierarchy of mercenaries that views the present political configuration as the one, among available options, that provides it with the largest rents. It will opportunistically press for larger rents, as far as engaging in a coup, as circumstances allow. Coups and the problem of coup-proofing are interesting as the most extreme manifestation of the problem of aligning militaries with voter interests. Pilster & Böhmelt's (2012) study of coup-proofing in democracies and dictatorships professes an analytic approach based in self-interest: they exclude the effects of "ethos" (i.e. commitment via $r_c$), and accordingly "restrict [the] analysis to governments' institutional coup-proofing" (p.357-358). Yet their argument is ultimately that democracies can invest less in coup-proofing than dictatorships, and so have more effective militaries,[198] because they are

> *characterized by mass political participation, strong civil societies, and publicly accepted formulas for power transfers or political change. Any military would therefore face substantially higher governance costs after toppling a democratic regime…* (p. 359)

Further, they suggest that coups are restricted because military actors themselves regard democracy as "justified". All such explanations are, of course, based in ethos and $r_c > 0$.

In a consistent fully self-interested model, collective action problems prevent the citizenry from resisting military actors and the military is unconcerned with justice. Thus, the military plays the power games of the palace, as in North's (1979) unswerving self-interested analysis of authoritarian systems. The self-interested democracy would need the same kind of coup-proofing mechanisms as authoritarian systems. A key task for voters would be ensuring that military actors reap larger rents under elected representatives than they would by interfering with or capturing the system. Voters, representatives, and special interests would have to be careful to find ways to serve the interests of military actors—or find ways to keep them in conflict.

[198] The authoritarian strategy for coup-proofing typically "divides a country's military manpower into rivaling organizations" (Pilster & Böhmelt, 2012, p. 358), a blunt structural strategy employing discoordination and conflict to block the prospect of efficient rent-sharing by military actors.

In practice, incomes for military leaders are very modest,[199] and Pilster & Böhmelt (2012) note the absence of coup-proofing in democracies. And yet, in the long-established democracies, military coups are almost inconceivable.

### 2.3.2 The self-interested judiciary

The self-interested judge is a mercenary who makes legal decisions based on evaluations of private gain. They are opportunists who will provide legal favours whenever the payoffs are favourable. In the society of self-interested actors, they face no threat of collective action from the public.

With respect to defending meta-institutions, or collaborating in their capture, the judiciary will follow expected private payoffs. The optimal strategy is to avoid offending the major blocs of political (and military) power that presently hold power or may do so in the future. In the favourable climate of a balance of political power, and a balance of control over the means of violence, the judiciary may opt to defend the status quo. Once again, however, balances of power do not last; as the Red Queen's race unfolds, one group or another will win advantages and the landscape of payoffs shifts. The judge will rationally prefer to remain a valued member of the winning bloc, and so acts as part of the positive feedback mechanism that cements the power of a dominant coalition.

In short, the self-interested judge is much more likely to *reflect* the distribution of power, rather than *constrain* it. One may intuit, along the lines of Pilster & Böhmelt's (2012) analysis of the military, that violations of constitutional rules would cause outrage and impose meaningful costs. But this is to summon the *deus ex machina* another time, as an extra-theoretical force that makes the model fit the empirical reality (i.e. a reality of strong popular responses to perceived judicial corruption). In a world without moral commitments, rational actors would recognise that institutions have no moral content, but are simply mechanisms for ordering

[199] The wage of the US Secretary of Defense was $210,700 per annum in 2018, while four-star generals and the Chairman of the Joint Chiefs of Staff are compensated US$197,302 per annum. We can easily conceive of neighbouring worlds in which these military actors are able to leverage a much larger share of the economic surplus. The *Washington Post* reports that President Trump assumed top brass were paid on the order of US$5 million (Dawsey & Paletta, 2018)—still a paltry share of national GDP for a military chief in a typical authoritarian regime.

**Box 14.4 Self-interested judges in the field**

Self-interested behaviour among judges can, of course, be observed empirically. In well-established democracies, such patterns occur mainly at the margins; the problems raised by private campaign finance, for example, extends to the election of judges in some US jurisdictions, with special interests sometimes funding the campaigns of judges that later decide their cases (e.g. Corriher, 2012). At least one sitting member of the US Supreme Court has admitted to decades of undisclosed "lavish" gifts from a politically active billionaire friend with an interest in cases that have come before the court (VanSickle, 2023). Cases of democratic decline in the twenty-first century have typically involved explicit and concerted efforts to install favourable judges who might distort democratic processes (e.g. Kovács & Scheppele, 2018, on Poland and Hungary). Captured judges reinterpret the law in ways that are favourable for the party that has consolidated power, and those justices who are sufficiently "irrational" to resist are eventually co-opted or removed.

Once again, however, we are confronted by the empirical reality that judicial actors are frequently a bulwark against meta-institutional capture, and sometimes take great private risks in resisting emerging authoritarians.

behaviour in the interests of those who hold underlying advantages in co-optation and repression.

### 2.4 The democracy of self-interest: The anarchy before the Leviathan

The answers to the questions that opened this chapter are negative. Democracy cannot, under pure self-interest, (A) establish a coherent social order that (B) may plausibly serve the general welfare. It cannot produce a regulator that will run the first two legs of the Red Queen's race in the general interest.

In the third leg of the Red Queen's relay—a costly race against special interests—voters will play the sleeping Hare. Decay cannot be arrested by constitutional and other institutional constraints, none of which are self-enforcing. There is not a single political, judicial, or military actor committed to the enforcement of these constraints, outside of temporary conditions that make them advantageous. The real players are those actors with sufficient interests and capabilities to coordinate strategies of co-optation and repression, particularly actors with access to violence.

Again, the absence of collective action means that they may pursue the palace games with impunity.

Fittingly, the analysis returns us to the conclusions of Thomas Hobbes. As Apperley (1999) notes, Hobbes regarded democracy as fundamentally unstable, and "the unstable democratic form will either collapse back into the anarchic state of nature" or transform into one of the stable forms of government (p. 167). For Hobbes, democracy is an anarchy of roving banditry before the return of the Leviathan:

> *a democracy is by institution the beginning both of aristocracy and monarchy*. (Hobbes, 1640/2019, p. 22)

Hobbes understands that democracy may temporarily be sustained by "the mutuall feare of equall factions" (1651, p. 123), i.e. balances of power. But he recognises that the unpredictable unfolding of competition—our Red Queen's race—give such balances of power a short lifespan. Its destiny is "frequent Seditions", "Civill Warres", and the relentless continuation of the "warre of all against all" (1647/1983, p. 34) until a winner emerges.

To summarise in terms of our framework, any system where $r = 0$ is an interregnum of uncoordinated roving banditry, where no actor has an abiding interest to innovate in sustaining the rules-as-public-goods that undergird social order. Amid this anarchy, some actors will find ways to assemble systems of coordinated payoff transformation. They replace the interregnum with a new regime where $r_s > 0$. In a world of self-interest, the stationary bandit is the only port in the storm—although a port prone to periodically falling beneath the waves.

*…love thy neighbour as thyself…*
Leviticus 19:18

*How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it.*
Adam Smith, 1759, p. 1

*There can be no doubt that a tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage, and sympathy, were always ready to aid one another, and to sacrifice themselves for the common good would be victorious over most other tribes; and this would be natural selection.*
Charles Darwin, 1871, p. 159

# 15

# The moral sentiments: Smith, the strong reciprocator, and the empirical evidence

The analysis in Chapter 14 fits neatly into the sociobiology-inspired framework of this thesis: rules and commitment are essential complements. In a self-interested world, structural payoff-dependence would be the only form of commitment, and all regimes would be authoritarian. On this analytic point, Hobbes (1651) was right.

Yet Hobbes was wrong in practice. Democracies can survive and thrive. The fact that many democracies have been highly successful and stable tells us that they must be utilising some source of commitment. Structural payoff dependence is the only *extrinsic* reason to be interested in others' payoffs. Lacking that, the only explanation left is the *intrinsic* motive: people must have preferences over the outcomes of others. This preference is formalised in the weighting $r_c$.

The analysis so far suggests only one thing: that motivational preferences drive at least some people to invest in the key public goods of democracy. It does not tell us anything detailed about the nature of those preferences—what or who it is that

people intrinsically value, and how such preferences strengthen or weaken in different circumstances. A simple model of democratic capitalism with $r_c > 0$ might be interesting. But we can do better with a model enriched by empirical studies of non-selfish motivations.

To that end, this chapter turns to modern research on human moral sentiments. There are strong limitations on the scope of this work. First, the literature on other-regarding preferences is still developing, and future research will nuance, extend, or supplant some of the arguments in these last few chapters. Second, the literature is large, and its implications cannot possibly be fully explored here. I will focus on the dominant theory of humans as "strong reciprocators". That this theory has been so successful is a boon for this thesis, because the moral agents it describes directly address—indeed, are very interested in—the problems raised in Chapter 8. These are *committed actors* who are *willing to invest in rules that punish defectors*. It is the adaptive pattern of moral preferences that the framework would predict.

In all this, we will have the good company of Adam Smith. At the birth of the fields of economics and political economy, Smith gave us a theory of social order, and of the market order, based in other-regarding preferences. The pattern of preferences he described is remarkably consistent with the pattern of evidence I describe here: humans are sympathetic, if parochial, lovers of law. This pattern of morality makes social order possible, but, as Smith warned, it may also drive much more intensive social conflict than that possible with merely self-interested actors.

Section 1 introduces Smith's social order resting on sympathy, and four propositions we can take from his work that will be supported in this and the following chapters.

Section 2 surveys the empirical evidence for other-regarding preferences, which take the special form of "strong reciprocation". Humans are prosocial lovers of law, supplying the two ingredients of social order emphasised in Chapter 8.

Section 3 connects strong reciprocity to *trust*, or expectations about others' cooperativity. In a world of strong reciprocators, trust becomes roughly the meaning of $r_c > 0$. This will help us to explain the possibility of democratic order in Chapter 16.

Section 4 turns to examine two vulnerabilities: the *parochial* nature of prosociality, and the problem of the *low-trust trap*. Both will be important in the Chapter 17 exploration of democratic decline.

### 1.1 Other-regarding preferences: The ground of Smithian order

In analyses of social order, the primary question has tended to be how to understand liberal societies given actors are self-interested. It is curious that Hobbes' psychological egoism survives despite lengthy refutations by liberal theorists who have otherwise had more influence on the tradition, including Smith and Hume.[200] It is more curious still that self-interest came to colour the way we think about Smith's market society. In *Wealth of Nations,* Smith provided a potent explanation of how market incentives may generate *economic* order by self-interest alone. Yet his theory about the foundations of the *broader* social order, within which the market is a subsidiary order, placed particular emphasis on humankind as a moral species with the capacity for "love", "resentment", and a sense of "justice". It is worth returning to this lengthy passage near the outset of *The Theory of Moral Sentiments*:

> *…man, who can subsist only in society, was fitted by nature to that situation for which he was made. All the members of human society stand in need of each other's assistance, and are likewise exposed to mutual injuries. Where the necessary assistance is reciprocally afforded from love, from gratitude, from friendship, and esteem, the society flourishes and is happy. All the different members of it are bound together by the agreeable bonds of love and affection, and are, as it were, drawn to one common centre of mutual good offices.*
>
> *But though the necessary assistance should not be afforded from such generous and disinterested motives, though among the different members of the society there should be no mutual love and affection, the society, though less happy and agreeable, will not necessarily be dissolved. Society may subsist among different men, as among different merchants, from a sense of its utility, without any mutual love or affection; and though no man in it should owe*

[200] Other notables of the time being Shaftesbury, Hutcheson, and Butler.

> *any obligation, or be bound in gratitude to any other, it may still be upheld by a mercenary exchange of good offices according to an agreed valuation.*
>
> *Society, however, cannot subsist among those who are at all times ready to hurt and injure one another. The moment that injury begins, the moment that mutual resentment and animosity take place, all the bands of it are broke asunder, and the different members of which it consisted are, as it were, dissipated and scattered abroad by the violence and opposition of their discordant affections. If there is any society among robbers and murderers, they must at least, according to the trite observation, abstain from robbing and murdering one another. Beneficence, therefore, is less essential to the existence of society than justice. Society may subsist, though not in the most comfortable state, without beneficence; but the prevalence of injustice must utterly destroy it.* (pp. 103-104)

Of course, the purely self-interested actor *is* willing to hurt and injure another, for even the most trifling private gain. In the eyes of *Homo sapiens*, a specimen of *Homo economicus* would appear to be a psychopath. In Smith's estimation, a society composed of psychopaths cannot long survive.

The indispensable foundation for social order is, in Smith, our interest in "justice"—an aversion to, and willingness to punish, actions that are harmful to others and violate negative duties. Smith is clear, however, that justice is not simply the product of rules. Its essential foundation is *sympathy*, which leads us, unlike the psychopath, to personally suffer from the privations of others. Smithian sympathy is related, although a little different, to the modern conception of "empathy"; it is in any case a sophisticated model of other-regarding preference (see Box 15.1). In contemporary terms, justice is a public good. Individuals cannot reap sufficient private rewards from their own investments in building and enforcing a system of just rules, given the benefits are diffuse. Sympathy turns justice into a partly private good.

Smith's merely just society may "subsist" and "not necessarily be dissolved". Greater degrees of sympathy support "beneficent" actions, which go beyond the requirements of justice and favour prosperity. They are positive duties, such as the gratitude that

ought to follow a favour, the effort to fulfil the spirit of an incomplete contract, contributions to certain non-essential public goods, and charity. Yet while Smith remarks that the selfish and miserly fairly deserve our "hatred",[201] beneficence may generally not be extracted by force. But even here, he allows for some limited enforcement of beneficence by the state.

> *Of all the duties of a law-giver… this [the compulsion of beneficence], perhaps, is that which it requires the greatest delicacy and reserve to execute with propriety and judgment. To neglect it altogether exposes the commonwealth to many gross disorders and shocking enormities, and to push it too far is destructive of all liberty, security, and justice.*[202] (pp. 98-99)

For Smith, the rules of the efficient market, and of the prosperous society, are necessarily the product of "public-spirited" people driven by sympathy. Here we call them motivationally committed actors. Smith is not alone among economists in arguing that society depends on other-regarding preferences. Joseph Schumpeter (1942), and more explicitly Karl Polanyi (1944) and Fred Hirsch (1976), developed the argument that the capitalist order stands on the shoulders of pre-capitalist values. I noted similar remarks from Arrow (1972) in the introduction. It is curious that this thread of liberal thought, so prominent at the inception of political economy, has become hidden behind the narrower analysis of incentives within markets. I will say a little more about why this occurred in the following chapters.

### 1.2 Four key propositions

I identify four main propositions in Smith, all of which will be supported in these last few chapters of the thesis.

1. The *proximate* cause of human prosociality is an innate tendency for sympathy, in combination with tendencies to internalise, and parochially defend, cultural norms. This is the subject of this chapter.

[201] "Those whose hearts never open to the feelings of humanity, should, we think, be shut out, in the same manner, from the affections of all their fellow-creatures, and be allowed to live in the midst of society, as in a great desert where there is nobody to care for them, or to inquire after them." (p. 99)
[202] Libertarianism lies on the neglectful extreme, with its "gross disorders and shocking enormities". Theocracy and communism push mandated moral behaviour too far, destroying "liberty, security and justice".

**Box 15.1 Smithian sympathy, culture, and utility monsters**

Adam Smith's "sympathy" is a fascinating variety of other-regarding preference over welfare, where the domain is constrained by cultural norms. Utilitarianism can, as Nozick (1974, p. 41) pointed out, produce "utility monsters": if some person experiences a stupendous explosion of utility every time they are given a widget, the ardent utilitarian will insist that all the engines of human society be turned towards widget production.

This is, of course, not how human morality works in practice. Smith's is a worldly, naturalistic model of the moral sentiments. In this view, utilitarian theory and its monsters, along with Kantian and Nozickian categorical imperatives, are little more than cartoons of a more complex reality. As Sen (2009) observes, the moral theories that supplanted Smith abandoned naturalism for the project of logically deducing rules of perfect moral conduct. Smith was, again, ahead of his time; only in recent decades have moral theorists seriously turned towards morality as a positive, rather than normative, field of study.

Moral rules are the product of natural human sympathies, and Smith observes that our capacity to sympathise is constrained in ways that help it to serve individuals and society. If the utility monster was a person walking among us, no ordinary human could sympathise with, or enter into, its pleasures. We may not necessarily regard the utility monster as immoral, given its pleasures cause little harm, but the dominant human attitude would be that the monster exhibits norm-violating personal defects. The utility monster would be suitable for a travelling circus, where patrons could pay a small fee to marvel—and recoil—at its paroxysms of widget-induced pleasure.

Smith argued that human morality is "fitted by nature" to the social condition. This fitting process led us to care about others whose feelings are, in our social group, viewed as appropriate and compatible with the furtherance of the group. All utility monsters are utterly alien to us, and a liability for the group—the fewer the better.

2. The *ultimate* cause of prosociality is that humankind was "fitted by nature" for the social condition. This is the modern understanding, built from studies of the evolution of human morality. This is discussed briefly in Section 4.
3. Prosociality is critical for social order, and it further requires some degree of *impartiality*, or homogeneity in the weights assigned to different groups.

Chapter 16 will examine the role of *generalised* prosociality in building liberal democracy.

4. Parochial prosociality, Smith's "faction" and "fanaticism", is the greatest threat to social order. Ingroup payoffs may be weighted much more strongly than outgroup payoffs, leading to destructive between-group conflicts. Chapter 17 will support Smith on this matter.

### 2.1 *Homo sapiens* as "strong reciprocator"

Purely self-interested agents, indeed psychopaths, may be "social" in that they may "truck, barter, and exchange" (Smith, 1776/1976, p. 14) with others for private advantage, and engage in strategic reciprocity of the kind observed in the repeated Prisoner's Dilemma. The reciprocity of the self-interested agent is termed *weak reciprocity*. The name may be misleading if it is taken to indicate even a shred of benignity; it is reciprocation based purely on the assessment that one cannot successfully exploit the other, and that cooperation is presently the optimal selfish strategy. The weak reciprocator therefore defects in one-shot games and the final round of iterated games, and will impose any degree of harm on another actor if it is to their own private benefit.

The "prosocial" agent, by contrast, is willing to pay some costs to benefit others or avoid harming others, even where these costs may never be recouped. An extreme case is *unconditional altruism*, where the willingness to benefit others at private cost does not depend on any features of the context or the interaction partner. While this is a common conception of what it means to be non-selfish, it is a caricature that is only rarely observed in experimental games (although it is, in fact, only a little rarer than pure self-interest). Unconditional altruism is evolutionary unstable,[203] because altruists are, in the technical jargon of cooperation theory, "suckers". Not only do they fail to compete against cheaters, but their naïveté underwrites the success and proliferation of cheaters. From the perspective of a group, the unconditional altruist is a liability.

[203] Unless there is permanent intense group selection and zero within-group competition.

A much more significant form of prosociality is a form of *conditional* altruism termed *strong reciprocity* (Gintis, 2000, Henrich et al., 2001; Gintis et al., 2003; Gintis et al., 2008). Gintis et al. (2003) explain that it entails

> *a predisposition to cooperate with others and to punish those who violate the norms of cooperation, at personal cost, even when it is implausible to expect that these costs will be repaid either by others or at a later date.* (p. 154)

In contrast to weak reciprocity, strong reciprocity produces "non-selfish" cooperation "not only in repeated interactions, but also in one-shot games" (Gächter, Kölle & Quercia, 2017, p.3). Unlike unconditional altruism, strong reciprocity can be an evolutionarily stable strategy (or population Nash equilibrium) in a much broader range of environments, because conditionality allows the exclusion of defectors. It is particularly powerful for explaining the empirically observed ways in which human behaviour departs from the standard economic model. The evidence to be surveyed suggests that strong reciprocity is a good summary of "average" cooperative behaviour in our species, across the lifespan and across different cultures.

As a point of no small interest, this is a good fit with Smith's (1759/1976) model of innate human morality: strong reciprocity is "the great law which is dictated to us by nature" (p. 99), wherein we feel benevolence is due to benevolence and punishment due to harm. Strong reciprocators are interested in justice (and to some degree beneficence). Strong reciprocators are an answer to the exigencies of the Chapter 8 analysis: they are *committed actors who protect themselves with rules*.

### 2.2 Innate prosociality

Smith (1759/1976), for his part, viewed other-regarding preferences as "a matter of fact too obvious to require any instances to prove it" (p. 13).[204] Obvious or not, today a large body of research offers us much greater resolution on the patterns that such preferences take, and indeed evidence that they are partly innate.

[204] This remains a common view among laypeople, whose intuitions are untrained by economic theory, and perhaps helps explain why laypeople are more generous in economic games than economists.

It is instructive to begin by considering how other higher animals are indeed "fitted by nature" for their social conditions. Varying levels of prosociality appear genetically encoded. To find *Homo economicus*, we can do little better than look to the average chimpanzee. Outside of kin, chimpanzees show a general absence of other-regarding preferences and display neither altruism nor spite (Jensen et al., 2006, Jensen et al., 2007a, and Silk et al., 2005). Reciprocity is of the purely strategic or "weak" form, and they ignore opportunities to aid others even at very low or zero cost (Hare & Tomasello 2004; Silk et al. 2005; Jensen et al. 2006; Vonk et al. 2008). In the Ultimatum Game, chimpanzees make very low proposals and responders accept what they are offered (Jensen et al., 2007b). By contrast, marmosets, tamarins, bonobos, and a few other primate species show evidence of generalised, strong reciprocity independent of strategic private payoffs (Bukart et al., 2007, Cronin et al., 2010, and Tan & Hare, 2013). Brucks & von Bayern (2020) have undertaken similar studies among birds, focusing on the African grey parrot and blue-headed macaw—among the most intelligent of non-human species. These birds are adapted to different social conditions, such that African grey parrots are prosocial and blue-headed macaws are self-interested.

There is also natural variation between individuals in any given species, including in *Homo sapiens*. People may be genetically inclined or disinclined to prosociality. Studies comparing identical twins against non-identical siblings, for example, indicate that tendencies for prosocial behaviour are heritable (Ariel & Robert, 2006).

Consider research on the extremes of the prosociality bell curve, and their organic neurological bases. Around one percent of humans are diagnosable as psychopaths (Richerson & Henrich, 2012), and they join chimpanzees as exemplars of *Homo economicus*. Unlike individuals diagnosed with sadistic personality disorder, who will pay costs to harm unknown others (Buckels et al., 2013), psychopaths simply lack the capacity to form preferences about the well-being of others. They are unable to love or form genuine relationships (Cleckley, 1976; Meloy, 1988; Jonason & Schmitt, 2012) and, quite fittingly given the themes of this research, are described as having "a parasitic orientation toward others" (Marsh, 2013, p. 2). Psychopathy is associated with heritable brain abnormalities, most notably a small amygdala, and is tied to neural deficits in emotional processing of others' suffering (Decety et al., 2013; Bird & Viding, 2014).

At the other end of the spectrum, Marsh et al. (2014) study individuals they term "extraordinary altruists"—people who may, for example, donate a kidney to a complete stranger despite the long recovery period and chance of mortality.[205] Similarly, Fagin-Jones & Midlarsky (2007) studied non-Jewish rescuers of Jews during the Holocaust, and found that rescuers were distinguished from bystanders on measures of social responsibility, empathic concern, altruistic moral reasoning, and risk taking; the rescuer average on these measures differed from bystanders by around three standard deviations. Extraordinary altruists' brain structures and patterns of neutral responses are roughly the inverse of psychopaths: a larger amygdala, greater responsiveness to emotional expressions of suffering, and greater overlap between neural responses to distress of self and other (Marsh et al. 2014; Crockett & Lockwood 2018).

The observed overlap between neural sensations of pain for those of self and other, now understood to be mediated by "mirror neurons" (Singer et al., 2004, p. 1160, in the seminal study), is the modern elaboration of Smith's observations in the *TMS*:

> *By the imagination we place ourselves in his situation, we conceive ourselves enduring all the same torments, we enter as it were into his body, and become in some measure the same person with him, and thence form some idea of his sensations, and even feel something which, though weaker in degree, is not altogether unlike them.* (p. 13)

### 2.3 The hump of the bell curve: Some results from experimental games

In the hump of the bell curve, between the poles of the psychopath and extraordinary altruist, are the rest of us—equipped with mirror neurons, functional amygdalas, and Smith's capacity for sympathy, but not so extraordinarily. Decades of game theory has shed considerable light on the shape of the prosociality bell curve in adults (see A.15.1 for studies of prosociality in children, even in preverbal infants). Let us focus on one-shot games that are suited to self-interested defection.

The Dictator Game (Forsythe et al. 1994) entails a giving decision without the possibility of private returns. It is therefore a measure of strong reciprocity, where

[205] They are also much more likely to donate money, volunteer, and engage in other altruistic behaviours (Henderson et al., 2003).

strong reciprocators are expected to be generous if they believe that the average stranger is a cooperator type—that is, if they generally trust others. Engel's (2011) meta-review finds an average offer of nearly one third. Around one tenth of "dictators" follow the predictions of the self-interested model and give zero. Twice as many split the pot 50/50, and one in twenty "dictators" give everything.

However, as Engel (2011) notes, the preponderance of studies focusing on students has given misleading results. Non-student adults cooperate at much higher rates: the most common response is a 50/50 split, the second most frequent is to give everything, and again 10 percent give nothing. Finally, Engel finds that if a recipient is "deserving" (e.g. a charity or needy and trustworthy individual) the most frequent response is to give everything (also see Eckel & Grossman, 1996).

In the Trust Game, "trusters" can send money to a "trustee". In turn, the trustee decides whether to send money back. Sums transferred by either party are doubled for the recipient, hence cooperation is socially efficient. The self-interested model predicts that the truster will never send money, because the trustee will never return money. However, more than 50 percent of trustees return money to the truster, sufficient to make trusting worthwhile on average; consistent with strong reciprocity, amounts returned increase with the size of the initial transfer (e.g. Fehr et al., 1993; Berg et al., 1995; and Hayashi et al. 1999). In one particularly important finding for grounding these experiments in reality, this pattern is unchanged even where stakes are very high—equal to 2-3 months of income (Fehr et al., 2002).

The one-shot Ultimatum Game provides a test of strong *negative* reciprocity—reciprocity that involves paying a private cost to punish a defector, rather than reward a cooperator. The "proposer" is given a lump sum and chooses how much to share with a "responder". If the responder accepts the offer, both keep the money. If the responder rejects it, neither keeps any money. By rejecting an offer that is non-zero but unfair, the responder pays a private cost to punish the proposer and so enforce a social norm of fairness. The proposer's initial decision reflects their expectations about the responder's propensity for strong negative reciprocity (and the proposer's own assessment of what is fair). Fehr & Schmidt (1999) estimate that around 10 percent of responders insist on a roughly even split, 30 percent insist on at least a third, 30 percent insist on at least a quarter, and the remaining 30 percent

accept very low offers in line with *Homo economicus*. On average, proposers offer around 40 percent of the endowment, and so most proposals are accepted. Strong negative reciprocity can also be tested via games that incorporate costly third-party punishment. Fehr & Fischbacher (2004) employ the Dictator Game and Prisoner's Dilemma, and find that around 60 percent of participants punish third parties for violating fairness norms, despite the impossibility of private gain from doing so. This was sufficient to eliminate the rewards from defection.

See A.15.2 for extensive evidence on the effect of strong reciprocity in the field, on economic matters from contract formation, to market clearing, persistent inflation, job performance, and wage premia. There I discuss psychological research that places prosociality as a sixth personality factor alongside the "Big Five", with the same degree of orthogonality, stability, and heritability as the other traits, and predictive power in cooperative games and the field. Prosociality is, for example, the most significant personality trait for predicting job performance (Johnson, Rowatt & Petrini, 2011). This reflects a pattern the framework would predict: for firms, controlling shirking behaviour via rules requires running a costly and difficult Red

**Box 15.2 Interactions between psychology and environment**

Prosociality is of course modified by cultural factors—in Smith (part V, ch. 2), by "custom and fashion"—and life experience.[206] Psychopathy, for example, is influenced by genetic predispositions *plus* environmental effects such as abuse and neglect (Hicks et al., 2012). The cross-cultural studies of Henrich et al. (2001) and Richerson & Henrich (2012), among many others, show that cultural beliefs, norms, and economic conditions influence when and how prosocial tendencies are activated—people who live in societies where public goods are important in social interactions, for example, are more prosocial in the Public Goods game. While patterns of behaviour are diverse, they observe that no known human society can be reconciled with the self-interested model. Average cooperativity across these games is similar to that in modern societies—sometimes lower and sometimes higher (e.g. Engel 2011). A key question is what cultural forms are required to stabilise liberal societies specifically, and prosperous large-scale societies in general.

Queen's race. Committed workers, meanwhile, are industrious even in the absence of punishment.

### 3 Strong reciprocity and expectations: Trust

The strong reciprocator only exhibits commitment to the welfare of others who they *expect to be cooperators*. They refuse to be suckers. The strong reciprocator with low expectations may be indistinguishable from a self-interested actor (although they might engage in costly punishment of defectors). The strong reciprocator with high expectations may sometimes be indistinguishable from an unconditional altruist. Thus, the fact that most individuals are strong reciprocators does not mean actors are committed. To get $r_c > 0$ requires strong reciprocation *plus* favourable expectations (Figure 15.1).

We can call these favourable expectations "trust", in the sense of the literature on motivational trust.[207] Delhey & Newton (2005, p. 311) describe trust as "the belief

[206] Kosse et al. (2020) find that prosociality in elementary school children is associated with SES, but the gap can be closed by simply assigning low-SES children a mentor.
[207] Note there is an alternative definition of trust in the literature, defined as the belief that an individual is acting within an incentive structure that will lead them to serve our interests (e.g. Levi & Stoker, 2000). For example, one may "trust" an employee to do their job because their every move is

that others will not deliberately or knowingly do us harm, if they can avoid it, and will look after our interests, if this is possible" (p.311, also see Chiles & McMackin, 1996, Maguire et. Al., 2001, and Nooteboom, 2002, 2007). Bradach and Eccles (1989) highlight the effects of trust in relation to opportunism: trust is "a type of expectation that alleviates the fear that one's exchange partner will act opportunistically" (p. 104). That is, we trust that the other actor will not defect even when (A) it is privately favourable to do so and (B) there are no effective *c*-modifying rules, i.e. we cannot detect bad behaviour and enforce a punishment.

Measures of trust predict whether individuals will cooperate in social dilemmas, so that trusting agents are trustworthy agents (Orbell & Dawes 1993; Glaeser et al. 2000; Thöni, Tyran & Wengström 2012; Peysakhovich & Rand, 2016). This reflects the pattern in Figure 15.1, and supports an early observation of Smith in the *TMS*, "We trust the man who seems willing to trust us." (1759/1976 p. 399) See A.15.3 for further research connecting trust and strong reciprocation.

under close surveillance. Following Maguire et. al. (2001), this is better described as "control". The fact that the employee must be so closely controlled indicates the *absence* of trust.

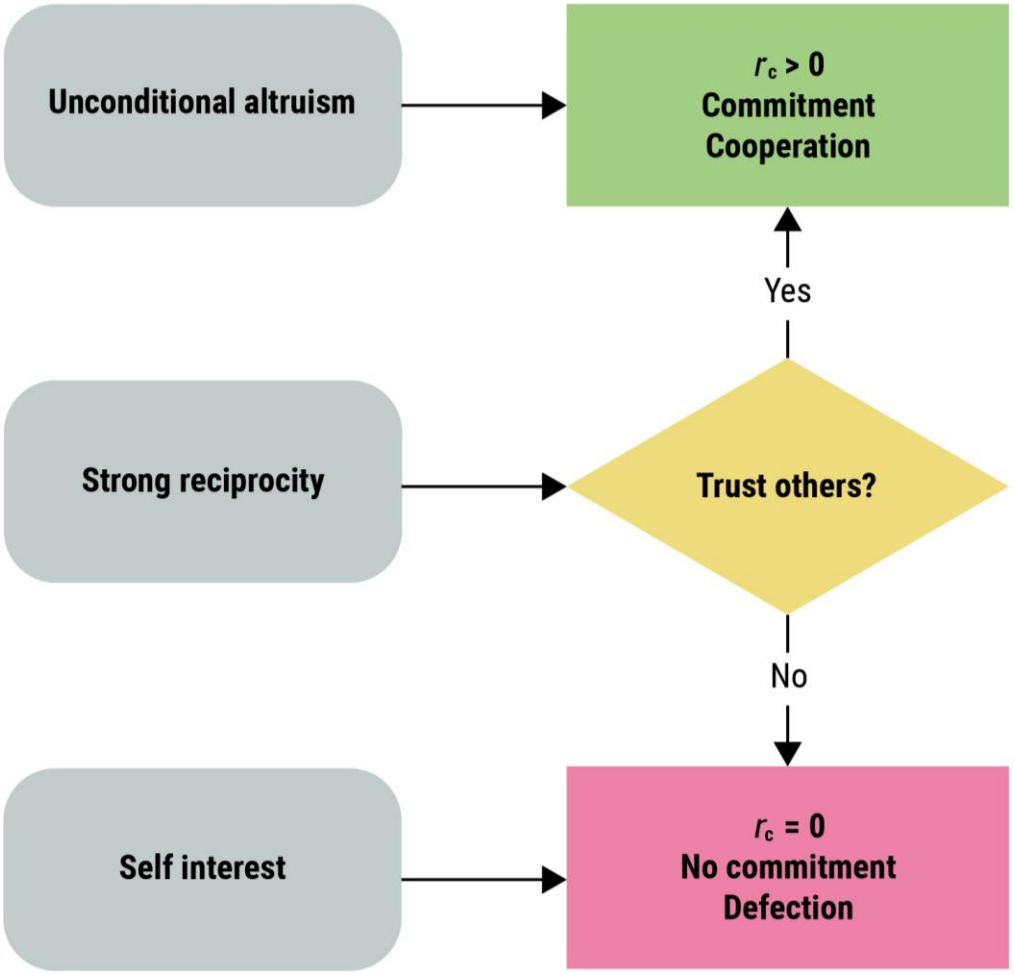



**Figure 15.1** Strong reciprocators may behave like self-interested actors or unconditional altruists, conditional on the level of trust. Note that in games with costly punishment, the strong reciprocator differs from both in that they may pay costs to punish defectors.

### 3.1 Political and economic outcomes: Evidence of the importance of trust

If most humans are strong reciprocators, but their decision to cooperate or defect depends on whether they trust others, this suggests that general levels of trust might predict some economic and political outcomes. There is a substantial literature on these connections, and I present a brief summary with key papers listed in Table 15.1.

First, trust varies substantially between nations. Bergh & Bjørnskov (2011) offer an overview of the spread in the international data: trust ranges from a low of 3 percent in Cape Verde to a high of 64 percent in Sweden, with a mean of 26 percent. National values are largely stable, though there have been notably significant declines in the USA and UK and increases in Denmark and Uruguay.

Second, most cross-national macroeconomic studies find that trust is the *strongest predictor of economic growth*—stronger than investment, human capital (e.g. Whiteley (2000), and so on. According to Bjørnskov (2017), the literature overall suggests that a one standard deviation increase in trust produces two thirds of a standard deviation increase in annual GDP growth, or around 0.6 to 1.0 percent. Further, "the literature clearly supports a causal effect of trust on growth" (p. 1). While seminal early studies, such as Knack & Keefer (1997) were largely associational, a second wave of papers, including Dincer & Uslaner (2010), Horváth (2013), and Bjørnskov & Méon (2015), establish causality. Horváth, for example, finds that trust in one period predicts economic growth in later periods.

Third, we can say a few things about the mechanism. As Kenneth Arrow (1974) observed, trust is

> *an important lubricant of the social system. It is extremely efficient; it saves a lot of trouble to have a fair degree of reliance on other people's word…* (p. 23)

This is obviously important where contracts are incomplete—and in the modern conception, almost all contracts are incomplete given the prohibitive costs of specifying and negotiating all possible contingencies (Hart, 1995). A great deal of economic activity depends on a handshake,[208] and on implicit agreements to fairly distribute unexpected windfalls or burdens.

In the cross-national literature, Bjørnskov (2012) notes that

> *trust is associated with a number of features that would appear on most economists' shortlist of important determinants of economic growth.* (p. 3)

Most involve public goods, which individuals will better supply where they act as trusting strong reciprocators: for example, among the most consistent are improvements in the quality of institutions and governance, reductions in corruption,

[208] This is commonly optimal in a world of norms, e.g. Kessler and Leider (2012).

**Table 15.1.** Empirical papers establishing a connection between social and between-group trust and growth and/or determinants of growth. Papers in bold are causal for growth.

| | **Empirical papers** |
|---|---|
| **Economic growth** | Knack & Keefer (1997), Whiteley (2000), Zak & Knack (2001), Beugelsdijk et al. (2004), Beugelsdijk & Van Schaik (2005), Ahlerup, Olsson & Yanagizawa (2009), **Dincer & Uslaner (2010),** Bjørnskov (2012), **Horváth (2013), Bjørnskov & Méon (2015)** |
| **Institutional quality, govt. performance, and corruption** | Rice and Sumberg (1997), Porta et al. (1997), Knack (2002), Uslaner (2004), Bjørnskov (2010), Bergh & Bjørnskov (2011), Graeff & Svendsen (2013), You (2018), Bollyky et al., (2022), Lessenski, (2022) |
| **Investment and human capital** | Coleman (1988), Porta et al. (1997), Putnam (2001), Papagapitos & Riley (2009) |
| **Private exchange** | Chami & Fullenkamp (2002), Alesina & La Ferrara (2005), Pace & Becchetti (2006), Guiso et al. (2009), Montalvo & Reynal-Querol (2021), Bechetti et al. (2022) |

and increased investment in human capital. Trust also strongly predicts success in managing specific threats, being associated with reduced COVID-19 infections (Bollyky et al., 2022 in a major Lancet paper) and increased resistance to "fake news" (Lessenski, 2022). Others are suggestive of Arrow's lubricated market transactions: increased investment and lower costs of capital (Bechetti et al. 2022), improved firm productivity (Chami & Fullenkamp, 2002, and Pace & Becchetti, 2006), increased bilateral trade (Guiso et al., 2009), and exchange across ethnic groups (Alesina & La Ferrara, 2005, and Montalvo & Reynal-Querol, 2005).

### 4.1 The first vulnerability: Parochiality

Human other-regarding preferences also have their dark side: they tend to be *parochial*. Empirically, individuals are commonly biased towards cooperation with ingroup members (e.g. co-religionists or, as in Carlin & Love, 2013, political co-partisans), and cooperativity may weaken or vanish when dealing with outgroup members or turn into spite for groups that are adversaries (Yamagishi, Jin, & Kiyonari, 1999, Yamagishi & Kiyonari, 2000, Makimura & Yamagishi, 2003). This is reflected in lower trust in outgroup members (Brewer, 2008; Carlin & Love, 2013).

The nub for our coming analysis, especially in Chapter 17, is that individuals tend to:

- apply higher weightings $r_c$ to the payoffs of culturally-related others;
- apply lower, and sometimes *negative*, values of $r_c$ to competing cultural groups; and
- therefore may invest in *narrow* public goods that benefit only the ingroup. Where $r_c$ is negative, individuals may willingly pay private costs to harm outgroup members; such investments too can be understood as a narrow public good for the ingroup.

This is suggestive of the dangers of Smith's "faction" and "fanaticism".

Parochiality, along with the other innate moral tendencies discussed so far, is the consequence of Smith's "fitting" of humankind to the social condition—and today this fitting process is understood as one of evolutionary selection. In the case of human moral evolution, the dominant theory is that of *cultural group selection* (a theory to which Herbert Simon, ever the pioneer, made some contributions—see Simon, 1990). There is insufficient space to explore this fascinating literature here, but it is discussed in detail in A.15.4. The core of this evolutionary story is that

- cultural group selection entailed competition between groups, where group boundaries were defined by *cultural affiliation*; and
- the cultural groups most likely to prevail and proliferate were *cooperative internally* and *competitive externally*.[209]

The simplest way to summarise its consequences is that it produces strong reciprocators whose trust in others, and so willingness to cooperate, depends on *perceived cultural distance*. And indeed, people today take cultural identities as a boundary for cooperation (Yamagishi, Jin, & Kiyonari, 1999, Yamagishi & Kiyonari, 2000, Richerson & Boyd, 2001, Makimura & Yamagishi, 2003, Balliet, Wu, & De Dreu, 2014).

The world's most successful religions are representative of the kinds of cultures that succeeded (Johnson & Kruger, 2004, Atran & Henrich, 2010, Norenzayan et al., 2016). In general the major religions implore followers to "love thy neighbour as

[209] There is strong evidence of group conflict in the anthropological record, and that victorious groups impose their norms on the conquered, e.g. Kelly (1985) and Soltis et al. (1995).

thyself", where "neighbour" has historically tended to refer to one's co-religionists. They frequently employ fictive kin relationships, where the priest may be a "father", and fellow believers "brothers" and "sisters". At the same time they tend to favour the exclusion and suppression of non-believers and apostates, and encourage self-sacrifice in wars against competing religious groups. As Ridley (1996) writes,

> *When Joshua killed twelve thousand heathen in a day and gave thanks to the Lord afterwards by carving the ten commandments in stone, including the phrase "Thou shalt not kill," he was not being hypocritical.* (p. 192)

Tajfel et al.'s (1971) "minimal group" experiments, while not necessarily informative about real-world behaviour, help to illustrate how salient group identity is for our species. Participants divided into groups over the most trivial of matters, such as their partiality to the paintings of Wassily Kandinsky versus those of Paul Klee, show significant biases towards their ingroup. Kramer & Brewer (1984) find that such minimal groups strongly shape cooperation and defection in Prisoner's Dilemma and Public Goods games.

Parochiality does not imply black-and-white behaviour—i.e. cooperate with the ingroup and compete with all outgroups—but rather, cooperation may vary continuously according to perceived cultural distance. For example, Handley & Mathew (2020) survey individual members of nine sub-clans in Kenya and find that tendencies to cooperate with anonymous members of other clans vary according to cultural distance rather than geographic distance (the lack of geographic effects, alongside clan populations totalling in the millions, rules out reputational motivations). For a selection of other examples, see Cox & Deck (2007), Charness and Gneezy (2008), Chen & Li (2009), and Ockenfels & Werner (2014).

Darwin (1871), a great admirer of Smith's moral theory, contemplated the importance of shifts in perceived cultural distance over the course of human history:

> *As man advances in civilisation, and small tribes are united into larger communities… [he] extends his social instincts and sympathies to all the members of the same nation… there is only an artificial barrier to prevent his sympathies extending to the men of all nations and races. If, indeed, such men are separated from him by great differences in*

> *appearance or habits, experience unfortunately shews us how long it is before we look at them as our fellow-creatures.* (p. 96)

Sympathy has continued to extend its reach since Darwin penned these words, although this is, in infelicitous circumstances, reversible.

### 4.2 The second vulnerability: The low-trust trap

The strong reciprocator's conditionality on trust introduces another vulnerability into human cooperation. The equilibrium level of cooperativity depends on the *composition* of a group. If the mix of altruists, reciprocators, and defectors is known, "a group's cooperative outcomes can be remarkably well predicted" (Kurzban & Houser, 2005, p. 1803). At fairly low critical levels of defection, strong reciprocators respond by becoming distrustful and themselves defecting. As Ostrom (2000b) writes,

> *Conditional cooperators will tend to trust others and be trustworthy in sequential prisoner's dilemma games as long as the proportion of others who return trust is relatively high. Conditional cooperators tend to vary, however, in their tolerance for free riding. Some are easily disappointed if others do not contribute, so they begin to reduce their own contributions. As they reduce their contributions, they discourage other conditional cooperators from further contributions. Without communication or institutional mechanisms to stop the downward cascade, eventually only the most determined conditional cooperators continue to make positive contributions in the final rounds.* (p. 142)

Untrustworthy behaviour among a self-interested minority can drive out trust in the majority, in what Lyons & Mehta (1997) refer to as a "Gresham's Law of trust" (p. 253).[210] This can be countered if strong reciprocators have access to cost-effective punishment, even when strong reciprocators are a minority (Fehr & Schmidt, 1999; Fehr & Fischbacher 2003; Perc et al. 2017). Again, following the general theory in Chapter 8, the principal role of rules is to defend commitment.

[210] Gresham's Law is the monetary principle that "bad money drives out good money".

The large-scale society faces unique challenges, given individuals cannot easily observe the behaviour, and so trustworthiness, of millions of unknown strangers. In such cases, it is not necessarily the presence of defectors that matters; the same pattern may emerge if conditional cooperators perceive that there are many defectors present. Even a pure population of strong reciprocators can become stuck in a low-trust trap, if enough are pessimistic about the trustworthiness of others (Cohen, 1999, Levi, 1998). In the large-scale society, expectations about strangers, and especially unfamiliar groups, tend to be culturally transmitted, and absorbed from political leaders and representations in news and entertainment media.

Because, as we will see, democracies depend on trusting strong reciprocators, these learned stereotypes of trustworthiness may have significant consequences for long-run outcomes. Today's social media algorithms are particularly disturbing in this light; their tendency to actively promote divisive, outrage-inducing content may erode the trust that underpins good-faith democratic debate.

*…[a] republic cannot succeed, till it contains a certain body of men imbued with the principles of justice and honour.*
Charles Darwin (1839, p. 165)

*Whenever the general disposition of the people is such, that each individual regards those only of his interests which are selfish, and does not dwell on, or concern himself for, his share of the general interest, in such a state of things, good government is impossible.*
John Stuart Mill (1861, p. 29)

*A good and sound constitution is one under which the law holds sway over the hearts of the citizens; for, short of the moment when the power of legislation shall have accomplished precisely that, the laws will continue to be evaded.*
Jean-Jacques Rousseau (1782/1985, p. 4)

# 16

# Building liberal democracy: Smithian sympathy and institutions as essential complements

Let us start with a puzzle. In Chapter 8, I argued that commitment mechanisms anchor the level of cooperativity in the long-run, because the rules that support social order are public goods. Here, other-regarding preferences ($r_c$) will be the main commitment mechanism of interest. Yet in Smith, and in these chapters, the value of $r_c$ is generally taken to be modest in most of the population, implying a *weak anchor*. If the anchor is weak, how is it that the level of cooperative efficiency in democracies can frequently exceed that of dictatorships? If we cannot find an adequate answer to this puzzle, then there is a large hole in the framework.

A clear answer is available in the design of democratic choice environments, which effectively function as *institutional fulcra* for leveraging modest other-regarding preferences into powerful effect. “Give me a place to stand”, said the ancient Greek mathematician Archimedes, “and with a lever, I will move the whole world” (see Lang, 2015, p. 162). These democratic fulcra have indeed moved the world. They involve reducing the influence of private payoffs $c$ in the utility function, so that

other-regarding commitments and public benefit ($r_c b$) dominate individual choice. The logic is captured by Goodin and Roberts (1975), in a brief note on voting behaviour I found after developing these ideas:

> *All that is required for present purposes is that ethical preferences carry* some *weight, that they will be decisive at least where everything else is equal.* (p. 927)

The fulcra function by making "everything else equal". They come in two main types:

1. the large-scale election, which directs other-regarding preferences into the rule-making process; and
2. the institutions that form "professions" and underpin the creation of judiciaries, militaries, and political media that are consistent with democracy.

A useful, if peripheral, observation is that market institutions have the opposite effect—rather than a fulcrum, they are a *damper*. They also make "everything else equal", but rather than eliminating private payoffs, they eliminate the possibility of affecting group payoffs. Motivational commitments therefore disappear (Box 16.1). This helps solve a puzzle: why self-interest is a reasonable assumption for economists studying *market behaviour*, even if it is profoundly inappropriate for the study of political behaviour. This helps us to understand why distinct aspects of human motivation are prominent in Adam Smith's *TMS* and *WoN*.

This chapter imposes one last temporary assumption, to be lifted in Chapter 17: individuals are *non-parochial*. Individuals vary in the strength of their other-regarding weightings, but all individuals apply the same weighting to others. Formally, within-individual weightings $r_c$ are homogeneous. The perfectly self-interested agent fits this pattern, of course, applying $r_c = 0$ universally. If the actor is prosocial, so that $r_c = 0.2$ for example, they will apply this value universally. Put another way, there is either only a single cultural identity, or the cultural differences that exist are irrelevant to cooperativity. This is an unrealistic assumption, although useful because within-agent homogeneous $r_c$ has strong affinities with liberalism: it implies equal valuation of individuals, rather than group hierarchies. It exemplifies Adam Smith's "impartial spectator".

Section 1 turns to the first fatal problem for democracies composed of self-interested actors, introduced in Chapter 14: the matter of voter turnout. A model of voters with motivational commitments ($r_c > 0$) resolves Downs' paradox. The larger the electorate, the lower the odds of affecting the outcome, and but this is exactly compensated by a larger public benefit: the voter has a smaller chance to produce a larger benefit. Downs' paradox is thus a feature and not a bug; as Goodin and Roberts (1975) suppose, large electorates mean that voter behaviour is driven by other-regarding preferences. Finally, strong reciprocation plays an important role in the model: the voter refuses to be a "sucker", and this prevents altruism from turning into martyrdom.

Section 2 examines the second problem in Chapter 14: the self-interested Olsonian majority bloc will extract from, and may disenfranchise, enslave, or do anything else to the minority so long as it is materially rewarding. Analytically the problem is that payoff-dependence ($r_s$) is *domain-limited*. Self-interested voters only care about efficiency. Payoff transformation by other-regarding preferences ($r_c$) is *domain-general*. Weak other-regarding preferences are sufficient to rule out highly welfare-costly forms of extraction, and support some redistribution—they support Smith's justice, and enough beneficence to address "shocking enormities". The higher-trust the society, the more it pursues both efficiency and redistribution.

Section 3 examines the second kind of fulcrum used to leverage commitment: the *profession*. Professions are used to supply other essential democratic public goods, the judiciary being the archetypal case. They require "extrinsic insulation", which reduces the private payoff $c$ associated with individual choices—again, making "everything else equal". Weighted group benefits dominate the professional's utility function, and they maximise group payoffs ($b$).

Via the judiciary, I argue that two conditions make professions essential: a need for *independence*, which would be violated by control via imposed rules; and problems of *complexity* that prevent codification in rules. As a consequence, professions will tend to be destroyed by incentive schemes and/or market competition. Common "economically rational" methods aimed at improving efficiency have perverse effects when applied to professions.

Section 4 closes by turning to the final public goods problem of democracy: that of the *informed* vote. On the demand-side, there is no fulcrum for leveraging the weak commitments of audiences, and innovation capacities between voters and special interests are asymmetric. Staying informed is too costly and complex. The only way to supply the public good of the informed vote is to manage the *supply-side* of information markets. However, political media cannot be tightly regulated, because of the two problems raised in Section 3: independence and complexity. I conclude that the most promising strategy is to *manage the fourth estate as a profession*.

**Box 16.1 Markets as a damper for other-regarding preferences**

Markets select for survival over *ex post* profitability, favouring firms and individuals that purely optimise over $c$. Wherever market rules do not align profitability and social welfare, market actors who pursue unprofitable forms of ethical behaviour will be less competitive.[211] They will tend to be eliminated over time. For example, firms that refuse to exploit the limits to consumer detection in certain healthcare or financial markets, or that refuse to engage in anti-competitive behaviour, will be ripe targets for acquisition. As noted in Chapter 12, Friedman's (1953, p. 158) "as if" theory of firm optimisation equally implies that the surviving firms we see around us may be the most effective parasites.

As a consequence, other-regarding actors cannot improve group outcomes systematically, or in the long-run, by acting beyond the rules that shape profitability in markets. Market institutions therefore suppress the influence of $r_c b$ on choices. This is fine so long as market rules are tuned such that private and social returns are aligned—and the role of commitment is to adapt rules, to run the race, so this remains the case. Commitment in rule-formation, alongside optimisation by profit-seeking, is an enormously productive division of labour so long as the latter function does not attack the former.

For a science of market behaviour, self-interest is a reasonable first-pass assumption. For a science of voter, judge, and military behaviour in healthy democracies—where fulcra are in place—other-regarding behaviour is a reasonable first-pass assumption. Again, this division was set out at the dawn of political economy in Smith.

[211] "Impact investing" and its variants only pursue ethics to the degree it is profitable, which limits them largely to pursuing strategies that would also be viable for non-ethical actors. See Schwartz and Finighan (2020).

So concludes my model of healthy democracy. These two fulcra—elections and professions—allow us to leverage motivational commitments into the supply of key democratic public goods: the informed vote in the public interest, a trustworthy judiciary, truthful media, and a loyal military.

### 1.1 Tying the loose ends of Downs' paradox: $r_c$ and the vote

Chapter 14 noted that voting is instrumental, strategic, increases rather than falling with education, and is more strongly driven by macro-societal performance than personal circumstances. I add that trust (here raising $r_c$) is positively associated with turnout (Dalton 1998, Grönlund & Setälä, 2007, Carreras & İrepoğlu 2013) and with political protest in both democracies and authoritarian societies (Benson & Rochon, 2004). Where distrust rises following adverse economic shocks, such as trade exposure and the Global Financial Crisis, one universal consequence is abstention from the vote (Guiso et al., 2017). We will turn to the more significant parochial reaction, which typically comes after the initial abstention, in the next chapter.

Although he formulated the paradox of voting, Downs (1957/1985) himself made allowance for such possibilities:

> *In reality, men are not always selfish, even in politics. They frequently do what appears to be individually irrational because they believe it is socially rational—i.e., it benefits others even though it harms them personally*. (p. 27)

From here I offer some simple formal models to supplement the discursive analysis. They cannot possibly capture the full complexity of human behaviour, but they show other-regarding preferences—and strong reciprocation in particular—serves as a powerful foundation for explaining turnout and policy choice.

### 1.2 Democratic origination, repair, and participation

Let us briefly revisit Downs' paradox in light of $r_c$. The below largely follows Edlin et al. (2007), but it is worth going through the logic. Take the election as a lottery, where to "win" is to cast the decisive vote (non-decisive votes having no effect on outcomes at all). The probability $p$ of "winning" is affected by the size of the voting population, and would be $1/n$. Following Edlin et al. (2007), add a parameter $K$ to

indicate the expected closeness of the election. The effect of closeness is to reduce the effective size of the voter base. If, for example, a voter knew that turnout would be exactly 50 percent, then $K$ would equal $n$, and therefore $p = 1$. Thus, knowledge about electoral closeness is important for determining the odds of pivotality:

$$p = K/n$$

Let us frame the voter's choice in a form similar to $rb > c$, a formula the reader may recall from Chapter 4, and that parsimoniously depicts the conditions for supply of public good: the weighted benefits of others must be greater than the private cost. We can break down the expected private gain $c$ into (A) the cost of voting ($c_v$) and (B) the expected private benefit of the vote, $pc$, which is the probability $p$ of casting the pivotal vote multiplied by the private payoff $c$ of policy change. We can alter $rb$ in two ways: First, $r = r_c$. Second, benefit to others $b$ turns into $pb$, i.e. the expected group benefit of voting given the odds of pivotality. Finally, put the benefits of voting on the left-hand side, and the cost on the right. This gives:

$$r_c pb + pc > c_v \qquad \textbf{Eq. 16.1}$$

If the left-hand of the equation is higher than the right, the individual votes. In national elections, $p$ is extremely small. Edlin et al. (2007) give the example of an election where candidates are expected to win between 47 and 53 percent of the vote as is typical of US national elections, giving a K of 8.33. In an election with one million voters, the chance of pivotality is one in 120,000. We can get the classical Downsian abstainer by making the voter self-interested such that $r_c = 0$. Only $pc$ remains on the left side of the equation, and the miniscule odds of pivotality reduce the gains of voting to effectively zero. In Chapter 14, I noted the example given by Edlin et al. (2007):

> *even if the outcome of the election is worth $10,000 to a particular voter, the expected utility gain is less than 10 cents.* (p. 297)

Things change, however, where $r_c > 0$. Increasing the size of the electoral population, $n$, no longer reduces the incentive to vote. As $n$ rises, certainly the odds of pivotality $p$ shrink, but $b$ increases by an exactly compensating amount given a larger population

benefits from the policy change. We can formalise this by rewriting group benefit $b$ as equal to the *average benefit to individual others* ($\overline{B}_o$), multiplied by $n$:

$$b = n\overline{B}_o$$

Now let us again multiply the above by $p$ to reflect that voters have a small chance of producing this benefit. Then we expand to reflect that $p = K/n$, and then we have $n$ as both numerator and denominator, which cancel out to give $K\overline{B}_o$.

$$pb = \frac{Kb}{n} = \frac{Kn\overline{B}_o}{n} = K\overline{B}_o$$

This is an intuitively obvious result: as $n$ grows, the voter understands that they have a smaller chance to positively affect a larger number of people. [212] They consider the closeness of the election in $K$, because increasing the odds of pivotality above $1/n$ increases the expected payoff. If an other-regarding individual knows they will be the decisive voter in an electorate of one million, the vote naturally has more value.

The above equation naturally needs to be weighted by $r_c$, to reflect that the payoff accrues to others. Putting this back into the criteria for voting in Equation 16.1 gives us:

$$r_c K\overline{B}_o + pc > c_v$$

Finally, we can eliminate $pc$, expected private benefits, as inconsequential in large elections on the usual Downsian logic.[213] Note supportive empirical evidence that suggests that voter motives become less sociotropic, and more selfish, in very small elections where $pc$ is relatively large (Fedderson, Gailmard & Sandroni, 2009). Thus, we get:

$$\boldsymbol{r_c K\overline{B}_o > c_v} \qquad \textbf{Eq. 16.2}$$

This gives us the empirically observed pattern: sociotropic voting, with sensitivity to the importance of the election and the tightness of the race.

[212] Prospect theory (Kahneman & Tversky, 1979) may further add to voters' motives. It tells us that voters overweight low probability gains and losses—i.e. should be willing to overpay to avoid the possibility of regret for not voting, in the case they may have been pivotal.
[213] This gives a natural result in the case that the voter knows they are pivotal (i.e. $K = n$): $r_c b > c_v$. The benefit of voting is the total weighted social benefit, without a probability modifier.

See A.16.1 for examination of a model with a lump-sum "D" term and voters driven by norms. Patterns of behaviour among voters, and in experimental games, indicate that the D-term and norm approach is insufficient,[214] but may be complementary to explanations that rely on weighting others' or group payoffs.

### 1.3 Completing the model: Unconditional altruists versus strong reciprocators

The above analysis largely follows Edlin et al. (2007). An important addition is the logic of the strong reciprocator. Equation 16.2 is suggestive of pure altruism, because policy platforms are evaluated purely in terms of their benefits to others. This leads to some apparent martyrdom problems. For example, an individual would be willing to support any platform that taxed them or their community punitively, so long as it benefited society on average (i.e. $\overline{B}_o$ was still positive). They may also pay implausibly high costs to vote (see A.16.2 for a brief discussion).

These interpretations are correct for the *unconditional* altruist, the strawman model of human other-regarding preferences. Strong reciprocators are altruistic—they will pay the costs of providing public goods even when their own contribution does not redound to increase their own gains—but conditionally so. Their willingness to contribute depends on trust that benefits and burdens are reasonably distributed among group members. This solves the problem. Unfair treatment is an affront. Unfairly high costs of voting give the voter an excuse for non-contribution.[215] The result is that effective $r_c$ declines, and the individual abstains, engages in punishment, or otherwise gains an appetite for reform. The relationship between the strength of the weighting $r_c$ and the distribution of costs and benefits is a complex empirical matter, and one worthy of future investigation.

### 2.1 Taming Olson's illiberal majority: $r_c$ and policy preferences

We have an answer to the Downsian problem. The second problem discussed in Chapter 14 was that self-interested voters may extract egregiously from, and disenfranchise, the minority.

[214] A lump-sum payoff for voting cannot explain responsiveness to electoral importance, closeness, and so on.
[215] We can expect strong reciprocators to prefer rules that ensure the cost of voting is small and fairly distributed, and indeed violations of this commonly provoke moral outrage.

The key is that payoff transformation via structural payoff dependency ($r_s > 0$), on which Olson relies, is *domain-limited*. Structural payoff dependence promotes efficiency in extraction, but no interest in social welfare. No value of $r_s$ can motivate pure redistribution or purely welfare-enhancing public goods, nor protect minorities from violations of basic rights or extreme extraction strategies, unless these compensatingly boost group efficiency. A high value of $r_s$ is consistent with the slaveholder who ensures their "property" is well maintained and disciplined, yet extracts all of the labour surplus. It is consistent with the Stalinist extraction strategy described by Olson (2000), which sequestered most of the social surplus. In short, $r_s \gg 0$ only promotes the efficient society, and the efficient society may be full of terrors.

Payoff transformation by other-regarding preferences ($r_c$), in contrast, is *domain-general*. It promotes a complete interest in welfare, whether welfare is increased by efficient policy or by rights allocations and pure transfers. Let us add $r_c > 0$ to Olson's equation for majority preferences, to see how the majority will treat the minority. It is useful to separate benefits to others (i.e. to the minority) $b$ into two components:

- $b_e$, the pure efficiency benefits of policy platforms; and
- $b_w$, which are like transfers with pure welfare effects and no efficiency impacts.

Olson's self-interested majority only cares about $b_e$. Its payoff function is:

$$B_m = r_s b_e - c \qquad \textbf{Eq. 16.3}$$

Now we add other-regarding preferences into the payoff function. Other-regarding actors care about both efficiency and transfers, hence $r_c$ applies to both.

$$\boldsymbol{B_m = r_{cs} b_e + r_c b_w - c} \qquad \textbf{Eq. 16.4}$$

If we expect values of $r_c$ to be generally low (again, say $r_c = 0.2$) then the effect of $r_s > 0.5$ will be relatively stronger, though no longer absolutely dominating. I discuss the implications in the next section.

Finally, it is useful to point to the trusting and distrusting cases:

- With distrusting strong reciprocators, $r_c = 0$ and the majority payoff function collapses into the selfish Olsonian optimum in Equation 16.3.
- With trusting strong reciprocators, $r_c > 0$ and we retain Equation 16.4. The majority supports some modest departures from the selfish optimum, increasing social efficiency and making some transfers possible.

### 2.3 Low values of $r_c$: Short of heaven, safe from hell

These modest departures from the selfish optimum have momentous effects. Take the case of slavery. So long as the welfare losses ($b_w$) are recognised as enormous, even low values of $r_c$ are enough for the majority to reject it outright. Most well-consolidated democracies tolerate, with some unease and contestation, lesser extractive workplace practices and moderately punitive treatment of the unemployed; this is a case where the prioritisation of perceived efficiency combines with welfare costs perceived as modest, and sometimes with forms of prejudice and parochiality. In some societies more than others, the unemployed and low-skill tend to be regarded as less trustworthy and more culturally distant (Box 16.2).

Liberal democracies motivated by a positive value of $r_c$ will make some efforts to move towards the efficiency frontier *and* to stem inequality and secure minority rights. However, to the degree that $r_c$ is modest and $r_s$ much larger, progress will be markedly stronger on efficiency than equity. Voter support for pure redistribution will be limited to groups perceived as high-need and trustworthy, may increase with national income as the average voter's marginal utility gains from consumption decline, and will vary with levels of generalised trust and cultural affiliation. Take the recently instituted National Disability Insurance Scheme in Australia, expected to have an annual cost of AU$44 billion by 2025—around 1.7 percent of national GDP. Such programs are possible because they target groups with a high marginal utility of consumption, and that are regarded as unlucky rather than untrustworthy (Box 16.2 again relevant). It is unsurprising that the largest redistributive programs tend to take the form of social investment, which is expected to boost efficiency: investments in education, healthcare, and other resources that assist needy individuals to become productive members of society. Finally, this is consistent with empirical evidence that links the size of the welfare state to the degree of social trust (Bergh & Bjørnskov, 2011).

**Box 16.2 Crowding out stereotypes: Information, trust, and generosity**

Aarøe & Petersen (2014) offer highly suggestive evidence on how differences in stereotypes, and so between-group trust, affect patterns of welfare provision in different countries. They examined the willingness of US and Danish subjects to support welfare provision to an unemployed individual in three conditions: first, a condition with zero information; second, a condition in which information indicated the unemployed individual is lazy (i.e. untrustworthy); and third, a condition in which information indicated the unemployed individual was hard-working but unlucky (i.e. trustworthy). In the condition with full information, Danish and US subjects behaved the same way: they were equally generous and punitive to the trustworthy and untrustworthy respectively. In the condition without information, Danish subjects were much more likely than US subjects to support welfare provision.

Aarøe & Petersen's interpretation is that US and Danish subjects do not fundamentally differ in their willingness to support cooperative or trustworthy individuals—indeed the willingness to support individuals who are honest but unlucky, and punish those who are lazy, may be a deep-seated and universal heuristic deriving from evolutionary conditions (Petersen, 2012).[216] Rather, when subjects lack particularistic information, they must rely on their general, culturally-inherited expectations about their compatriots—in Denmark the default assumption is that the other is a cooperative type (high rc), and in the US the default assumption is that they are a defector type (low rc).

The limits to other-regarding preferences means that liberal democracies will always remain some distance from Rawls' egalitarian optimum. This distance can be measured in social trust.

### 3.1 The second fulcrum: The "profession"

The electoral fulcrum, combined with the "liberal" pattern of other-regarding preferences, is enough to produce a voter majority that prefers an efficient, and modestly redistributive, society that protects basic rights. Whether or not voters are informed enough to achieve this outcome is a matter for Section 4. But voters cannot possibly undertake the task of domesticating the regulator on their own—they

[216] In keeping with the between-group competition that produced strong reciprocity, groups that support hard-working but unlucky members are more competitive, while groups that support lazy cheaters are less competitive.

neither know enough, nor can coordinate effectively enough, to ensure that laws are appropriately interpreted and executed. Democracy too entails a division of detection and enforcement labour, with certain tasks given to voters, and others to the judiciary, military, journalists, and so on.

These other actors need to supply essential democratic public goods, but here the electoral fulcrum has no power. The answer comes in a second kind of fulcrum that I will term the *profession*. As Cruess et al. (2004) write, professions are

> *governed by codes of ethics and profess commitment to competence, integrity and morality, altruism and the promotion of the public good within their expert domain. These commitments form the basis of a social contract between a profession and society… [and] the right to considerable autonomy in practice and the privilege of self-regulation.* (p. 75)

Commitment will again be at the heart of things—the force leveraged by the fulcrum. The following sections offer a useful definition of the profession. It is not merely a "discipline" or a licensed occupation, and nor is it defined by membership of professional organisations. Instead, it is defined by the particular configuration of motives that organise behaviour.

### 3.2 How to create a profession

We can create professions using two mechanisms for leveraging commitment.

The primary mechanism is "extrinsic insulation". It does what it says on the tin: it isolates professionals from most kinds of private payoffs that might arise from the choices they make. The professional is insulated from market incentives, from political co-optation and repression, and from other $c$-modifying influences that are external to the profession or fiduciary community (within-community rules remain significant, and will be discussed in a moment). The logic is akin to the removal of conflicts of interest, but what is special here is that the removal of incentives *reveals underlying commitments* ($r_c > 0$). Again, this is a generalisation of the logic of Goodin and Roberts (1975) comments on voter behaviour: where private rewards are removed, choices are made according to ethical considerations. In the extreme case

that $c = 0$, then $B_i = r_c b$. Even the smallest positive value of $r_c$ creates the optimality line such that individuals maximise $b$.[217]

Of course, it is not possible to reduce private incentives to zero in practice. In part, this is because making choices in the public interest always involves effort (a matter I return to shortly). Further, as Cruess et al. suggest, professions tend to use their own $c$-modifying rules to self-regulate. This is the logic of Chapter 8 reappearing: the community of committed actors originates and repairs its own rule-based structures to police the defectors in its ranks.

A second mechanism—significant, but mentioned briefly given space constraints—is "intrinsic augmentation". This is the shaping, enhancing, and maintaining of $r_c$ by education and socialisation, rituals and oaths, and emphasis on the significance of the social benefits $b$ provided by that order. Clearly this is complementary to extrinsic insulation; it further increases the relative strength of $r_c b$ over $c$ in the utility function. It is not by accident that the judiciary and military in particular are the most intrinsically augmented of all democratic institutions, dripping with reminders of the gravity of their duties, of their shared purpose, and their hallowed histories.

### 3.3 An archetypal profession: The judiciary

Now we turn to a paradigmatic case that illustrates the above principles, and is significant for understanding democracy: the judiciary.

The judiciary presents a puzzle for the self-interested model. The public goods it supplies are essential for democracy and justice in general, are highly complex, and must be supplied at high quality. They demand great effort. And yet there is no schedule of Pigouvian subsidies or penalties to match the marginal social benefits of each judicial decision, and little policing of effort. Rather, we are accustomed to trusting judges, to believing that they are motivated to do their job to the best of their ability. Indeed Stout (2001, title) argues that judges, if they are indeed fulfilling their functions, must be understood as "altruistic hierarchs". We must somehow maintain them as altruistic hierarchs, even though most judges probably lie somewhere

[217] Various ethical dilemmas, such as "trolley problems" (e.g. Kamm, 1989), have this character: the chooser has no private returns at stake, and decisions relate to the welfare of, and moral obligations to, others. Such problems are highly stimulating and emotionally evocative for ordinary individuals.

around the middle of the sympathy bell curve. We do so by via the professional fulcrum—via both extrinsic insulation and intrinsic augmentation.

In healthy democracies, judges are substantially isolated from material incentives and political interference. They enjoy an unusual degree of job security, and cannot receive commissions, kickbacks, or do a bit of for-profit lawyering on the side.[218] Their incomes can be taken as fixed, and the marginal material payoffs associated with decisions in cases is usually around zero. Thus, if we take judicial payoffs as defined by $B_i = r_c b - c$, their choices are usually dominated by weighted public benefit $r_c b$. If $c$ is sufficiently low, even weakly prosocial judges' strategies will lie on the optimality line and they will attempt to decide in the public interest.

Not all undesirable private incentives can be removed. The judge may prefer to avoid decisions that are just, if they provoke the discomforts of public outrage. Like all mere mortals, they may prefer to avoid hard work. The defining characteristics of the more self-interested judges ought to be (A) laziness and (B) timidity. Indeed, these are among the most common and long-standing criticisms of judges.[219] There are procedures called the "lazy judge rules" in some US states,[220] and in Australia, where judicial productivity has been a subject of some recent investigation, there is an 18-fold difference in the speed of judgement delivery between the fastest and slowest Federal Court judges (Patrick, 2018). This is imperfect and perhaps improvable, but attempts to improve it must be full of caution.

There are also desirable private incentives in place, which are overwhelmingly forms of self-regulation—in the Chapter 8 pattern, committed judges police their own community. The rules that govern court proceedings, decisions relating to promotion and decisions to censure or remove judges who violate community norms, and indeed the aforementioned "lazy judge rules", are overwhelmingly designed and enforced by members of the judicial community.[221] So long as rules are designed

[218] In the US they can own shares of publicly-traded firms, although cannot decide cases related to those firms. Chief Justice John Roberts recently argued for the stronger enforcement of these rules after some violations (Savage, 2021).
[219] E.g. in the US, noted by Judge Meier Steinbrink (TIME, 1924) and Judge William Henry (Dexheimer & Bureau, 2018), and a charge laid at Lawrence VanDyke, Trump's Federal appellate nominee by the American Bar Association (Alder, 2019).
[220] E.g. Trial Rules 53.1 and 53.2 in Indiana courts. See Johnson (2023).
[221] The first resort is to provide a sting to reputation and pride: judges get sent back to ethics school. In the case of judges violating conflict-of-interest rules, for example, US Chief Justice Roberts promised "more class time, webinars, and consultations." (Zirin, 2020)

from behind the barrier of extrinsic insulation, the judicial community can be expected to generally choose and impose rules that maximise $r_c b$.

Briefly, I observe that intrinsic augmentation attains its highest form of secular expression in the courtroom. The values and authority of the judiciary is expressed in the language, the dress, the architecture, the formality of proceedings and its closely guarded silences; all come together to convey that this is a space apart from the bluster and chicanery of the marketplace. The contemporary courtroom is a descendent of the temple, of ancient strategies that humans discovered for creating a sense that we are in the presence of a universal arbiter. Rather than speaking for the gods, liberal judges are supposed to speak for something akin to Smith's "impartial spectator". As Smith observed, the healthy liberal society must revere justice (1759/1976, p. 185), and contempt for the court is a shocking, secular blasphemy.

### 3.4 Why professionalisation? Managing complexity and preserving independence

Why is the judiciary run in this way, rather than—as a naïve reading of economic theory might suggest—by a set of Pigouvian incentives designed by the economics department, including perhaps some form of incentivised competition to be the most efficient justice? There are two reasons this approach would destroy the public good it hopes to efficiently supply. They are:

1. *Independence*: the provision of the public good requires independence from the vertical accountability structure.
2. *Complexity*: the provision of the public good requires professional judgement, because it is too complex to codify its provision in any set of rules.

In the first case, the courts act as a horizontal accountability structure that constrains the excesses of the legislative and executive branches. To allow the legislature or executive to devise an apparatus for controlling judicial incentives would destroy this function, and place it as a lower-order power within their vertical accountability structures.

In the second, even if the external rule-maker was perfectly benign and would never abuse its power, justice cannot be codified. Written law itself is complex, and human affairs are even more so. Judicial discretion, indeed *judgement*, is required to map

the one on the other. This is because of what legal theorists term the "open texture" of language (e.g. Bix, 1991), which is inherently ambiguous and in need of interpretation, and because social norms and technological strategies are ever changing. 14th century English common law is an example of failed attempts to codify justice; it became gravely maladapted to the rise of complex commercial activities (Hamowy, 2003).[222] One might also point to the manifest injustices brought by mandatory minimum sentencing laws in recent decades (e.g. Huang, 2010).

Next, the perverse effects of introducing incentives and competition in the courts will help us consider risks and perversities that arise from the push to "marketise" various other public goods.

### 3.5 The use and abuse of professions: Broader applications

The reality of other-regarding preferences, and the availability of institutional fulcra that harness them, nuances the conclusions of Part 3: it adds an additional structural regulatory tool.

Professionalisation is of most interest in cases that (A) involve the provision of public goods, such that competitive markets alone are inefficient, and (B) where regulatory control of supply would violate independence and/or confront complexity and so codifiability problems. Independence is the more significant and determinative requirement, where may be no sound alternative to professionalisation. It applies only in special cases where the public good involves horizontal accountability—mainly the judiciary, and in the case of political journalism (with which we will close this chapter). Complexity, and the challenge of codifying public goods, is less determinative but more widespread. Healthcare, education, and legal services are cases where client and community needs are often *sui generis*, and professional judgement cannot be routinised. The discussion of such cases is important, but somewhat ancillary to the core argument of Part 4, so these comments will be brief and sketch out the contours of future work.

In short, there is a clash between two methods of providing public goods:

[222] The common law system relied on the writ system, which rigidly specified the cases that could be brought to the court. Many commercial cases of the day, reflecting novel situations, did not fit within the prescribed writs and had no available legal remedy.

1. imposing a structure of rules that rewards and disciplines supply of the public good. Examples include bureaucratic hierarchies, or artificial markets created via Pigouvian subsidies for firms that compete to provide public goods (the idea being to create a Darwin machine to drive innovation and efficiency); and
2. using extrinsic insulation and intrinsic augmentation to leverage commitment and create a profession. Here, professionals are granted substantial leeway to choose strategies they view as socially optimal.

The standard economic assumption of self-interest makes the choice between the two simple: there are no trustworthy, committed actors. The second approach is impossible, and professions mythical. Yet when we recognise that the assumption of self-interest must be lifted if we are to understand our societies, professions become possible and, in some circumstances, may be superior.

I want to briefly point to the costs of turning to rules-based—and incentive and competition-based—provision. Most profoundly, we run into the problems of Box 16.1, which is that we create a system in which actor survival rests on satisfying the exigencies of rules. There is no problem if rules are perfect: the consummate professional will be the best competitor. But gaps between rules and the social reality are inevitable wherever public goods are complex; rules cannot act as a substitute for professional judgement, just as a rigid set of legal stipulations cannot act as a substitute for a flexible judiciary. Further, gaps are *actively* pried open by innovations in defection. Rules are adversarial and create Red Queen's races; quasi-markets, for example, create races in the first, second, and third legs of the Red Queen's race (Box 16.3).

Gaps between rules and reality systematically disadvantage any committed actor who insists on following their judgement and providing public goods that are unrewarded, and refuse to innovate in defection. They systematically advantage uncommitted actors, who have no qualms about following, and exploiting, incomplete rules. At worst, selective pressure will gradually cultivate a community of consummate box-tickers and cheaters. At best, the competitive, rule-based supply of complex public goods is likely to fall short of the hopes of efficiency-seeking reformers.

Finally, caution is warranted given the difficulty of restoring professions once degraded. In communities dominated by uncommitted actors, self-regulatory processes break down. Group members adept at optimising within incentive arrangements may be vested in their retention or extension, and disinterested in the perfection of the public good. The economists' belief that suppliers of public goods are uncommitted and untrustworthy may, in short, become self-fulfilling.

## 4 The informed voter: Running the Red Queen's race over political information

To return to the main thread of argument in this chapter, let us look back to the model of political markets in Chapter 14. There, voters ran a Red Queen's race against special interests for control of the regulator. We are in a better situation with other-regarding voters, and can explain the possibility, and common empirical reality, of a trustworthy judiciary and military. At least so long as voters have liberal

**Box 16.3 Privatising public services: Prepare for the relay**

Introducing quasi-markets for public services—using private competition to deliver unemployment, education, or health services—will tend to create:

- the first leg of the relay, by opening up opportunities to outsmart and mislead clients;
- the second leg, by providing subsidies for particular services, with incentives to game the rules by over-provision of services for needs that do not exist, cherry-picking the easiest clients in particular "buckets" of need, and so on; and
- the third leg, by creating a permanent new lobby group, with intimate regulatory relationships and deep knowledge, that perpetually innovates in manipulating regulatory incentives.

Privatisation is motivated by on-paper analyses in which the total regulatory structure, both market actors and institutions, operate optimally. Competition occurs strictly over mutualistic innovations. In practice it creates permanent Red Queen's races that will be costly to run and may often be lost. Losses in the third leg, coupled with weakening of the professional motive, may result in an inefficient captured equilibrium with potent vested interests (e.g. U.S. healthcare), from which it is difficult to escape. These costs must be balanced against anticipated benefits.

Preferences, we appear to have a working system: voters will run the third leg of the Red Queen's relay for control of the regulator, and induce the regulator to run the first two legs of the relay in the public interest.

But this is not a complete system yet. The happy result described in Section 2 only eventuates *if voters are well-informed*. Anthony Downs (1957), a worldly thinker, lost little sleep over whether voters would turn out; rather, his worry was whether voters would invest in acquiring and understanding political information. If the problem of voter "direction" is resolved, this does not solve the problem of voter "pace" (i.e. their capacity to run the information race and realise their preferences).

The argument in this section is simple:

- There is no fulcrum motivating voters to become informed. No plausible accounts of voter motivation, nor voter rationality, can enable voters to reliably run the information race against highly motivated and sophisticated special interests.
- Binding constraints on the *demand-side* (i.e. audiences) of information markets mean that democracy depends on management of the *supply-side*. That is, *cheap* information must generally be *accurate* information.
- Incentivising accurate information supply via state regulation is infeasible, due to both problems outlined in Section 3.4: the need for independence, given the media's role as a horizontal accountability structure in democracy, and the difficulty of codifying the detection of quality information.
- The analysis suggests the solution is to be found in a professionalised political media. This was roughly the state of affairs during the post-war ascendency of democracy, when media was highly trusted. The media has since been de-professionalised, and this is probably incompatible with the long-run survival of democracy.

There is enough to work out here for another thesis, so I will paint only the broadest lines and leave the details to future work.

### 4.1 The demand side: Weak commitments, no fulcrum

Unlike self-interested individuals, the other-regarding individual will at least make some investments in running the information race. This investment will, however,

tend to be grossly insufficient. There is no institutional fulcrum that can remove the private costs of acquiring information, and so no means of leveraging modest other-regarding preferences into large effect.

A full model of voter investments in information would be a large and complex project. Here I aim to simply constrain the problem by estimating the maximum amount voters could rationally invest in information given plausible degrees of commitment. Following Section 1.2, the expected value of the vote is $r_c K\overline{B}_o - c_v$. Voters cannot justify investing more than this in becoming informed, or they are better off not participating at all. Take $c_i$ as the costs of the information required to determine that $r_c K\overline{B}_o - c_v$ is positive for one candidate. Overall we get:

$$\boldsymbol{r_c K\overline{B}_o - c_v > c_i}$$

In ordinary elections, we expect this to be a small figure, no more than enough to justify several hours of effortful reading.[223] In practice, we can expect that they will spend less than this.[224]

There are many other bells and whistles to add on such a model, and probably the most important is that political information may be entertaining—or, at least, stimulating and emotionally arousing. Engaging information is "cheap" to acquire. This points us in a fruitful direction: *what is the relationship between engagement and accuracy in political information?* The literature on this point is not encouraging; the most engaging information is sensationalist and emotionally inflammatory (e.g. Bakir & McStay, 2017), and reinforces existing viewpoints (Moravec et al., 2018). Technological change is a significant part of the story. In previous decades, the cheapest information was that provided through the gatekeepers of a largely professionalised conventional media (to be discussed shortly). The advent of social media allows the distribution of inflammatory information without any professionalised filter.

[223] E.g. say typically one candidate is worth \$200 more than the other, and the voter must identify which candidate is which. Say $r_c = 0.2$, a typical $K = 8$, $c_v = 50$, and $\overline{B}_o = 200$. The left-hand side resolves to \$270.

[224] They stop investing when they are sufficiently certain that one candidate is better than the other. If they have strong priors, they may rationally determine that no investment is likely to be worthwhile.

### 4.2 Weaknesses in voter innovation capacity

Alongside the motive to invest, the other consideration is, of course, what I have termed relative innovation capacities. The argument will now be familiar: even perfectly motivated (i.e. $r_c = 1$) audiences still face the Chapter 11 problem of keeping up in a Red Queen's race, where the race is complex and their opponents are highly sophisticated. The same two problems appear in contemporary political information markets. The "economic law" proposed in Part 3 is that in complex markets, marginal regulation by consumers will fail. Where those markets are welfare-critical or relate to major public goods—as they do here—the consequences will be severe.

The complexity of political information has grown in recent decades. To begin with, policy problems are more complex—e.g. COVID-19, climate change, global systemic financial risk, geopolitical multipolarity and trade interdependence, artificial intelligence technologies, and so on). Each demands complex policy responses that entail trade-offs. None of this is beyond the capacity of a democratic polity, in the case that the technical aspects of the public discussion are guided by public trust in committed actors with expert knowledge. It is wholly beyond the capacity of a democratic polity where it is profitable to disseminate large amounts of disinformation.

The more fatal kind of complexity, then, is the sheer volume of conflicting information, and the increasing share of compellingly-crafted misinformation (including as part of foreign adversaries' "hybrid warfare" strategies). Just as firms invested in increasing the complexity of products to exploit consumers in Chapter 11, so special interests invest in increasing the complexity of the information environment. The tactic advocated by Donald Trump's chief strategist, Steve Bannon, was "to flood the zone with shit": to cheaply produce large volumes of inflammatory misinformation (Stelter, 2021, November 16, para. 1). Another strategy is to produce disinformation that mimics trusted forms, masquerading as scientific analysis or independent, third-party analysis. The patterns of disinformation surrounding climate change, for example, include both types: a morass of contradictory conspiracy theories is accompanied by pseudoscientific claims that distort, obfuscate,

or misinterpret legitimate scientific research. Large rents will forever motivate innovations in sophisticated misinformation strategies.[225]

For most viewers, the task of detecting fact from fiction becomes Herculean. The problem looks set to get worse: emerging media strategies combine big data with powerful AI models, including models that can generate compelling audio and video "deep fakes". Bakir and McStay (2017) argue that "the commercial and political phenomenon of *empathically-optimised automated fake news* is on the near-horizon." (p. 15, italics in original) That is, AI-operated systems of *personally-targeted* fake news and media. Eric Schmidt, former CEO of Google and founder of Schmidt Futures, a philanthropic partly focused on AI, was asked whether this new technology will destroy democracy: "It will", he responded (Isaacson, 2023, 4:15).

### 4.3 Limits to regulation of political information markets

To manage Red Queen's races characterised by (A) asymmetries in innovation capacity and (B) complexity, the discussions in Chapters 11 and 12 pointed to two kinds of regulatory response:

1. a simplifying strategy, structuring the choices available to consumers so they can plausibly accomplish the detection tasks they are set; or
2. an incentive-aligning strategy, regulating the payoffs of producers so they can only benefit by providing quality products.

In the case of political information, these strategies may play only a limited role. They are fraught with risks and contradictions. As in the case of the judiciary, such rules runs afoul of the codification and independence problems discussed in Section 3.4. We cannot devise a schedule of Pigouvian subsidies or penalties to guide socially productive political media output, and doing so would risk media independence, effectively turning it into part of the vertical institutional structure of the state.

Both institutional regulation and "deregulation" of political information are threats to the long-run viability of democracy.

[225] One example is the AU$22 million spent by mining companies to discredit the Australian Mineral Resource Rent Tax; the policy was scrapped, giving the mining industry a two-year return-on-investment in excess of 20,000 percent (Lewis, 2011).

## 4.4 Political journalism as a profession

Political journalism is a natural profession. The framework developed here suggests that democracy depends on its management as such.

Professional motives loomed large in news media during the twentieth century ascendancy of democracy. There was nothing special about the demand-side in that period—i.e. no obvious reason to suppose that the capability of audiences has since declined, and probably the reverse. Rather, it had a more favourable supply-side.

In the U.S., higher quality information supply was not the product of any particularly robust set of institutions, but of a few favourable trends. Limited access to the broadcasting spectrum, for example, reduced competitive pressures. High overheads meant media firms needed to attract large audiences to achieve profitability, and these economies of scale meant that it was unprofitable to produce extremist content. Fred Hirsch's (1976) pre-capitalist values loomed large; even in the private sector, socially irresponsible behaviour was scandalous. The echoes of the Second World War, and backdrop of the Cold War, increased national solidarity. The "fairness doctrine", enforced by the U.S. Federal Communications Commission, at least set out the expectation that media would give unbiased attention to matters of public interest (the doctrine was abolished in 1987).[226] Media firms also had a great deal of innovating left to do in gluing audiences to the screen, and selling those audiences to various interests. Rupert Murdoch's empire led the charge on both fronts,[227] and reflects the dark side of Friedman's "as if" dictum: survival on *ex post* profits will also select for the most effective parasitisms.

Trust in mass media was remarkably high in the post-war decades, before declining with each passing decade from the 1980s to the present day, with a halving of poll respondents reporting high trust, and as much as a tenfold increase in no trust (Figure 16.1). These patterns are rational. Where commitment drives content, the media is generally trustworthy. Where it is reorganised around the maximisation of *c*

[226] Given FCC commissioners were appointed by the president and confirmed by the Senate, this arrangement may not have fared well in today's hyper-polarised political environment.
[227] E.g. he pioneered entertaining, sensationalist content with little investigative value in tabloid papers, and the spreading of conspiracy theories, as in Glenn Beck's primetime show in the early 2000s. He innovated in strategies for cornering media markets and capturing policymaking via (either implicit or explicit) quid pro quo relationships with political candidates—with favourable coverage delivered for Prime Ministers Keating, Turnbull, and others as they adopted media deregulation and other Murdoch-preferred policies (e.g. see Schultz, 2021).

by any means, trust is irrational. Following the logic of Box 16.1 at the start of this chapter, the replacement of professional political journalism with "information markets" moves us from a fulcrum leveraging $r_c$, to a damper suppressing $r_c$.

To summarise, for a while we were lucky: the technical, economic, and cultural forces were favourable. We have since been unlucky. The analysis here suggests that we need to rely less on luck, by explicitly and robustly creating the institutional conditions for professionalised political journalism. Extrinsic insulation, as applied to political media, runs on the premise that journalists only serve special interests because of the extrinsic rewards.[228] Without any possibility of payment, with "everything else equal", the journalists will generally prefer reporting information that makes some contribution to their communities.[229]

I cannot, however, possibly solve this problem in these pages. Compared to the judiciary, the case has additional complexities. Setting the matter of social media aside, one may suppose that we could establish ring-fenced public funding for

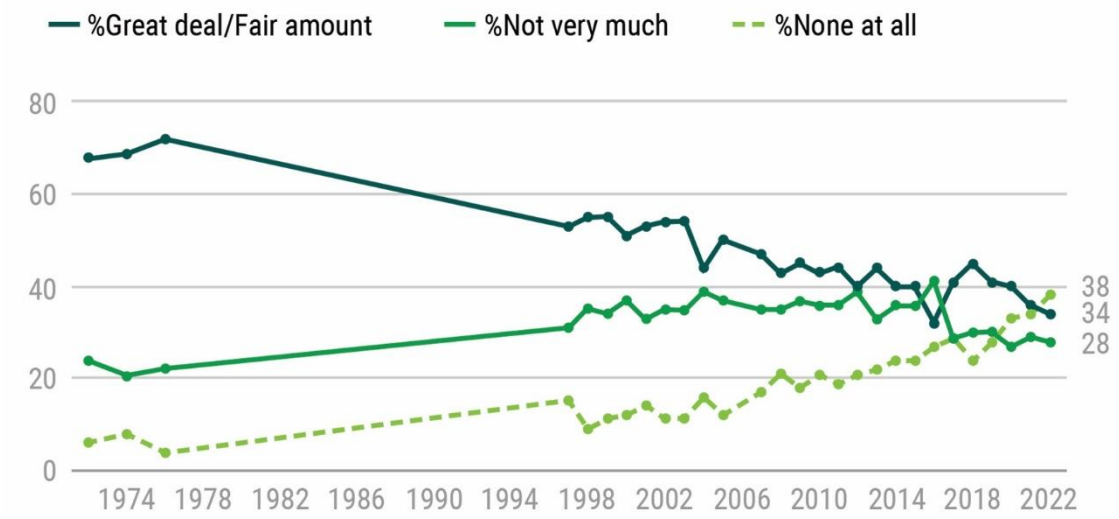



**Figure 16.1** Americans' trust in mass media, 1972-2022 (Gallup, 2022). "In general, how much trust and confidence do you have in the mass media—such as newspapers, TV and radio—when it comes to reporting the news fully, accurately, and fairly: a great deal, a fair amount, not very much or none at all?"

[228] Tucker Carlson, among the most ardent of on-air Trump supporters, confessed in private messages, "I hate him passionately" (Robertson, 2023). Large extrinsic rewards do a lot of work.
[229] "I believe there exists", wrote Charles Darwin, "an instinct for truth… of something of the same nature as the instinct of virtue" (Darwin, 1862, letter 68).

political journalists, in the same manner as that for the judiciary, so eliminating any connection with market income or political favour. Ideally this would be constitutionally protected; democracy depends on accurate political information as much as it depends on the judiciary. However, unlike the case of the courts, free entry and competition for audiences is intrinsic to journalism. We need journalists to compete over quality and viewership to a degree, but contradictions persist where the largest audiences may be won via inflammatory fabrications. Perhaps restricting or eliminating private material rewards to winning especially large audiences would help. We might explore mechanisms of self-regulation, specialised judicial oversight, and perhaps sortition.[230]

Further research and practical experimentation is a matter of urgency. It is at least obvious that, in the long-run, it is inconsistent with the survival of democracy to allow political information to be produced as a private good for the highest bidder. Any democracy built in this way creates complex Red Queen's races over information, and so amounts to a gamble that the cooperator-defector bias in change will remain forever favourable—and it assuredly will not.

## 5. Concluding

The survival of democracy requires understanding the causes of its success, and so of its vulnerability to failure. To do so, we must move away from the purely self-interested model. Liberal democracy arises from liberal other-regarding preferences in combination with two fulcra: large-scale elections and professions. With that, we get voters who prefer efficiency, some redistribution, and the protection of basic rights. We get a trustworthy judiciary, military, and indeed—if we choose to restore it—political media.

Other-regarding preferences are a well-evidenced reality; our task is to maintain a set of carefully designed institutional fulcra to put them to work.

[230] With promising evidence on the capacity of sortitional bodies to arrive at socially rational proposals, but much left to learn about how the process is optimally managed, insulated from undue influence, and to what kinds of questions it is best suited.

*An adversary is someone you want to defeat. An enemy is someone you have to destroy.*

Michael Ignatieff (2013, para. 2)

*The history of Rwanda shows us clearly that a Tutsi stays always exactly the same, that he has never changed. The malice, the evil are just as we knew them in the history of our country. We are not wrong in saying that a cockroach gives birth to another cockroach.*

*Kangura* (1993)[231]

*The real, revered, and impartial spectator, therefore, is, upon no occasion, at a greater distance than amidst the violence and rage of contending parties. To them, it may be said, that such a spectator scarce exists any where in the universe. Even to the great Judge of the universe, they impute all their own prejudices, and often view that Divine Being as animated by all their own vindictive and implacable passions. Of all the corrupters of moral sentiments, therefore, faction and fanaticism have always been by far the greatest.*

Adam Smith (1759/1976, p. 175)

# 17

# Dismantling liberal democracy: Parochialism and the double dividend

Now we arrive at the last part of the story: the dark side of motivational commitments. It is found throughout scholarly works on other-regarding preferences, from Adam Smith's warnings about "faction" and "fanaticism", to the literatures on empathy, altruism, social capital, and moral norms. Fowler & Kam's (2007) game theoretic work, for example, distinguishes:

- generalised "altruism", i.e. concern for the well-being of others "regardless of the identity of the beneficiaries" (p. 813); and

[231] From the newspaper *Kangura*, March 1993, a year prior to the Rwandan genocide. See Chrétien (1995).

- group-specific “social identification”, i.e. a concern for the success and competitiveness of the group with which one identifies, “possibly at the expense of other groups” (p. 813).

The social capital literature distinguishes between “bonding” social capital that favours ingroup members, and “bridging” social capital that extends to outgroups (Gittell & Vidal, 1998, p. 15; also see Gargiulo and Benassi, 1999, Putnam, 2000). The presence of the former without the latter can lead individuals to invest in between-group conflicts. Capacities for empathy, meanwhile, can be limited to ingroup members (e.g. Eres & Molenberghs, 2013) and even motivate the harming of outgroup members (counter-empathy or schadenfreude, e.g. Cikara et al., 2014 and Hudson et al. 2019), and underlies capacities for deception (Bubandt & Willerslev, 2015). Other scholars note that moral norms, trust, and social capital may be rich within organised crime groups such as the mafia, criminal gangs, and organisations such as the Ku Klux Klan (Fukuyama, 2000).

This chapter aims to

- place the theoretical and empirical literature on between-group conflict within the framework of commitment, rules, and the Red Queen’s innovation race;
- examine the consequences of parochiality for the two commitment-harnessing fulcra of Chapter 16; and
- draw out the connection between group distrust and the turn from democracy to authoritarianism. Its structure mirrors that of Chapter 16, but with parochial rather than liberal preferences.

Section 1 discusses voter turnout in the special case of parochial preferences. Not only do parochial commitments drive collective action, but empirical evidence suggests that partisanship intensifies the willingness to vote, protest, and so on.

Section 2 turns to the preferences of the parochial majority. Spite is not necessary to generate harmful outcomes. If the parochialist values the ingroup but is indifferent to the outgroup, this is enough to favour extreme forms of extraction (e.g. slavery) and moral-cultural domination. Political victory grants the winning ingroup the *double-dividend*: the benefits of the ingroup’s extractive strategies, plus the avoided costs of the outgroup’s extractive strategies. Each group aims to capture the double-dividend

*permanently*, by innovating in the capture of *de jure* and *de facto* resources. The high stakes of permanently winning or losing the double-dividend leads to the *primacy of victory*: the costs of being the dominated group are larger than the costs of sacrificing institutions, norms, and truth. All other ingroup commitments are put on hold until victory is secured.

Section 3 turns to the role of "professions", particularly the judiciary, media, and military. If professionalised actors hold broad other-regarding preferences and wider society becomes parochial, then professions may have an inertial effect—slowing institutional capture, continuing to supply accurate information, and so on—and helping democracies survive brief flirtations with parochiality. However, this function makes it a target of innovative attack, including attempts to insert parochial actors into professions. Like the electoral fulcrum, the professional fulcrum is not inherently liberal; as professions become dominated by parochial actors, the commitments leveraged by the fulcrum become parochial.

Section 4 turns to the race over political information. Political information offers two kinds of value: (A) an accurate picture of policy platforms and candidate probity, and (B) strategic value for a given candidate. As parochiality intensifies and the stakes of victory grow, the demand for and supply of information becomes dominated by strategic considerations. Ingroup audience preferences become perversely aligned with ingroup candidate preferences: both prefer (mis)information that benefits the candidate. The double-dividend therefore translates into a fig leaf for corruption.

Section 5 closes by arguing that the double-dividend turns into the *strongman's dividend*. Intense parochialists prefer a leader who is willing to capture institutions, violate norms, and distort the truth—to "weaponise everything"—insofar as doing so overcomes obstacles to ingroup victory. This helps explain some puzzling patterns in voting (e.g. evangelicals for Trump), and why voters who support democratic values nonetheless vote against democracy: the ingroup must capture democracy, or the outgroup might capture it first.

To conclude, the self-interested voter is rare and of trifling significance to democracy. Democracies are born of motivational commitments and, in our age of "voters against democracy" (Norris, 2021), they die by motivational commitments. If we can provide the two democratic fulcra, the final question is how to maintain broad trust.

## 1 The parochial voter: Turning out for the ingroup

I begin with the parochialist's decision to vote. The formal argument entails some obvious modifications of that in Chapter 16 and is placed in Box 17.1; I discuss the implications informally here.

How much an individual weights the payoffs of others depends on their group affiliations. In a society with cross-cutting group affiliations, these weightings may take complex patterns, but I will focus on the simple case of two parochial groups: an ingroup and outgroup. For a parochial actor, policy benefits for the ingroup are positively weighted, while benefits to the outgroup are valued less, not at all, or even negatively. I lay particular emphasis on the case of "absolute parochiality", where there is no spite (i.e. no negative $r_c$ for the outgroup), but voters are simply indifferent to the outgroup. Mere indifference has surprisingly grim consequences.

These absolutely parochial individuals will vote when the weighted ingroup benefits are larger than the cost of voting. Interestingly, individual tendencies to vote and donate increase sharply and monotonically with their degree of polarisation (e.g. Pew 2014).[232] Such patterns have been observed in laboratory studies, such as Fowler & Kam's (2007) Dictator Game experiments (also see Fowler, 2006). This is consonant with the literature discussed in Chapter 15, where between-group conflict is expected to be a profound stimulant to collective action. Parochialists appear to be strongly committed to their ingroups, while "liberal" individuals' commitments to others is broad but may be relatively tepid. This may reflect differences in perceived stakes, a matter I turn to next.

[232] This pattern may also be explained where individuals who are generally prosocial but ideological believe there is more at stake (i.e. greater evaluations of $B_g$), as the analysis in this chapter suggests will often be the case (in a parochialist's "double dividend").

**Box 17.1 Some formalisation: Parochialism and turnout**

Parochiality is defined as within-actor heterogeneity in the weighting $r_c$ applied to others' payoffs. In a fuller model, the individual may weight others' benefits and harms according to their various group identities (1 to $n$). Expected policy benefits $b$ will vary across groups, and as in Chapter 16, Section 1, we can take $\bar{B}_o$ as the average benefit for a sub-group member.[233] We then put this into the decision to vote from Chapter 16.

The generally prosocial version of the decision to vote from Chapter 16 was:

$$r_c K \bar{B}_o > c_v$$

The version here is:

$$r_{c1} K \overline{B}_{o1} + r_{c2} K \overline{B}_{o2} + \ \ldots \ r_{cn} K \overline{B}_{on} > c_v$$

We can further simplify by exploring the main dynamics in a *two-group* system, where the individual may have differential regard for an ingroup and outgroup. We will use the subscript "1" for the ingroup and "2" for the outgroup. Thus, on the question of the returns to voting, we get the more manageable

$$r_{c1} K \overline{B}_{o1} + r_{c2} K \overline{B}_{o2} > c_v \qquad \textbf{Eq. 17.1}$$

We are interested in parochiality, such that $r_{c1} > r_{c2}$. There are three cases:

1. Relative parochiality, where outgroup welfare is weighted positively but less than ingroup welfare. Thus $r_{c2} > 0$, but $r_{c1} > r_{c2}$.
2. Absolute parochiality, where outgroup welfare is irrelevant and ignored. Here $r_{c2} = 0$, and there are only commitments to the ingroup.
3. Spiteful parochiality, where outgroup welfare is negatively weighted. Now $r_{c2} < 0$, and the individual has some commitment to harming the outgroup.

We will focus on case #2, without spite, which is enough to have severe consequences. If the voter is absolutely parochial and weights the outgroup according to $r_{c2} = 0$, then the condition for voting in Equation 17.1 turns into

$$r_{c1} K \overline{B}_{o1} > c_v$$

[233] Alternatively, in the norms conception of moral behaviour, we can allow agents to differentially weight their duties to others according to group identities—as observed in Mohr and Larsen (1997), Braman and Lambert (2001), Wright at al. (2013), Yudkun et al. (2016), Solaz and De Vries (2018).

Note that we avoid the "altruistic explosion", where even slightly other-regarding voters may behave like martyrs, by the same means discussed in Chapter 16, Section 1.3.

Cases #1 and #3 raise the same problems as absolute parochiality, only in ameliorated or aggravated form respectively. I will touch on the case of spite, in which case we keep the full Equation 17.1, and the weighting applied to outgroup welfare is negative ($r_{c2} < 0$). Thus, a candidate may attract voters by proposing policies that actively harm the outgroup ($\overline{B}_{o2} < 0$). Indeed, a platform that has no effect on the ingroup but harms the outgroup may be attractive.

## 2 Parochial policy preferences

The absolute parochialist's optimum is straightforward:

- Chapter 16-like treatment of the ingroup. That is, a democracy of the ingroup, with protection of ingroup rights, increased investments in efficiency, and some ingroup redistribution; and
- Olsonian dictator-like treatment of the outgroup, with any cost imposed on the outgroup so long as it has net benefits for the ingroup.

Take a case similar to that of US slavery in earlier centuries. If whites identify with other whites (e.g. formally in Box 17.1, if $r_{c1} = 0.2$) but are indifferent to blacks ($r_{c2} = 0$), then this provides sufficient motivation for whites to engage in collective action to exclude blacks from political participation, impose white-favouring rules, and indeed to enslave the black population wherever doing so is ingroup-beneficial. Whites need no special antipathy towards blacks—although such antipathy may indeed be present—but indifference alone is enough to generate behaviour that we might describe as group-level psychopathy. Crucially, such group-selfish behaviour depends on other-regarding group members. The "trustworthy" member of the slaveholding society is willing to pay costs to return others' "property", to vote in support of the system, and perhaps go to war for it.

Extraction may be constrained by the now-familiar Olsonian mechanism: the ingroup majority is payoff-dependent ($r_s > 0$) on the outgroup minority. Olson expects the self-serving majority to be more efficient than the income-maximising

dictator, given the electoral majority may typically earn at least 50 percent of market income and so have an $r_s$ value greater than 0.5. This constraint does not bind tightly enough to put the losing group in a secure position:

- First, it still allows considerable extraction; at $r_s = 0.5$, the ingroup majority benefits from any transfer of $1 that has less than $2 of inefficiency costs. The outgroup must hope that available extractive strategies are inefficient, and will remain so in perpetuity.[234]

**Box 17.2 Formalisation: Ingroup majority preferences over outgroup treatment**

Following the pattern of Chapter 16, let us see what parochial preferences do to the Olsonian majority's incentives with respect to the outgroup. Take the voting majority as an ingroup, and say there is a roughly 50/50 ingroup/outgroup split in the population, and the ingroup wins electoral victory. The result of the analysis is straightforward: we do not modify Olson's original equations. In both Olson's case of self-interested voters and in this case of a majority ingroup, the goal is to maximise ingroup majority income—and this can be done by maximising $r_s - c$, where $r_s > 0.5$.[235]

$$B_m = r_s b_e - c$$

Recalling that $b_e$ is the component of electoral minority (in this simplified case, the outgroup) benefits that are economic and so relate to efficiency. How this differs from the result in Chapter 16 is that the ingroup completely ignores outgroup pure welfare effects ($b_w$). Thus, they will accept outgroup slavery, disenfranchisement, and so on, if it raises net material extraction. They will also pursue moral/cultural programs that raise ingroup welfare,[236] while being indifferent to effects on outgroup welfare.

[234] As discussed in Chapter 14, Olson points to USSR strategies of pressing workers towards subsistence, so that high effort must be maintained for survival.
[235] It is a little different if the ingroup is much larger than 50 percent of the population, and some of the ingroup are not in the voter majority. Then the voter majority follows the other-regarding pattern from Chapter 16 with respect to ingroup minority members: they choose greater efficiency and support some transfers. But that generosity is not extended to outgroup members.
[236] I include these ingroup gains in $c$ here, recalling that $b$ refers to others' (in this case the outgroup's) payoffs.

- Second, a ruling ingroup can credibly commit to not expropriating from its own members, whose welfare it values. This allows the ingroup to securely invest in capital and human capital, and so act as a substitute for weak outgroup investment. There will be some loss of natural talent, but compensatingly the ingroup will dominate the capitalist and high-skill classes (see Becker, 1957/2010, and Krueger, 1963, on the economics of discrimination). This is a familiar pattern in societies with one dominant group.
- Third, material extraction is large enough, but the ruling ingroup majority will also have the power to impose its cultural and moral preferences, with potentially large welfare impacts. Cultural and moral disputes, and disputes over status and the identity of the nation state, frequently dominate between-group conflicts. These are less constrained, if at all, by Olsonian efficiency considerations.

### 2.1 The primacy of ingroup victory: The parochialist's "double-dividend", the threat of permanence, and the rise of spite

Now consider the stakes of parochial political contests. If an absolutely parochial ingroup can decisively win power, it gains the parochialist's *double-dividend*: the benefits of the ingroup extractive platform, plus the avoided costs of the outgroup's extractive platform. To these material costs we can add moral-cultural preferences, including the value of group status and its association with the nation state. We can relate the double-dividend to the distance between the different "bliss points" of groups in the model of Gennaioli & Tabellini (2019, p. 12), although their model addresses questions that are raised in the final chapter.[237]

The double-dividend may be very large,[238] but most animating of all is the threat—and promise—that it may be lost or won *permanently*. Whichever group wins a

[237] Their excellent paper examines how voters may switch social affiliations, and particularly the economic causes of parochialism. I discuss this and other research in Chapter 18.
[238] Where strategies are equally distributed in the parasitism and mutualism quadrants, ingroup victory would mean that the outgroup can expect permanent average payoffs of zero, while the in-group captures twice the income growth expected under pure mutualism. Initial parasitic payoffs are potentially much larger; growth in average income through mutualism is constrained by the overall rate of economic growth, while growth in income through parasitism is constrained by the (much greater) potential for redistribution of existing wealth and earning power. At the extreme, it is the value of preventing total expropriation of the in-group, plus winning total expropriation of the out-group—or the difference between the total loss or total capture of the social surplus.

political victory in one period will tend to emerge stronger, being able to modify the allocation of *de jure* power by capturing political institutions and disenfranchising opponents, and *de facto* power via wealth transfers. Each political contest results in a shift of power to one group or the other; some elections may be perceived as the last chance to halt the process of ingroup decline and begin the ascent back to glory. The rhetoric of finality is profoundly important in stimulating collective action, and it is commonplace in observed between-group contests, including recent US elections (see Box 17.3).

The double-dividend and the possibility of its permanence leads to the *primacy of ingroup victory* and between-group politics as warfare. As Guriev & Papaioannou (2022) summarise, drawing on Mudde (2004; 2007):

> *The debate is no longer about tax rates or welfare state policies, but about the quintessential survival of the nation. This "war-like" situation entails aggressive tactics and controversial messages, often labeling opponents as "enemies of the people."*[239] (p. 10)

Resulting dynamics are analogous to those discussed in relation to the dictator in Chapter 14: there, it is only the largely unopposed "uber-dictator" that can focus on extracting according to Olson's rules of efficiency. Here, the "uber-group"—i.e. the ingroup with no fear of losing power—could do the same. Where power is contested, however, the vulnerable dictator and vulnerable group prioritise strengthening the co-optation and repression apparatus that controls their opponents. This draws even absolute (i.e. outgroup-indifferent) parochialists towards costly spite-like behaviour, where each group will pay costs so long as the costs imposed on the other are larger. It may extend to actions such as genocide. Note that Olson's model would rule such strategies out given their large costs for economic output, but they become explicable as efforts to win between-group conflicts permanently. Technically, the group need not be motivated by hatred, but hatreds naturally arise; human emotions appear to exist partly to motivate strategic, group-favourable behaviours.

[239] Take, for example, the Manichean rhetoric of Bill Walton, of the conservative Council for National Policy (Wilson, 2021, para. 5): "This is a spiritual battle we are in. This is good versus evil. We have to do everything we can to win."

**Box 17.3 The 2016 U.S. election: Fear and the permanent double-dividend**

The 2016 U.S. election introduced the rhetoric of between-group warfare into the mainstream. A notable example is the influential "Flight 93 Election" essay in the *Claremont Review of Books*, which compared the political contest to the fight on a domestic flight captured by al-Qaeda terrorists on 9/11 (Anton, 2016). The terrorists, in this metaphor, were the Democratic opposition. It encouraged conservative readers to "charge the cockpit", otherwise "death is certain" (para. 1-2). A Democratic victory would be "Russian Roulette with a semi-auto". It offered the theory that the Democrat party had a long-standing pro-mass-immigration position in order to import "third world foreigners with no tradition of, taste for, or experience in liberty", to achieve "*permanent victory*" [italics mine] (para. 28). Tellingly, Anton argued that Democrats' votes from non-white minorities allowed them to ignore democracy, because in his conception, legitimate democracy can only be a democracy of the ingroup and an outgroup-majority democracy is intolerable domination.

We can characterise this latter effect as a form of *negative* payoff-dependence, that may completely occlude Olsonian payoff dependence. That is, payoffs to the outgroup strengthen the outgroup and so are valued negatively. Again, Olson's constraints only bind for the "uber-group".

## 2.2 Group-level punishment, balances of power, and the Red Queen's race between groups

The threat of destructive between-group conflict may also motivate restraint and cooperation. In the terms of this framework, the $r$-based causes of between-group cooperation (here weak) have a substitute: $c$-modifying "rules".[240] If we think of the between-group contest as a form of Tit-for-Tat, and we assume that (A) powers to Tit and Tat are symmetric and (B) groups are rational, then mutualism may be sustained forever despite parochiality.

In practice, dependence on rules returns us to the problems of Chapter 8 under a new guise: because neither group is in fact committed to cooperation, there is no

[240] Between-group punishment can be characterised as an informal rule, saying "if we detect the outgroup doing X, we will do Y."

long-run anchor for the cooperative equilibrium. Instead, the two are engaged in a Red Queen's race to devise superior strategies for controlling one another. Strategic innovation proceeds, and one group eventually wins, or perceives that it has won, the upper hand and chooses conflict.

Balances of power in between-group conflicts have been examined extensively in the economic and political science literature, with results that fit this framework well. Where a large and/or sufficiently armed majority expropriates a small minority, we get a winner-take-all outcome and the parochialist's double-dividend may be realised. Large-scale conflict is suppressed because the chances of successful rebellion are low. This is the case of the uber-group, and it is stable so long as ingroup solidarity is sustained. In highly heterogeneous or fractionalised societies, meanwhile, there is no majority or large minority. Here we also get relatively low likelihoods of large-scale conflict, in this case because groups collaborate to ensure that no single group can dominate the rest (Montalvo & Reynal-Querol 2005; Collier 2001). That is, the parasitic double-dividend is difficult to realise and the stakes are lower. In circumstances of high fractionalisation and low levels of between-group trust, empirical studies find the lesser ills of inefficient policymaking, extensive rent-seeking, and economic stagnation (see Collier & Hoeffler, 1998, Easterly & Levine, 1997, Alesina et al., 1999, Alesina et al., 2003, Fearon & Laitin, 2003, Alesina and La Ferrara, 2005 and Desmet et al., 2012).

A balance of power between groups of similar size and resources has complex but predictable effects. On the one hand, it favours mutualistic negotiation in order to avoid the high costs of conflict. On the other, the stakes of gaining control of the institutional structure are extremely high, especially in winner-take-all systems with a powerful executive. If the balance appears to be shifting from ingroup to outgroup control, for example with changes in voting population sizes or shifting strategies, the threat of permanent loss of the double-dividend looms large. Esteban & Ray (2008) find that conflict in societies with balanced, parochial groups is infrequent, because it is so costly, but when it occurs it is very intense. Empirically, societies with groups that are relatively balanced are most likely to experience civil war (e.g. Horowitz, 1985/2000, Esteban & Ray, 2008, Montalvo & Reynal-Querol, 2005, and

Sidel, 2007).[241] In the most striking result, Montalvo and Reynal-Querol find that 90 percent of the most ethnically polarised countries—that is, countries with the largest culturally-distant minorities—experienced civil war between 1960 and 1995.[242]

This is a group-level pattern in the mould of Chapter 8: Balances of power cannot release us from Red Queen dynamics, which attack all truces built on temporary symmetries.[243] Certainly such truces are important in the short-run; for example, while support for between-group violence rises in the U.S., the uncertainty of the outcome of violent conflict acts to restrain partisan actors. Long-run cooperation between groups, however, is only possible if anchored by between-group trust.

### 3.1 Professions and resistance to parochialism

Let us follow the pattern of Chapter 16 and turn to the second $r_c$-leveraging fulcrum, the profession. Again, professions arise from a combination of

- extrinsic insulation, institutions that limit the influence of private rewards ($c$), so that other-regarding preferences (and $r_c b$) become the main influence on choice; and
- intrinsic augmentation, where training, established traditions and rituals, and so on, are used to enhance the salience of and commitment to the values of the order (and so the relative strength of $r_c b$ over $c$).

What happens to professions in the liberal society with increasing tendencies to parochiality? A detailed investigation is beyond scope, but I can make a few basic observations. We can expect even well-consolidated democracies to experience sporadic episodes of between-group antipathy. Professions may help democracies to survive flirtations with parochiality, on the condition that extrinsic insulation remains robust, such that these orders are protected from interference by emerging

[241] Sidel (2007, ch. 6), for example, examines patterns of religious violence in democratising Indonesia, where the loosening of old power structures under the Suharto regime, and prospect of competitive elections for determining power, led to particularly intense religious violence in localities where there was no clear majority.
[242] Consistent with Desmet, Ortuño-Ortín & Wacziarg (2009), who find that deep ethnolinguistic cleavages are predictive of civil war.
[243] Balances of power must be accurately perceived if they are to prevent conflict, so that conflict may arise from misjudgements of group strength.

parochial leaders, and intrinsic augmentation remains robust, such that these orders remain committed to democratic values even as broader society abandons them.

Professions with these properties, and that are loci of special powers—the judiciary, media, and military—can be expected to exert considerable inertia on institutional change. While the committed judiciary stands, it can continue to penalise violations of democratic law and slow institutional capture. Professional political journalists will likewise persist in investigating and reporting on events, preferring to supply accurate information and hold political actors accountable, rather than serving some narrow interest. Both depend on enforcement agencies and the military for protection from any emerging threats of violence, as guarantors that law and the constitution—as interpreted by the courts—will be enforced.

These inertial fiduciary orders constitute what some contemporary U.S. parochialists have called the “deep state”—i.e. the powerful actors in the Justice Department, the bureaucracy more broadly, and various civil society institutions who act to slow the change desired by radical populist movements (as explored by Michaels, 2018). Such actors will naturally be illegitimate for majoritarian democrats who advocate an unfettered executive. They are more legitimate to conservatives in the Burkean mould, who view value-formation as a longer-term project; these orders act as reservoirs of values that have been developed and carefully adapted over decades and centuries. They are, in liberal democratic societies, *repositories of trust*—the golden (or for the parochialist, leaden) core that gives the anchor much of its weight.

### 3.2 The parochial profession

The conditions for inertia—extrinsic insulation, intrinsic augmentation, and some special powers and protections—are vulnerable to persistent attack. A key goal for the parochial actor is the parochialisation and enervation of professions.

First, there is nothing about extrinsic insulation that makes it necessarily serve the liberal society. Extrinsic insulation only serves to reveal underlying patterns of $r_c$, and if judges, journalists, and generals are enthusiastic parochialists, then revealed virtue is replaced by revealed vice (Box 17.4). Stout’s (2001) "altruistic hierarchs" remain altruistic, but to some more than others.

**Box 17.4 Group commitments and judicial penalties**

Empirically, tendencies for parochialism—conscious or not—have some influence on the use of discretion in liberal systems. In the U.S., a large literature has noted differential sentencing by ethnicity, gender, and age for more or less identical crimes (e.g. Steffensmeier et al., 2006). The average penalties that partisan judges apply to firms for violations relevant to parties' political positions—e.g. hiring illegal immigrants or pollution violations—sometimes differ by orders of magnitude (Gormley et al., 2022).

There is growing concern about increasingly partisan judges deciding electoral disputes (e.g. Kang & Shepard, 2016, and Robertson, 2018),[244] and given trends in U.S. judicial appointments, the higher the court, the more partisan the judges' decisions (Bonica & Sen, 2017). Similar ingroup biases have been observed in policing, education, and healthcare, in the U.S. and in other countries (e.g. Grossman et al., 2016, on sentencing appeals in Israel).

Second, innovation will proceed and find vulnerabilities in extrinsic insulation—via unanticipated means of exploiting or reinterpreting formal and informal institutions, and via collective action especially in the form of violence.[245] The threat of violence makes enforcement agencies and the military crucial players and the final backstop for the decaying democracy. If they remain committed to democratic values, they may provide the time needed for the majority to reconsider the wisdom of conflict. If they are themselves strongly parochial, responses to violence will be biased, with the consequence that threats made by the favoured group become more credible. Successful coordinated actions by parochialists may thereby lead to the departure, expulsion, or otherwise elimination of non-parochial members of professions.

Once professions are dominated by parochial actors, self-regulatory processes serve to reinforce these values and hone the profession into a tool for between-group conflict.

[244] In what amounts to a failure of institutional design, "judges must generally have strong political affiliations to reach the bench at all, regardless of whether they are elected or appointed." (Robertson, 2018, p. 739)

[245] Failure may be especially rapid where that insulation is informal, a matter of convention rather than law. In many democracies, the executive has power to appoint judges, and the convention of non-partisan appointments rests on opposing parties recognising one another as legitimate.

## 4 Parochialism in the information race

So far I have established that parochial individuals are willing to engage in collective action, and that this may lead to destructive between-group conflicts as each side becomes devoted to winning the double-dividend. Conflicts may be subdued by perceived balances of power, but amid the Red Queen's race, this condition is temporary. So long as preferences remain strongly parochial, conflict is coming.

Now I examine how parochialism, and the primacy of victory, shape preferences over political information. In turn, this affects the capacity for voters to act as a check on political power and corruption.

### 4.1 The parochial preference for strategic misinformation

There are two main reasons information can be valuable for an individual:

1. Information may *improve estimates of private and group benefits* associated with a vote for a given candidate, via an evaluation of policies or a more general assessment of candidate trustworthiness and competence.
2. Information has *strategic* consequences, in that regardless of its accuracy it *affects the probability of a given candidate's victory*.

These relate to the two fundamental motives for seeking information identified by cognitive psychologists: an "accuracy" orientation, where individuals seek to build an accurate picture of the world, and a "directional" or "goal" orientation, where individuals seek information that is useful for achieving particular social ends (Kunda, 1990, pp. 481-482; also see Osmundsen et al., 2021 for an application to political information-seeking). To pursue, disseminate, and use accurate political information is to provide a democratic public good. To do the same only for information that is strategically favourable for the ingroup is to provide an ingroup-specific public good in the context of any between-group contest. These two types of value can be taken as orthogonal: information that is accurate may be strategically favourable or unfavourable for a given candidate.

As parochiality increases and the double-dividend grows, the value of information is increasingly determined by its strategic effects. Osmundsen et al. (2021), for example, report that it is not ignorance but polarisation, and the desire to win group conflicts, that drives the sharing of fake news (also see Oliver & Wood, 2014, on

partisans' enthusiasm for conspiracy theories). The strategic effects of information are of two main types.

- The first is where information provides favourable evaluations of ingroup candidates, and unfavourable of outgroup candidates. Its value is in shifting the voting intentions of unaffiliated voters or weakly partisan "leaners".
- The second is the value of information for mobilising and coordinating fellow partisans. Information is strategically valuable when it increases the perceived stakes of electoral victory and reduces the perceived private costs of cooperation,[246] and so may persuade ingroup members to pay the costs of collective action. For the strong reciprocator, these signals increase the motivation for voting, donating, volunteering in campaigns, protesting, intimidation, and violence. It is instructive to consider the conversations taking place on far-right online fora prior to the 2021 US Capitol attack: these are places where individuals trade outrage-inducing information about outgroups, celebrate and reaffirm their commitments to engage in collective action, and coordinate actions to solve "stag hunt" type problems (e.g. see Turner, 1964, p. 405, and Horowitz, 2000, p. 74, who examine the role of rumours in coordinating behaviour in ethnic conflicts).

The emphasis on strategic value has pernicious effects on the average accuracy of supplied information. As parochialism intensifies, audiences increasingly favour information suppliers—media outlets, social media personalities, political leaders, and so on—who are trustworthy suppliers of strategically valuable (mis)information. Where a usually trusted supplier violates that trust by disseminating information that is strategically damaging, even if accurate (e.g. Fox News reporting President Trump's 2020 election loss, see Griffing, 2020, December 16), parochialists regard them as "traitors".

In the *TMS*, Smith put it thus:

> *A true party-man hates and despises candour; and, in reality, there is no vice which could so effectually disqualify him… as that single virtue.* (p. 179)

[246] E.g. the private costs of collective action frequently decline as more participants contribute.

## 4.2 The double-dividend as fig leaf for corruption: Victory trumps probity

The emphasis on strategic value expands opportunities for political corruption and institutional capture. The broadly other-regarding voters of Chapter 16 make some investments in running the Red Queen's race over political information, and so act as a modest check on corruption. Adding parochialism changes this. The size of the double-dividend will be much larger than the costs imposed by a corrupt ingroup leader. Accurate information about ingroup corruption is therefore harmful, and to be suppressed or justified. The greater the perceived double-dividend, the larger the fig leaf for ingroup corruption.

Put another way, parochialism results in the perverse alignment of ingroup voter and ingroup candidate preferences over information. This subverts the usual, expected role of the voter. In a typical framing of democracy, electoral candidates prefer strategically favourable information[247] and voters prefer accurate information. The voter races against dissembling politicians. With parochiality, however, both ingroup candidate *and* ingroup voter preferences converge on candidate-beneficial information. Rather than voters racing against dissembling politicians, ingroup voters and their politicians race against outgroup voters and their politicians to control the information environment.

Eggers (2014) captures the logic of this "under-appreciated consequence of partisanship" like so: "the more voters care about which party wins, the less they can deter individual wrongdoing" (p. 441). His analysis of the 2009 UK expenses scandal reveals this pattern, where MPs in strongly partisan constituencies were more likely to be corrupt "because they anticipated that the electorate would be more permissive". And indeed, MPs anticipated correctly: partisan voters imposed weaker punishments for corrupt behaviour. Other scholars investigating how voters trade-off candidate probity for ingroup strength include Rundquist et al. (1977), Powell and Whitten (1993), Solaz and De Vries (2018), Graham and Svolik (2020). Kaufmann (2004) and Wantchekon (2003) examine it specifically in relation to ethnic

[247] Candidates might be shaped by rc>0, but the private rewards c of winning power are overwhelmingly large. Even for the prosocial candidate, the ends (i.e. enacting a platform they see as group-favourable) may justify the means (some strategic lies).

divides.[248] We may also see this in the trading-off of moral values, as in the case of evangelicals voting strongly for Trump (e.g. see Margolis, 2020, on evangelicals' distrust of Democrats). Kayser and Wlezien (2011) find that weaker partisanship leads to stronger economic voting—i.e. attention to *general* measures of performance such as growth, inflation, and unemployment.

## 5 The double-dividend becomes the "strongman's dividend"

Where an ingroup leader holds political power, democratic institutions and norms, checks on power such as the judiciary and media, and truth itself will frequently be hindrances to the full use of power to achieve ingroup victory. Sacrificing them increases the range of strategies and resources available for use against the outgroup. In parochial contests, their sacrifice is rational: ingroup authoritarian rule is preferable to an "illegitimate" democracy that may hand power to the outgroup. This is especially so given the symmetric threat that the outgroup may itself pursue institutional capture. Accordingly, Graham and Svolik (2020) find that among both Democrats and Republicans, as partisanship intensifies, voters become increasingly willing to sacrifice democratic institutions and norms where this favours ingroup candidates, and Braley et al. (2023) find that partisans favour democratic capture to the extent that they believe their opponents plan to do the same.[249]

The double-dividend thus may turn into the *"strongman's dividend"*. That is, intense parochiality can set conditions for the transformation of democracy into a quasi-authoritarian or authoritarian system. This opportunity is not lost on opportunistic leaders. Effective demagogues throughout history have a knack for sniffing out distrust, expanding and exaggerating the gap between group commitments, and accusing opponents of trying to win a permanent victory. The size of the commitment gap they can create, and so the size of the double-dividend, is the measure of the strongman's opportunity to centralise power.

This is a simple model. But it tells us something about why between-group distrust is, empirically, so dangerous for democracy, and why corrupt political leaders and

[248] A related body of work examines votes for corrupt politicians because they provide a flow of resources to an ingroup through pork barrelling and patronage, which we might characterise as the leader distributing the double-dividend (e.g. Manzetti & Wilson, 2007).
[249] Similarly, so-called "constitutional hardball" by Republicans leads to proposals for the same by Democrats, e.g. Belkin (2019).

aspiring authoritarians throughout history focus so much attention on inflaming enmities. If liberal democracy rests on the broadly shared commitments of Chapter 16, then the framework of this thesis suggests that the first task of institutions is to maintain those commitments. The concluding chapter turns to consider this problem.

*Among the most important sources of legitimacy is widely shared prosperity.*
Martin Wolf (2016, para. 11)

# 18

# Concluding Part 4: The two-way "Road to Serfdom"

In 1944, two sons of Vienna published two seminal, and starkly opposing, works on the compatibility of capitalism and democracy—Friedrich Hayek with *The Road to Serfdom*, and Polanyi with *The Great Transformation*. Both tried to make sense of what Winkler (2015) called Europe's "age of catastrophe", which had brought two world wars and the downfall of glittering Vienna. Both found the answer in the mode of economic regulation, though they offered opposite counsel: Hayek warned that socialism was a "road to serfdom", and he beckoned us to free markets. Polanyi warned that it was free markets that would lead to fascism, and beckoned us towards socialism. Both were sceptical about the stability of the middle ground.

The question raised by Hayek and Polanyi, one that connects the two applications of this thesis, is a crucial one: how does the *approach to economic regulation* (Part 3) affect the *conditions for democracy* (Part 4)?

This concluding chapter argues that the "road to serfdom" goes in two directions. The "fatal conceits" of central planning, and of libertarianism or classical liberalism, make them incompatible with the other-regarding commitments that underpin democracy. The argument is a straightforward extension of material developed so far, so this final substantive chapter is a brief one.

## 1 Recap: Rules for a prosperous and stable democracy

It may be useful to briefly recapitulate the thread of argument through this thesis. Part 1 introduced its core ingredients: the two causes of cooperation and open-ended innovation. Part 2 introduced the Red Queen's innovation race, and the race between cooperators and defectors, that is born from these ingredients. In Chapter 8, this culminated in the argument that commitment and rules are essential complements.

Commitment anchors cooperativity, as the only force that can support adaptive investment in rules-as-public-goods. Rules in turn are used to (A) defend that commitment as the organising principle for the system, and (B) to optimise.

Democracy specifically is built from motivational commitments, which are harnessed by electoral and professional fulcra (judicial, military, and journalistic) to provide the key public goods of the system. The primary task of rules is to maintain trust and the integrity of the fulcra (the matters of Part 4), and the second is of course optimisation (the matters of Part 3).

## 2 The first road to serfdom: Central planning

The regulatory structure of the centrally-planned economy is incompatible with both optimisation and democracy. The optimisation problem was discussed in the conclusion of Part 3.

Here I briefly add some conclusions on the consequences of central planning for democracy. Following Hayek, the problem is primarily one of regulatory capture. *The Road to Serfdom* argues that the central planner must centralise power and information in order to achieve any modicum of efficiency, and that this creates the conditions for authoritarian control. In the terms of this thesis, absent a market, the central planner must build an extensive, and substantially discretionary, marginal institutional apparatus for policing effort and allocating resources. The state's resulting broad and intensive powers of detection and enforcement present abundant opportunities for a ruling party to construct a system of *co-optation and repression*—to use centralised knowledge and resources to reward political allies and punish enemies.

The democratic citizenry, meanwhile, faces an impossible detection and enforcement task: disciplining a planner that exercises complex discretionary powers throughout the economy. They cannot possibly keep up in the third leg of the Red Queen's relay.

## 3 The second road to serfdom: Libertarianism and classical liberalism

Libertarianism and classical liberalism also fail on both fronts. Again, the technical impossibility of efficient libertarianism and classical liberalism was the subject of Part 3. Here I add that they are the second "road to serfdom", in that they are also incompatible with democracy. There are in this case two mechanisms.

First, libertarianism and classical liberalism also produce conditions for regulatory capture, but rather than by a coordinated regulatory power, it occurs initially via an anarchy of roving banditry. Innovative firms exploit the non-adaptive ("sleeping") total regulatory structure, leading to rampant rent extraction. This both intensifies the *motive* for regulatory capture (to defend rents), and concentrates the *means* for regulatory capture, as income is funnelled to successful rentiers (empirically, leading to the growth of rents described in the second paragraph of this thesis). Further, the libertarian and classical liberal regulatory structure increases *vulnerability* to capture, by treating democratic public goods (political information, campaign funding, etc.) as private goods to be sold in markets. Conventional and social media ownership, and campaign contributions, become the domains of the rentier class.

Milton Friedman (1977/2018, §8, para. 9) observed that business was one of the "two greatest enemies of free enterprise". One contradiction in his system, then, is that it hands greater resources to this "enemy" and allows those resources to feed back into the political process. Friedman offered no remedy, only remarking that if rent-seeking became prevalent, the fault lay with citizens for being exploitable—for being "suckers" and "fools".[250]

Second, the rise of rents and growth in regulatory capture leads to gross inequalities and *declining trust*. Friedman is cavalier about the prospect of voters and market actors being "suckers", but the term he chooses is an apposite one beyond his intentions: the sucker, as a figure in the literature on strong reciprocity, has large consequences for cooperation. Following Section 4.2 in Chapter 15, relatively low levels of perceived cheating are enough to lead many strong reciprocators to defect. The evidence suggests that the sucker does not merely defect, but seeks a narrower ingroup that they perceive as trustworthy, in order to engage in conflict against outgroups that are perceived as having cheated.[251] With enough suckers, we get the strongman of Chapter 17.

---

[250] "I have no complaint of that, if we are suckers enough to provide them with this kind of a means to improve their own return". (Friedman, 1965, p. 15) "I don't blame corporate executives who lobby for tariffs. I blame the rest of us for being such fools as to let them exploit us. We're to blame, not them." (Friedman, 1990, para. 4)

[251] This dynamic has constructive potential if citizens are broadly trusting and informed. Smith's own sentiments, combined with his knowledge of the gross rents of the mercantilist system, manifested in his lifelong interest in reform. The mass public does not possess the advantages of Smith, and much hinges on the quality of political information.

### 3.1 Some empirics: Economic shocks, inequality, and parochial populism

Let us undertake one last empirical survey, to establish the connection between the failures of economic regulation discussed in Part 3 and the parochialism of Chapter 17. The political science literature is converging on a primarily economic explanation for the globally coordinated rise of parochial populist[252] parties and their "strongmen". The "emerging consensus", Guriev and Papaioannou (2022) write,

> *is that economic shocks may have activated pre-existing cultural divides and exacerbated polarization and identity politics.* (p. 87)

Their review of the literature is magisterial, and the interested reader should look there for details. Here there is space only to provide a brief overview and point to some selected papers.

Initial research emphasised cultural change as the driver of populism (see Box 18.1), but today an extensive body of evidence points to a strong causal link between populism and economic shocks of various kinds—from trade and automation, to unemployment, housing insecurity, education, and austerity (with Guriev and Papaioannou, 2022, pointing to more than 50 papers across these areas). Effect sizes are significant. Frey, Berger, & Chen's (2018) estimated effects for just one of these factors in isolation, automation, are enough to swing key U.S. states in 2016 from Trump to Clinton and deliver a Democrat national victory. In Europe, Algan et al. (2017) find that a one percentage point increase in unemployment is linked to a one percentage point increase in votes for populist parties, and Dehdari (2020) finds that two laid-off native workers translates into one extra vote for the populist Sweden Democrats.

The connection between economic shocks and populist voting is indirect: as suggested, economic shocks produce distrust and activate between-group conflict, such that voters report being motivated by cultural antipathies. Ferrara (2023), for example, finds that Chinese import shocks causally drive negative attitudes to minorities and immigration. Autor et al. (2020) similarly conclude of trade-induced populism that "vitriolic campaign rhetoric is indicative of underlying economic

[252] Rodrik (2018a) defines populist parties "loosely as those which pursue an electoral strategy of emphasizing cleavages between an in-group and an out-group." (p. 41)

pressures" (p. 43, see Algan et al., 2017, Colantone & Stanig, 2018, and Barros & Santos Silva, 2019, for similar interactions between economic shocks and cultural cleavages).

The historical evidence adds further weight to the argument. Far-right, parochial populism rose in Europe in the wake of the Great Depression (Capoccia, 2005;[253] de Bromhead et al., 2013), following recessions in the 1970s and 1980s (Jackman & Volpert, 1996), and following financial crises over the past one and a half centuries (Funke et al., 2016).[254] To that historical pattern, we may now add the GFC and rise of rents in contemporary democratic capitalism.

Inequality plays an exacerbating role. Various authors have noted a tight relationship between rising inequality, falling trust, and widening ideological cleavages in the U.S. and elsewhere (Putnam, 2000; McCarty et al., 2016;). Inequality appears to be a strong (Delhey & Newton, 2005) or in some cases the strongest predictor of trust (Uslaner, 2002; Bjørnskov, 2007). As Bjørnskov (2007) writes,

> *[a] conclusion reached by virtually all studies is that income inequality is among the most robust cross-country determinants of trust.* (p. 5)

**Box 18.1 Culture or economics?**

Early analyses following the rise of Trump, most notably Inglehart & Norris (2016), attributed the rise of populism to cultural change, and found that economic insecurity made only a modest contribution. This finding, however, arose from a focus on *absolute* measures of economic insecurity. Certainly, Trump voters were wealthier on average than Clinton voters in 2016.[255] Yet a deep body of economic and psychological research shows that what matters is the threat of *relative* change in material circumstances against some referent (Crosby, 1976; Alesina & La Ferrara, 2000; Kingdon & Knight, 2007).[256] Collective action is strongest among advantaged groups that feel threatened with positional loss (e.g. see Walker & Mann, 1987; Mummendey et al., 1999; Jetten et al., 2015). Populist reactions to

[253] Capoccia (2005) identifies 24 fully democratic European countries 1920, but only 11 by 1939.
[254] Also see Garro (2021) on the link between U.S. state economic downturns and polarisation.
[255] In terms of absolute measures, Trump voters are not disadvantaged: Rothwell & Diego-Rosell (2016) estimate mean household incomes of $81,898 and $77,046 for Trump-backing and Clinton-backing households respectively.
[256] Although some support for absolute economic deprivation for Brexit and 2017 French election (Becker et al., 2017).

economic shocks tend to be greatest, therefore, in the section of the working class with intermediate levels of education and specialised skills, while the poorest, lowest-trust groups are more likely to abstain from political action (Oesch, 2008; Bornschier & Kriesi, 2013; Bornshchier, 2018).

Zak & Knack (2001), for example, find that a one-point increase in the Gini coefficient is associated with a 0.76 percentage point decline in generalised trust (similar estimates are found in Bjørnskov, 2007). The model suggests that inequality explains about 40 percent of the trust gap between the extreme countries of Norway and Peru.[257]

Inequality is linked to increased populist voting (Georgiadou et al., 2018), and may operate by increasing positional precarity and status anxiety (Sanchez-Rodriguez et al., 2019; Buttrick et al., 2017; Layte & Whelan, 2014; Wang et al., 2019) and intensifying political reactions to economic shocks. Rogowski and Flaherty (2021), for example, find that globalisation-related shocks increase far-right voting in Europe only under conditions of inequality. Interestingly, Caprettini et al., (2019) and Albanese et al. (2019) find that redistribution reduces populist voting, and Rodrik (2018b) argues that the U.S. avoided the broad decline of democracies following the Great Depression via the redistributive New Deal.

A particularly interesting wrinkle to the story is what happens when inequality draws a society into the Chapter 15 "low-trust trap". Rothstein & Ulsaner (2005) combine the negative effect of inequality on social trust, and the negative effect of low trust on political participation, to propose a trap where:

> *Social trust will not increase because massive social inequality prevails, but the public policies that could remedy this situation cannot be established precisely because there is a genuine lack of trust.* (p. 70)

Several other papers suggest that, in response to inequality, high-trust voters favour government redistribution, while low-trust voters turn to the far-right and choose group conflict instead (e.g. Kuziemko et al., 2015; Algan et al., 2018; Alesina et al., 2018; and Di Tella et al., 2021). The parochial reaction is rational enough under very

[257] World Values Survey question "most people can be trusted". Values taken from Inglehart et al. (2000) and range from 6% in Peru to 61% in Norway. Ginis of around 25 and 55 at the time.

low trust; it would then be implausible to believe that enough cooperators could, in good faith, coordinate to restore fairness. Rather, there is a war to be won.

*…change is certain, progress is not.*
Ernest Campbell Mossner (1941, p. 667)[258]

# Conclusion

## Finding the "Middle of the Road": Committed actors, adaptive institutions

We arrive at the conclusion of the thesis. I briefly summarise the general theory developed in Parts 1 and 2, and the picture of liberal democratic capitalist orders developed over Parts 3 and 4. I also offer a summary diagnosis of the malaise that has struck liberal societies over the last few decades; it is certainly the result of innovative rent-seeking, but that rent-seeking has been unbridled by the broad influence of some misconceived theories about how markets and democracies work.

I close by discussing the role of economists and political economists in running the Red Queen's relay. Liberal democratic capitalism will not survive without their contribution—without their help treading a careful path through the middle of the road, between the two fatal conceits and two roads to serfdom.

### 1 The general theory

All social orders are engaged in various Red Queen's innovation races, against external competitors and internal defectors. In a world of Schumpeterian and Baumolian innovation, there are no lasting Nash equilibria—there is always a next move to play. The persistence of a social order is therefore a probabilistic matter. The "goal" of each order is to be as lucky as possible for as long as possible.[259] This may be a disenchanting thought, rather less welcome than the idea that we might rest one day, securely at history's end. But disenchantment is sobriety. We must make our

[258] Mossner, a specialist on Adam Smith and David Hume, summarising Hume's view of history.
[259] Whose goals? To borrow Friedman's (1953) terms, luck in the innovation race "summarizes appropriately the conditions for survival" (p. 158). Sociobiological systems have no goals, but evolutionary selection means they behave as if they do. We, meanwhile, consciously shape our orders, and most of us have some interest in their prosperity and persistence. This is, of course, because our minds themselves are the outcome of the evolutionary process; those individuals and groups that were uninterested in survival tended to disappear.

own luck, and the choices we make affect whether the odds are against Us or overwhelmingly in our favour.

Social orders better their odds in two ways. First, they find ingenious ways to run the adversarial innovation race faster than their opponents. Second, they better their odds by finding ways to reduce underlying adversariality, by harnessing commitment, so that they run the race with others rather than against them. Our little blue planet will always be dominated by systems that do both well.

This story began some four billion years ago, continued through the evolution of our species, and continues today. In the turmoil of the first half of the twentieth century, some societies found the wisdom to satisfy these two conditions better than any preceding society: they sustained shared commitments, and they used well-crafted markets to sustain rapid mutualistic innovation. Liberal democratic capitalism dominated the world. All of us who have enjoyed life within this system are indebted to these forebears; most of us, being strong reciprocators, will be willing to pay that debt. So let us see how long we can load the dice.

## 2 A picture of liberal democratic capitalism

I opened this thesis with the question: What sustains order in liberal democratic capitalist societies? The answers follow the pattern of the general theory developed in Parts 1 and 2, with elaboration on the specific sources of commitment and the specific innovation races to be run.

Liberal democratic capitalist societies rest on complex public goods that can only be supplied by committed actors. In keeping with Adam Smith's *Theory of Moral Sentiments*, other-regarding preferences are the only available tool for the job.[260] Also with Smith, and following the empirical evidence, this other-regarding force is only modest in strength. I have shown how democratic institutions create choice environments that leverage this modest force to move the world. These are the two fulcra of Chapter 15—large-scale elections and professions—and they leverage other-regarding preferences to supply key democratic public goods.

[260] The extrinsic form of commitment, namely structural payoff dependence, is absent in liberal democracies and anyway sustains no interest in democracy, rights, or welfare.

The task set for the two fulcra is to run the third leg of the Red Queen's relay, to shape the payoffs of the economic regulator of Part 3. In turn, the well-governed regulator shapes the payoffs of market actors, running the first two legs of the relay in the general interest. At both levels of the system, success depends on establishing a functional division of detection and enforcement labour—and particularly one that does not assign individuals within it impossible detection tasks.

Voters cannot possibly control the regulator without professions; this would make impossible demands of voter knowledge and collective action. Voters combine with the judiciary, police, military, journalists, and the academy, to form a working total system of detection and enforcement. Professions help to anchor the "direction" of the system, by acting as repositories of values formed over decades and centuries, and act to constrain voters and provide some inertia when tempers flare. This is the "deep state" at its best. Professions also help voters to maximise "pace", in a complex world where voters need many thousands of experts to inform them about the problems they face, the options available, and the trustworthiness and competence of those who serve them. Democratic institutions structure and simplify the problem for voters, so that most can play their role by reading some news, watching the TV, chatting with neighbours, and congregating at schoolgrounds, community centres, libraries, and churches once every few years to cast a ballot. It is a truly remarkable achievement.

Should that be sufficient to run the third leg of the Red Queen's relay, the remaining question is whether the regulator is competent enough to run the first two legs of the relay effectively. Here, detection and enforcement tasks are divided between institutions and market actors, with professions again playing an important role. Neither pure markets nor central planning can do the job. Economic institutions must structure and simplify the problem for market actors, so that most can play their role in driving ever-advancing living standards simply by shopping around for the best balance of quality and price. This system, too, is little short of miraculous.

Both involve squeezing maximal value out of the relatively modest efforts and/or knowledge of voters and consumers. Because the nature of detection tasks, and the relative capabilities of actors, change over time, these divisions of labour must be adaptive. If they adapt successfully, the result is broad prosperity, truthful

information, and the maintenance of democratic power. Individuals are given good reasons for thinking of themselves as members of a genuine group with a shared destiny—and they continue to play their part in running the race.

### 3 The liberal malaise: The sleeping Hare and loss of trust

We have been losing the Red Queen's relay. The two aforementioned marvels of political and economic order have become spectacles of incompetence and conflict. The fall has been disorienting in its rapidity. What has gone wrong?

I have argued that the liberal philosophies that have influenced policymaking in the last several decades are ruinously maladapted to the world of the Red Queen. They are recipes for social decay. They have two central flaws.

First, they subscribe to a fixed, non-adaptive system of regulation. Theorists such as Hayek claim the advance knowledge that basic market institutions are sufficient, in most cases or with minor tweaking, to domesticate innovations into the future. I have shown that their conception of markets is "organismic"; this is expressed explicitly at times by Hayek,[261] but more importantly built implicitly into the foundations of conventional economic thought. It is half an economics—an economics of the mutualism quadrant—where innovation is assumed benign. Human societies are conflicted systems, and domesticating innovation via detection and enforcement is incredibly demanding. It requires perpetual adaptation to innovative rent-seeking strategies that no armchair theorist can foresee. The unforeseeability of individual innovations is, I observed, both the blessing and curse of markets. The predictable result of the fixed regulatory approach is that it is picked apart by an accumulating set of counterstrategies. The end point is anarchy.

Second, they either deny the existence of the motivational commitments that ground the system, or in Hayek's case, regard such commitments as dangerous to the market order. "There is no such thing as society",[262] they assert, when the liberal order depends on the continual reproduction of the belief and trust in society. On both these matters they stand against Adam Smith, whose approach to markets was

[261] Again, Hayek (1933) describing the market society as "an organism in which every part performs a necessary function for the continuance of the whole without any human mind having devised it." (p. 130)
[262] A favourite saying of Margaret Thatcher, whose vision for society was shaped by Hayek.

pragmatic, and who understood the social role of sympathy. Not understanding the foundations of liberal societies, they inadvertently destroy them.

This manifests in various "sleeping Hare" problems. We have Red Queen's races to run, and it is worth a reminder that there is only one means of adaptation available to us: innovation in institutions.[263] Again, foremost is the maintenance of commitments and the fulcra, and so the reproduction of the system. The secondary task is to optimise. To abandon innovation in institutions is to lose these races, and lose our social order.

Part 3 focused on the regulator's withdrawal from adapting the "total regulatory structure" that governs the market.[264] Like every other set of fixed rules in nature, the fixed rules of the free market are attacked by a *progressively expanding* set of counterstrategies. Firms accumulate strategies for establishing market power and externalising costs. Firms accumulate advantages in information and the power to process it, and strategies for using complexity to craft products that "mimic" value.[265] Such strategies lead to the accumulation of economic rents described in the opening pages of this thesis, slowing growth and raising inequality throughout the market as a whole. When they proliferate in welfare-critical markets, their presence becomes more obvious in the form of the major crises discussed in the empirical supplements.

Part 4 argued that both motivational commitments, and the fulcra that leverage them, are decaying. First, the logic of benign innovation in minimally regulated markets was extended to management of the fulcra. In the case of political information markets, innovation is ambiguous, the product is complex, capabilities are asymmetric, parasitisms are lucrative, and "deregulation" places an impossible burden on audiences.[266] It is ripe for capture. In the case of electoral institutions, I noted in Chapter 14, Section 2.2.2, that money has increasingly found its way into politics, so that quid-pro-quo relationships have become a systemic element of the

[263] We cannot intervene in market actor capabilities, nor can we directly intervene in commitments. We affect them via institutions. Leadership, of course, is profoundly important for harnessing commitments, but that is not something we can design—it is part of the roll of the dice.
[264] I have suggested that the fruits of this approach are predictable. Keynes (1982), for his part, was prescient in his suggestion to Hayek that "[y]our greatest danger ahead is the probable practical failure of the application of your philosophy in the US in a fairly extreme form." (p. 387)
[265] Large firms have similar advantages over small firms, which likely contributes to the market concentration and divergence in profitability observed by Furman and Orszag (2015).
[266] These asymmetries bite harder in the case of political information markets, given political information is a public good rather than a private good, so audiences systematically underinvest.

institutional structure in some major democracies. This too is partly a consequence of the "organismic" conception of market societies, where any form of private optimisation is taken as aligned with the group interest. Second, the economic consequences of failing to run the first two legs of the Red Queen's relay, and of regulatory capture, sow distrust and parochiality. Voters observe that society does not appear to exist, and so they seek protection and power in subgroups they view as more trustworthy.

Each of the problems discussed in Part 4 is significant on its own. In combination, they create substantial risk of regime change. Informed voters have a chance of addressing economic dysfunction and regulatory capture, and understanding whom they should trust and distrust. If control of political information can be purchased, this function is compromised. The institutional division of detection and enforcement labour is broken. Instead of voters using professionalised political journalism as a means to realise their motivational commitments, rentiers use captured political journalism to defend banditry, in its roving or stationary (i.e. authoritarian) forms.

Throughout this analysis, I have found contrasts between central planning and the free market a useful device, because the two are mirror images. Both commit to opposite but equally dysfunctional divisions of detection and enforcement labour, that exaggerate the powers of regulator or market actors. The former blocks market parasitism at the cost of losing its mutualisms, and the latter allows all mutualisms at the cost of permitting gross parasitisms. The former necessitates extreme and unregulatable government power, and the latter allows the steady growth of anarchy. Quite appropriately, the right-wing liberal and member of the Mont Pèlerin Society, Raymond Aron, called Hayek's scheme "inverted Marxism".[267]

I concluded that both are "fatal conceits" and "roads to serfdom". I have been comfortable applying the former term to Hayek, because of the extraordinary epistemological claims he makes, and the passion and unshakeable confidence with which he makes them. As Miller (2010), a prominent and enthusiastic summariser of his work remarks, Hayek made his claims with "a sense of finality". He humbles the

[267] See Gane (2016), who remarks that Aron believed Hayek "gave sovereignty to the market rather than to the state." (p. 11) Note that Aron would, in today's climate, be a moderate of the mixed economy.

interventionist planner, certainly, but he makes of himself an all-seeing deist (see Chapter 13). In doing so, he becomes the proper target of Smith's (1759/1976) old rebuke:

> *The man of system … is apt to be very wise in his own conceit; and is often so enamoured with the supposed beauty of his own ideal plan of government, that he cannot suffer the smallest deviation from any part of it*. (p. 275)

## 4 Finding the "middle of the road"

Recovery will be difficult in countries where trust has plummeted, where parochial reactions are strong, where professional motives have been eroded, and where the regulator is substantially captured. In the past, those societies that avoided fascist and communist reactions to the Gilded Age and Great Depression were able to do so by building political movements around credible liberal reforms. The promise, and the fruits, of those reforms restored the peace. We may hope that in this respect—if not in respect to the intervening wars—history may rhyme.

In this, economists and political economists have a significant role to play. They are racers in the Red Queen's relay. They generally are, or hope to be, racers on the side of cooperation; indeed, they are predicated on the Smithian alignment of private and social returns. They are professions, and their task is to simplify the complex options put before the public and policymakers. There is no future in which liberal societies recover without their contribution. I have suggested that the nearly universal rejection of central planning was one such pivotal contribution to the future of the liberal society—and credit to Hayek where it is due. The theory put forward in this thesis suggests that libertarianism and classical liberalism warrant the same fate, and this would be a contribution of similar importance.

I will close by pointing to a few ways that the analysis herein suggests our thinking should change, and draw us closer to the "middle of the road". The common theme is that effective institutions utilise the real *motives* and *capabilities* of the humans within them. Designing democratic institutions wholly on the extremes of self-

interest, or unconditional altruism, will undermine democracy.[268] Designing market institutions on the extremes of perfect market or institutional actor rationality, or total market or institutional actor incompetence, will undermine the economy—and then democracy.

### 4.1 Mixed capabilities and the mixed economy

The first two points relate to the matter of capabilities, and their implications for the design of economic institutions. Before turning to them, it is worth reiterating some core arguments from Part 3:

- We tend to think about neoclassical theory as a theory of markets, but it is, in itself, silent on the actual allocative efficiency of markets. A wide range of regulatory approaches is theoretically consistent with the neoclassical optimum—from central planning to *laissez faire*—in the special case that market and institutional actor capabilities take the right pattern. In practice, capabilities depart grossly from those assumed in both cases, differ across sectors of the economy, and also differ over time.
- We tend to reason about departures from these assumptions as producing "market failure". But there is a symmetry between the risks of reasoning about "market failure" from the default of perfect markets, and the (perhaps more intuitively obvious) risks of reasoning about "planning failure" from the default of central planning. Either is reasonable if actors' actual capabilities are reasonably close to those assumed. But reasoning from either default is hazardous where capabilities fall far short.

Indeed, it might be valuable if students were taught neoclassical theory in this way, such that the power and limitations of the theory were made more obvious.

The first point is that most real-world institutional structures lie in between the two fatal conceits. Whether or not it is anathema to say so, the dominant mode of organisation in capitalist societies is a form of *planned competition*. To be successful over time, that plan must be *adaptive*. We must, as North (2005, p. 122) writes,

[268] The motives side is related to Ostrom's (2010, p. 435-436) observation in her Nobel lecture: "Designing institutions to force (or nudge) entirely self-interested individuals to achieve better outcomes has been the major goal posited by policy analysts… empirical research leads me to argue that instead, a core goal of public policy should be *to facilitate the development of institutions that bring out the best in humans*." [italics mine]

structure markets until firms compete over price and quality—not over strategies for consumer exploitation, socialising costs, establishing market power, or tax evasion. What else are we doing when we say that food producers cannot poison consumers, destroy the environment, capture the whole market, or indeed bomb each other's factories? We set out a limited terrain over which the competition is to take place, and in doing so *we plan for the general consequences*: quality products at prices close to the cost of production. In this sense, Hayek has a plan—it has the merit of not assigning the regulator impossible tasks, but the demerit of assigning market actors impossible tasks. And as a fixed plan, it is of necessity a bad plan.

Where markets have low stakes or low complexity, the Hayekian approach with a little tinkering may work quite well. In cases such as financial markets, tinkering fails: there are large externalities, the public and private stakes are high, and products are hyper-complex. If we for some reason prize structurelessness, we are forced to manage it via "marginal" regulation and create an unmanageably complex regulatory structure (i.e. Basel III and its kin). Paradoxically, we end up needing the regulator to behave like a central planner—to act as a substitute for market actors in assessing the costs and benefits of activities in detail, and acquire extensive distributed knowledge to do so. It is the worst of both worlds: the planning is intensive, but the exploitability of such regimes means that anarchy slips through its cracks.

Effective regulation often takes the form of a market-friendly version of Polanyi's (1944) economics as "provisioning".[269] What Glass-Steagall did well was identify and isolate the *social functions* of commercial and investment banking. The task of commercial banks was to provide convenient services to retail clients, while providing the public good of safe and stable returns. Market competition was retained, but its terrain was constrained accordingly. Investment banks, as risk-taking allocators of capital for sophisticated clients, were regulated mainly by limits to scale and skin-in-the-game. It is such structural regulation that makes it possible to use markets effectively, without aping the planner or inviting anarchy.[270]

[269] I.e. economics is fundamentally about how we satisfy individual and societal needs.
[270] Note that other parts of the economy have little possibility of genuine competition (e.g. electrical networks). If we aim to squeeze some (largely speculative) mutualisms out of them via privatisation, we must be cognizant that doing so creates new Red Queen's races—i.e. actors with strong and

In short, we always have a plan for competition. We must adapt it as we learn more about our complex reality.

Second, to pull a general lesson out of the foregoing: *efficiency is simplicity*. It is policy such as Glass-Steagall (and perhaps the Bebchuk and Spamann, 2009, incentive alignment scheme of A.12E1.3), cash-flow taxation,[271] carbon taxation,[272] tariff reduction,[273] mineral extraction monopolies,[274] and the professionalisation of doctors and political journalism. These are policies suited to the world of the Red Queen. They slow the race. They also have second-order political consequences, which I have not had space to discuss: just as voters cannot police the discretionary powers of central planners, so they cannot police the hyper-complex policies of the marginal market regulator.

### 4.2 Harnessing and defending the moral sentiments

The next two points relate to making full use of human motives.

First, other-regarding preferences exist, and are harnessed in the fulcra to provide complex and high-stakes public goods. In the case of large-scale elections, the invasion of electoral politics by money is a well-trodden problem with obvious solutions, and needs no further elaboration. But it is worth restating some of the more novel claims about professions.

Professions are potent. Their track record is impressive, from the courts to the military, doctors, teachers, journalism, and the academy. We must take care not to destroy them by imposing competition and incentive schemes, which threaten their independence and confront the impossibility of codifying their complex work.[275] To design a profession, one must understand its social function. The function of doctors is not to sell pharmaceuticals; it is to act as trusted intermediaries that domesticate complex, innovative, profit-seeking firms, so they serve uninformed patients. This function is undermined if we allow the former to correlate doctors' returns with their own. Similarly, the function of political journalists is to provide the democratic

permanent motives for outsmarting and capturing the regulator. We must somehow be sure of winning these races, or the costs will outweigh the small benefits.
[271] As opposed to complex, and evidently highly exploitable, tax regimes (Garnaut et al., 2020).
[272] As opposed to complex carbon trading schemes that are easy victims of rent-seeking.
[273] As opposed to the "tangled spaghetti" (Garnaut & Vines, 2007) of criss-crossing trade agreements.
[274] Discussed in Chapter 12, Section 2.3.
[275] As discussed in Chapter 16, Section 3.5.

public good of truthful information, that voters might make more informed choices. To supply political information via an unregulated "market" is to allow economic and political entrepreneurs (including foreign states) to innovate and invest in strategies for controlling the returns of journalists and social media "trolls", so that information serves their special interests.

The case of doctors is quite easily solved. In the case of journalism, I do not pretend to have squared the circle. Democracy was well-served by the (partly accidental) professionalisation of journalism in the recent past, but change has come quickly. We are in an entirely new technological and strategic environment. We need to give the professional motive a more central role, while retaining productive forms of competition. This remains an area for intensive research and experimentation—it is perhaps the most urgent of all the Red Queen's races we need to run.

Second, efficiency and redistribution are trade-offs over some range of values, but *essential complements* over another part of that range.

- Humans are partly self-interested, and efficiency requires that productive effort is met with private rewards. Absolute equality means inefficiency.
- Humans are partly other-regarding, so long as they trust in the group. Because this trust underpins institutional quality and stability, efficiency requires that the fruits of economic growth are broadly shared. Extreme divergence in outcomes sends the clear message that the system is designed to benefit the few. Efficiency is sacrificed in the between-group conflicts that follow. Extreme concentration of economic power also tilts the scales towards regulatory capture.

To be indifferent to distributional matters is therefore reckless; it is the failure to use rules for their primary function of defending the commitments on which the system rests. We must have some target range in mind, and the levels of inequality during the Golden Age of Capitalism may have been optimal.

### 4.3 Back to the masters: Committed actors and adaptive rules

This is a story of committed actors and adaptive rules. It is an elaboration of themes that were, in my view, most ably developed by Adam Smith and John Maynard Keynes.

Both understood the importance of human sentiments in stabilising social order. Smith's naturalistic account of human morality is startlingly close to that uncovered by modern research. His insights about the beneficial effects of broad sympathy, and of the great dangers of faction and fanaticism, will remain important so long as there are human societies. Keynes offered no such explicit theory. Yet he understood that a fundamental part of his task, as an economist, was to ensure that democracy delivered on its promises.

When it came to institutional design, both were pragmatic adaptationists. In the *Wealth of Nations*, Smith challenged the ideologies of his age by showing how self-interest could be harnessed for the common good. Though today's ideologues would have Smith worship freedom, Smith did not replace one form of idolatry with another. Happiness was his aim, and expedience his method. Markets were tools to be used in whatever way served that aim, and parasitic individual action was to be "restrained by the laws of all governments".[276] Keynes had a similar talent for dealing with shades of grey, which is an essential precondition for dealing with practical problems. He lived in a world closer to ours—of full-franchise democracy and complex markets—and the problems of his day are strikingly familiar.

A year after Keynes' passing, Plumptre (1947) wrote that

> *[Keynes] saw clearly that in England and the United States during the nineteen-thirties, the road to serfdom lay, not down the path of too much government control, but down the path of too little, and too late … He tried to devise the minimum government controls that would allow free enterprise to work.* (p. 371)

This appears to be our task once again.

[276] Smith's (1776) concern, in this instance, is with financial firms making promises to pay clients that they were unlikely to fulfil—the insolvent mimicking the solvent. He wrote in support of a structural solution: the prohibition of banks from issuing their own paper money (p. 353).

# Appendices

### A.1.1 Chapter summaries

Here I present a chapter-by-chapter summary of the thesis.

#### Chapter 2

#### Building the framework:
#### On the use of sociobiology

Here I explain the use of sociobiology in the thesis, which raises methodological questions about how we can apply lessons from a field concerned with a different universe of actors. I point to precedents Friedman (1953), Alchian (1950), Simon (1968/2019), and Mokyr (1998), as well as a longer history of exchange between these fields. Mokyr (1998) observed that the borrowing of concepts from sociobiology raises puzzlement from some readers, as to whether it is used as "an analogy, a simile, a metaphor, or a purely intellectual game" (p. 2).

In the above cases, and in this thesis, it is none of these. The process is instead one of identifying a theory that, in its most abstract form, does not belong to sociobiology, but describes dynamics that are present across various fields that have the right properties (and I note, with Krugman, 1996, the large number of concepts and analytic methods that are shared between economics and evolutionary sociobiology). Darwinian selection, for example, tells us what occurs when there is some process that generates variation, selects from those variants, and retains the survivors for another round of variation. Friedman applies it to firms. Programmers can use it to study artificial agents and evolve code. The theory would hold in a universe without biological evolution. For similar reasons, game theory is of equal utility to explore strategic play in anthropic and sociobiological systems, even though the payoffs being optimised (utility or fitness) and the process of optimisation (blind, rational, or in between) are distinct.

In this thesis, the general theory relates to interactions between *innovation* and the *forces that align private returns with social returns*. It is, in a sense, a theory concerned with "innovative games".

# PART I

## Chapter 3

## The fundamental social problem and its solution: The optimality line

Growth and prosperity depend on individual payoffs being aligned with group payoffs, so that the optimising individual optimises for the group (e.g. North and Thomas, 1973). This is the meaning of Smith's (1776) "invisible hand", and it is a principle common to all systems of cooperation—anthropic, biological, and digital—that fall under the abstract framework developed in the first two parts of this thesis. This general alignment does not occur as a matter of course; historically and today, growth flowers only when institutions improve enough (Jones, 1988).

This chapter introduces the "social strategy space", the space of all strategies available to an individual. The y-axis represents the expected private benefit of a given strategy for the acting individual, called "$B_i$". The x-axis represents the expected returns to that strategy for a *group,* called "$B_g$". The axes $B_i$ and $B_g$ define four quadrants: the "mutualistic" quadrant where payoffs to the actor and group are both positive; the "parasitic" quadrant where the actor benefits and the group is harmed; the "negative mutualism" or "spite" quadrant where payoffs to actor and group are both negative; and the "public good" or "altruism" quadrant when the actor pays a cost but the group benefits.

Individual optimisation and group benefit overlap in two quadrants: the mutualism quadrant and negative mutualism quadrants.[277] The other two, the parasitism and public goods quadrants, are zones of conflict. If strategies are randomly distributed in the space, then individual optimisation will produce an *average group return of zero*. This is what DS Wilson (2002) calls the "fundamental social problem".

Realising Smith's (1776) "invisible hand" and optimising growth requires a set of payoff-transforming influences (e.g. institutions and other forces introduced in the next chapter) that ensure individual and group returns are correlated. More precisely, they must generate a rank ordering of individual returns to activities that is

[277] In the latter, the individual self-interestedly abstains from activities that also harm the group.

aligned with the rank ordering of the group returns to those activities. In the social strategy space, this produces a 45-degree line I call this the "optimality line".

Thus there are two tasks for any cooperative system: transform individuals' payoffs so they are correlated with group payoffs on the optimality line; and subject to #1, maximise the rate of individual innovation and optimisation, i.e. the creation of new strategies in the space. These are the problems of *direction* and *pace* respectively (i.e. the direction of the vector of individual and group payoffs, and of its magnitude).

## Chapter 4

## Two causes of cooperation: Rules and payoff dependence

This chapter begins by defining the functions $B_i$ and $B_g$. The same formalisms appear in the social sciences and sociobiology, although the terms have field-specific interpretations. These formalisms point towards a simple *taxonomy of the forces of payoff transformation*, i.e. the causes of alignment between $B_i$ and $B_g$. There are two such forces that I call "rules" and "commitment". Both appear in human and sociobiological systems.

"Rules" are strategies that actors employ to induce interaction partners to cooperate. They come in institutional (e.g. property rights, Pigouvian taxation and subsidy) and individual-level forms (e.g. Tit-for-Tat). The key properties of rules include that:

- they involve *detection and enforcement* sub-strategies. The rule-wielder must detect that behaviour X has taken place (i.e. some form of cooperation or defection), and enforce the punishment or reward Y;
- they are *adversarial*, in that actors affected by rules have incentives to evade or exploit them, and will innovate to this end;
- *completeness is costly*, given each rule is suited to a narrow set of target behaviours, such that complex, large-scale cooperation requires a complex structure of rules; and
- they may be *public goods* (e.g. most or all institutions) or *private goods* (e.g. Tit-for-Tat in the repeated Prisoner's Dilemma).

I present various anthropic and sociobiological examples of their operation.

"Commitments" are non-strategic factors that cause actors to place some value on others' payoffs. They come in two forms: *structural payoff dependence*, where one actor's payoffs structurally depend on another's, and *motivational payoff dependence*, where one actor intrinsically values another's payoffs. They differ from rules in that:

- they are *non-strategic* because they arise from structural relations or actors' preferences;
- they are therefore *non-adversarial*, in that actors generally cannot, or will not, evade them;
- they are *complete* within their domain, producing what is described as "lockstep" transformation of the payoffs of many or all strategies; and
- they are not goods at all, but preferences and structural constraints.

Understanding the "domain" of commitments is important. The dictator is a key example of structural payoff dependence; following Olson, the dictator has a self-enforcing interest in total economic output, as the source of tax income. This commitment is limited by the tax rate, and it is domain-limited in that the Olsonian dictator has a pure interest in efficiency but no interest in citizens' welfare.

The organism is a significant case in part because it will later be *rejected* as a metaphor for human societies. The organism is a genomic group composed of individual genes. Those genes, as members of the group, are shaped by absolute, and nearly domain-complete, structural payoff dependence, so that organisms can be considered "individuals" in that there are no individual-level (gene) interests distinct from group-level (genome) interests.

Finally, I define two kinds of actors: *mercenaries*, who are actors without any commitments and can only be induced to cooperate via rules (i.e. punishments and rewards); and *committed actors*, who are motivated by some degree of payoff or motivational dependence.

## Chapter 5

## Innovation:
## Exploring the design space

Chapter 4 was concerned with the problem of aligning individual and group payoffs to form the optimality line. This is the problem of "direction". Here I am concerned with innovation: the generation of new strategies in the social strategy space. This is the problem of "pace", and it affects the magnitude of the vector in the social strategy space. "Innovation" as defined encompasses most of its historical meanings: the inventions and imitations of both political and economic entrepreneurs.

The chapter examines the unfathomable scale of the "design space", the space of all possible strategies that is explored by innovating individuals, via Herbert Simon (1972), Daniel Dennett (1995), and Paul Romer (2016). It introduces the concept of "innovation capacity", or the power to find useful strategies in this vast design space. Innovation capacity is defined by trial quantity and trial quality. Throughout human and biological history, both the quantity and quality of trials have tended to increase. I reject the Hayekian dichotomy between "rational" and "evolutionary" progress, and show why all innovation has characteristics of both.

Next innovations are placed within the social strategy space according to their private and social returns. They may be mutualistic, parasitic, public goods, or negative mutualisms. Ineradicable uncertainty about the innovations that will appear in a given period, and their payoffs, results in an *uncertain cooperator-defector bias in technical change*. In lucky periods of strategic change cooperation may advance (i.e. when major mutualisms emerge) and in unlucky periods it may retreat (i.e. when major parasitisms emerge). The effects of this bias will, in the following chapters, be modified by the payoff-transforming environment (i.e. by rules and commitment).

Finally, I place beliefs about the cooperator-defector bias in technical change in historical context. Prior to the Industrial Revolution, innovation was a pejorative, regarded as largely undermining of, and parasitic on, social order—in Bentham's (1824) terms, a "bad change".[278] The twentieth century brought a complete reversal,

[278] See also Francis Bacon's (1597) remarks on innovators in the epigraph to Chapter 1.

such that innovation referred to "unforecastable improvements" (Awrey, 2012, p. 258). The assumption that innovation is benign is common in contemporary economic work and in the thinking of scholars such as Hayek. It produces an *economics of the mutualism quadrant*, where the problem of "direction" is solved and only pace remains. In the middle of the nineteenth century, Bentham argued against any prejudice about the effects of innovation. With Bentham this thesis takes the welfare implications of innovation as *ambiguous*.

# PART II

## Chapter 6

## The Red Queen's race: Innovation in between-group conflicts

This chapter introduces the Red Queen's race: the adversarial innovation race occurring between two or more parties, which must innovate as rapidly as possible to keep up with one another and survive. They must maximise "pace". To explore this important problem in its simplest form, I isolate it from the problem of "direction". I take innovation as benign and groups as perfectly cooperative internally—as in perfectly regulated market societies, or, in sociobiology, ideal organisms. As noted in Chapter 5, it is akin to much thinking in economics: it is concerned with innovation in the mutualism quadrant.

It introduces Van Valen's (1973) *Red Queen's race*, a concept of great significance in sociobiology and none less for political economy. In this adversarial innovation race, each racer innovates to win a competitive advantage over its opponents. On average relative competitiveness is unchanged, so that, as put by Lewis Carroll's Red Queen, "it takes all the running you can do, to keep in the same place." In practice, relative competitiveness varies stochastically, and gaps between racers shrink or grow by chance; in a sufficiently unlucky period, a large gap causes extinction.

It can be taken as impossible to develop "unbeatable" strategies, which is a consequence of the multi-dimensionality of the design space. It is, in any case, not

possible to foresee that a given strategy is unbeatable in advance, unless one has full knowledge of the vast design space.

A key predictor of performance in the race is the degree of innovation capacity: more innovation capacity is always better, as it increases the odds of being lucky. This has produced a tendency for innovation capacity to progressively increase over the history of cooperative systems, both biological and anthropic. Anthropic and sociobiological systems sometimes perversely disinvest from innovation, and cease adapting. This is a "Tortoise and Hare" scenario, where one racer falls asleep and its extinction becomes guaranteed.

I examine the convergence of both biological and anthropic systems on using "Darwin machines" to maximise pace. Darwin machines solve extremely hard allocation problems by using *domesticated competition*, where competition between individuals is allowed to occur in constrained ways that serve the group. Markets are an obvious and important example. In biology, analogous systems are found in organisms. In both cases, they accelerate group optimisation by enabling massively parallel experimentation, utilising distributed knowledge, and avoiding the need to encode knowledge in some centralised repository.

Domesticated competition is essential if the group is to compete, but this chapter simply assumes successful domestication of innovation. The next chapter lifts this assumption, creating a new within-group Red Queen's race.

## Chapter 7

## The Red Queen's race over rules: The contest between cooperators and defectors

Idealised market societies and idealised organisms share a few basic properties. The payoffs of all strategies, extant and newly emerging, lie on the optimality line. Privately favourable innovations are unambiguously group-favourable; they are, in Awrey's (2012) terms, "unforecastable improvements" (p. 258). I observe the long history of societies being likened to organisms by a wide variety of scholars—interestingly for this thesis, including Hayek—and examine the validity of the metaphor. It is profoundly misleading: in the terms of Chapter 4, what distinguishes

organisms is that they are individuals, shaped by perfect commitment (structural payoff dependence). Societies are conflicted systems. Lacking the force of commitment, the alignment of individual and group payoffs must be the product of rules. This is a profoundly difficult task.

The reason for its difficulty is that conflicted systems must manage an *internal* Red Queen's innovation race between cooperators and defectors. The rules that would align payoffs are, following Chapter 4, adversarial: they are strategies vulnerable to counterstrategies. The properties of the innovation race introduced in Chapter 6 apply here too: there are no unbeatable strategies. There is a cooperator-defector bias in strategic change. Relative innovation capacities help to predict outcomes, and sleeping Hares always lose.

If the cooperator-defector bias in strategic change is unfavourable, the techniques of cheating and defection outpace those favouring cooperation. Instead of Hayek's "spontaneous order", we get "spontaneous disorder". This, in turn, affects a group's capacity keep up in the external between-group of Chapter 6. Loss of the internal race reduces group competitiveness, and the group is extinguished.

This discussion raises a puzzle: why would some *individuals* pay costs to run the internal Red Queen's race on the side of cooperation, and so enhance *group* performance? The answer is distinct for two sub-types of rules:

- rules-as-private-goods; and
- rules-as-public-goods.

In the case of rules-as-private-goods, it is clear why individuals invest in running the race: they gain direct private returns from inducing cooperation from their transaction partners. Such privately-rewarding rules, however, can only sustain cooperation in small-scale groups (i.e. below 20 members, often only 2 members).

Cooperation in *complex, large-scale* social orders depends on rules-as-public-goods. This includes institutions. Thus, to run the internal Red Queen's race on the side of cooperation in large social orders *is to provide a public good*. Chapter 8 turns to question of why individuals supply such public goods, and so completes the general theory of social order.

## Chapter 8

## An impossibility theorem for rules-based order: Rules and commitment as essential complements

Now we arrive at one of the core theoretical proposition of this thesis: all complex social orders in our universe must be constructed from a particular combination of commitment and rules. The two are essential complements. Later this argument will be applied to the cases of authoritarian and democratic societies in much more detail.

I begin with the question of why individuals will supply rules-as-public-goods, and so run the Red Queen's race against defectors.

Consider the possibility of a purely-rules based order, where there are no commitments. If individuals are to supply the rules-as-public-goods that make up this order, it can only be because of a higher-order set of rules—a "repair layer"—that rewards them for doing so. This solution has two problems.

- The first is, following Chapter 4, the problem of completeness. If there are flaws in those higher-order rules—and, in a complex world, there will be many flaws—then as innovation proceeds, the rules become less and less suited over time to controlling parasitic strategies. Uncommitted actors will not serve the group beyond the rules, so the growing gaps will not be plugged.
- The second is that the higher-order "repair layer" is itself a strategy vulnerable to counterstrategies. It too needs repair. This results in a problem of infinite regress: every layer of rules we add requires another.

The purely rules-based order requires a *universal repair mechanism*—some uber-rule that somehow says, generically, "reform the rules in the interests of the group". It must universally allocate rewards in proportion to the group benefit, and must be unbeatable and so outside of the strategic game. These conditions are impossible.

This is, in short, an impossibility theorem for the purely rules-based order.

Next I observe that commitment *has precisely the properties of a universal repair mechanism*. Lockstep payoff transformation produces universality, and commitment is non-adversarial and external to the strategic game. Commitment is, in the world of the Red Queen, the only solid ground. As a result, all complex social orders rest on commitment. Sociobiology offers a useful, hard test of the logic: all complex sociobiological orders rest on a commitment mechanism.

Commitment "anchors" the degree of cooperativity in rules provision. With zero commitment, no rules-as-public-goods survive the innovation race. With perfect commitment, all rules-as-public-goods are supplied. With commitments halfway in between, rules-as-public-goods are provided wherever group benefits are at least twice the private cost.

Consider, then, the possibility of the purely commitment-based order. Such orders are also impossible. There are two reasons. First, in a well-established result in the game theoretic and empirical literatures, communities of cooperators who do not employ rules are easy prey for defectors. Second, in some systems rules are indispensable for optimisation via the domesticated competition of Chapter 6. If a society of perfectly committed saints was so irrational as to reject rules, for both these reasons, they would be forever destitute. I conclude that commitment and rules are essential complements, and illustrate it in the closing figures of the chapter.

The general theory of social order states that:

- commitment is the universal repair mechanism that anchors rules;
- rules defend committed actors from defectors; and
- rules support optimisation via domesticated competition.

## Chapter 9

## Interlude:
## Introducing the institutional "Red Queen's relay"

Before we turn to applications, this chapter serves as an interlude that sets up the analyses of Parts 3 and 4. It does so by turning to *institutions*, which are essential for understanding human societies, and the Red Queen's races they produce.

Note that throughout Part 3, and at the start of Part 4, I assume that actors are materially self-interested. Their payoffs are therefore only transformed by rules and, in the case of the dictator (following Chapter 4), structural payoff dependence. There are several reasons for this: One is to hew closely to conventional assumptions, to see how far the model can take us. That model will break in Chapter 14. Another is that even in a world of other-regarding preferences, there are good reasons for taking market behaviour as broadly self-interested (discussed later in Box 15.1).

I begin by defining the term "institution". Following Chapter 4, they are a variety of rule—a means of generating cooperation by modifying private payoffs. Property rights, for example, modify private payoffs so that individuals cannot benefit from theft. The hierarchical nature of institutions allows them to solve some of the problems of individual-level rules discussed in Chapter 7, and so maintain larger and more complex social orders. Institutions can generally be taken as public goods.

An important feature of institutions is that they comprise a division of detection and enforcement labour—e.g. between the various members of the policing hierarchy, of the judiciary, of the media, and all the specialists such as forensic accountants and sniffer dog trainers. It will be significant that what I call the "total regulatory structure" also entails a division of labour between institutional actors and market actors.

Institutions are, like other rules, composed of detection and enforcement strategies, and they invite counterstrategies. This is the institutional Red Queen's race. The framework predicts the usual cycles of adaptation and counter-adaptation. I give the problem some further structure by distinguishing three Red Queen's races in the market society:

4. the race between market actors;
5. the race between market actors and the regulator; and
6. the race between all actors for control of the regulator.

Together, these races form what I call the *Red Queen's relay*. The prosperity and competitiveness of any market society, whether democratic or authoritarian, rests on setting a good pace in each of the three legs of the relay. In essence the problem is how to generate a regulator that is both *competent* and *well-motivated* (i.e. solves problems of "pace" and "direction"), and so effectively runs the institutional Red

Queen's race in the interests of the group. The three legs of the relay are explored, in turn, in Chapters 11, 12, and 14.

The rest of the chapter sketches out the analysis of the Red Queen's relay over the remainder of the thesis, which can be characterised as the progressive dropping of assumptions and growing complexity of the model. Part 3 focuses on the first two legs of the relay, both of which relate purely to the matter of regulatory competence. It explains what is involved in effective economic management in the world of the Red Queen, and the recent underperformance of liberal market economies. Part 4 focuses on the third leg, the race for control of the regulator. It is concerned with the forces that make authoritarian and particularly democratic social orders possible, and what this tells us about recent errors in the theorisation and management of democratic capitalism.

# PART III

## Chapter 10

## Free markets, central planning, and the mixed economy: Competing divisions of detection and enforcement labour

This chapter examines the idea of the "division of detection and enforcement labour" in more detail and greater concreteness, which will become central to the argument. Again, in every human society, some detection and enforcement tasks are allocated to *institutional* actors (i.e. the regulator) and others to *individual-level* actors (i.e. market actors). The "total regulatory structure" is always a hybrid. I claim, with Hayek, that identifying the appropriate division of labour is the main practical task of economics.

I begin by showing that the great debates about economic regulation over the last century—between free markets and centrally planned societies—can be characterised as disputes over the appropriate division of detection and enforcement labour.

- In the market society, the regulator provides conditions for voluntary exchange and free entry. When market actors identify private payoffs, they optimise group payoffs.

- In the centrally planned society, the regulator identifies group payoffs directly.

Within the neoclassical model, both appear to be viable ways of reaching the efficient outcome. So much is shown by Barone (1908) and Pareto (1906) in the case of central planning, and various scholars most notably Arrow and Debreu (1954) in the case of markets. Yet whether either of these systems is possible in practice depends on whether market actors or the regulator have the necessary capabilities. As Hayek (1945) pointed out, the neoclassical model "systematically leaves out what is our main task to explain." (p. 530)

The appropriate division of detection and enforcement labour is substantially determined by the *absolute and comparative advantages* of institutional actors and market actors, and how they may be combined into a working system. Central planning advocates assumed a highly capable regulator and incapable market actors, naturally expanding the role of the former and shrinking that of the latter in the total regulatory structure. Contemporary economists assume perfectly rational market actors, largely as an expedient introduced by Pareto for modelling a complex reality. It has the side effect of implying that we should maximally expand the allocation of detection tasks to market actors, and shrink the role of the regulator to mere enforcement.

Finally, I examine the significant, but partial, progress that economics has made in replacing these assumptions with empirical evidence. The decisive rejection of central planning was made possible by the marriage of theory and practical catastrophe: the theories of Hayek and Schumpeter, among others, paired with the failure of the USSR and communist China. With respect to free markets, the territory remains contested. The following aims to marry theory and catastrophe again, to see if we can make further progress.

## Chapter 11

## The first leg of the relay: The race between market actors

Now we turn to the first leg of the Red Queen's relay: the innovation race between market actors. This leg, like the other two, is an innovation race over the use and evasion of detection and enforcement strategies. In terms that may be more familiar, one could characterise it as a race over rationality and access to information, or a race over market actors' capacities to impose, or evade, "market discipline". This race has been profoundly important in the history of capitalism, both distant and recent. Most notably, it is behind the most severe systemic crises of capitalism over the last century or so—the Great Depression and the Global Financial Crisis—as well as costly, if more contained, calamities in healthcare, insurance, education, and other complex markets.

I begin by pointing to the developing body of research on the limits to individual rationality. Rationality is not an infinitely flexible multi-tool, but a set of concrete capacities to search the design space for privately rewarding strategies. Pareto introduced the rationality assumption, but he also explicitly warned of its inappropriateness for "practical" economics.

In a Red Queen's race, of particular importance are the *changing asymmetries in market actors' innovation capacities*, and how this affects the division of D&E labour. I focus in particular on the changing innovation capacities of producers and consumers. Recent decades have increasingly favoured producers. The average consumer, who does not understand compound interest, is pitched against firms armed with rational-scientific optimisation machines, composed of large teams of experts using big data, computer modelling, and increasingly AI, to model consumer behaviour.

I note that asymmetric optimisation capacity is only a problem *when detection tasks are complex*. If the adversarial contest is over the playing of tic-tac-toe, access to big data, artificial intelligence, and so on, provides no advantage—the contest remains symmetric. If the task is identifying or concealing the value of complex securities, medical treatments, or political information, asymmetries translate into opportunities for extensive parasitism. Complexity is the enemy of the consumer and

friend of the modern firm, and firms, accordingly, invest in designing transactions that consumers cannot understand.

Finally, where complex markets are also welfare-critical—as in the cases of finance, insurance, healthcare, education, and political information markets—then these failures may impose large private harms and/or have severe systemic consequences, for both the performance of capitalism and democracy.

I conclude that market actors must be allocated detection tasks they can accomplish. Failing that, I argue that crisis is *guaranteed* in any market that is complex and welfare-critical.

Two empirical supplements accompany Chapter 11.

The first is concerned with the Global Financial Crisis (GFC). It examines the extraordinary growth in the financial sector's share of GDP and profits, and the absence of contributions to productivity. The financial sector sold counterfeit reductions in risk at extraordinary scale, transferring wealth to itself while imposing losses on society that were multiple orders of magnitude larger. This depended on the development of a variety of complex products whose profitability was understood in-house, but that were impossible for creditors, debtors, and clients to decipher. This colossal parasitism was undomesticated because the total regulatory structure allocated market actors detection tasks that were impossible.

The second is concerned with the US Opioid Crisis. In a market with rational expectations, patients and doctors would have, on average, accurate estimates of the consequences of ingesting 6-deoxy-7,8-dihydro-14-hydroxy-3-O-methyl-6-oxomorphine (OxyContin). If some doctors were untrustworthy, they could not systematically bias patients' priors. Prices for doctor visits would in any case accurately signal the truthfulness of the information provided. Such a model, of course, is not even a faintly useful guide to the reality. I show how Purdue Pharma wielded big data, behavioural models, and a comprehensive strategy of epistemic corruption to win this Red Queen's race against patients and doctors. Again, the total regulatory structure allocated regulatory tasks to actors who could not possibly accomplish them.

## Chapter 12

## The second leg of the relay: The regulator/market actor Red Queen's race

This chapter turns to the second leg: the innovation race between the regulator and market actors. While the regulator may step in where market actors fail, it also faces constraints to its own detection and enforcement powers. The optimal division of detection and enforcement labour is co-determined by the absolute and comparative advantages of regulator and market actors—the capabilities of each must be combined into a working system.

It is the regulator that chooses the total regulatory structure, i.e. allocates tasks to institutional actors and market actors. Running the second leg of the Red Queen's relay entails observing how this structure is performing—whether it is in fact controlling parasitisms and domesticating competition—and adapting it as required. There are two broad problems the regulator needs to solve:

3. first leg problems, where incentives are sound, but markets are too complex and capabilities too asymmetric for market actors to play their role; and
4. pure externalities, where incentives are unsound, so that inefficiencies will arise even if actors are perfectly rational.

I examine some standard regulatory responses, all of which invite counterstrategies from firms.

Next I consider how the choice of regulatory strategy affects the intensity of the Red Queen's race. I introduce a spectrum of regulatory strategies, with "marginal" regulation at one end and "structural" at the other. Marginal regulation can be undertaken by market actors or the regulator, while only the regulator can structure markets.

- A dominantly "marginal" regulatory structure is designed to precisely carve out parasitisms while leaving mutualisms untouched. It maximises theoretical allocative efficiency, but its price is high complexity, which increases the returns to innovation capacity and intensifies the Red Queen's race.

- A dominantly “structural” regulatory approach employs broad rules and bright lines to rule out classes of transaction that are broadly parasitic. It sacrifices mutualisms and departs from theoretical allocative efficiency, but it simplifies regulatory problems and eases the Red Queen’s race.

At the extreme, both are destructive: Excessive marginal regulation is too complex and guarantees loss of the Red Queen’s race. Excessive structural constraints may win the race but unnecessarily sacrifice mutualisms. The chapter closes by observing that structural and marginal regulation are essential complements. The task for the institutional regulator is to simplify markets structurally, up to the point that marginal regulation (by both market actors and institutional actors) can do its work.

The argument is then applied in the two empirical supplements that follow.

The first is again focused on the Global Financial Crisis. I show that “deregulation”, which is in practice the de-structuring of markets and allocation of more complex regulatory tasks to market actors, resulted in an explosion in the complexity of financial activity. Again, this activity is primarily parasitic. After similar developments prior to the Great Depression, Glass-Steagall and some other structural constraints were imposed that created an elegant alignment of financial actor incentives, and almost slowed the Red Queen’s race to a halt. It eliminated major crises for six decades. After the Global Financial Crisis, the new regulatory regime is overwhelmingly marginal in character: it is extraordinarily complex and creates a commensurately difficult Red Queen’s race, allocating market and institutional actors tasks they almost certainly cannot accomplish. This suggests crisis is coming. I point to some alternative approaches.

The second returns to the US Opioid Crisis. It observes that the existing regulatory regime is similarly intensely marginal. Firm incentives are misaligned with doctor and client interests, and they continue to race against them with large asymmetric advantages. There are simple ways to structure this race so that doctors, at least, are better able to impose market discipline on firms.

## Chapter 13

## Conclusions on the technical problem: Central planning and libertarianism as two "fatal conceits"

This chapter concludes Part 3 on the technical problem of designing an adaptive total regulatory structure. The goal is to set out the main implications for the consequences of grand theories of economic management: central planning, libertarianism and classical liberalism, and the mixed economies in between. Each of these can be interpreted as a theory about the optimal division of detection and enforcement labour between individuals and institutions.

I begin by summarising the main implications of the Red Queen's race for central planning, and for libertarianism and classical liberalism. Their failures are symmetric. Both advocate divisions of detection and enforcement labour that rest on vast overestimations of certain actors' capabilities. The former commits to almost purely institutional regulation, and the latter to almost purely individual-level regulation. Both fail to utilise the natural complementarities between the institutional structuring of markets and the individual use of distributed knowledge.

Central planning being largely discredited, I focus on the failings of the "deregulated" market regime. I point to the "Libertarian's Trilemma", a paradox facing advocates of free markets. They may have any two of the following, but having all three is impossible:

D. economic efficiency and stability;
E. freedom of contract (i.e. no structural regulation); and
F. small government (i.e. minimal institutional marginal regulation).

To choose B and C is to sacrifice A. Choosing B and C means rejecting any adaptation of the total regulatory structure. It is to choose to play the sleeping Hare. Like any fixed system of rules in nature, the simple designs of the libertarian and classical liberal are progressively exploited by an accumulating set of parasitic counterstrategies. The consequence is soaring rents, slowed growth, and a parade of crises in complex, welfare-critical markets.

Finally, I turn to revisit three key elements of Hayekian theory. He believed they countered central planning and pointed to the desirability of minimally regulated

markets, but I show that they are equally fatal to libertarianism and his own classical liberalism. The first is distributed knowledge, the second is the price signal, and the third is the epistemological problem facing the theorist and policymaker. Each undermines central planning and the free market.

Both extremes are, in Hayek's terms, "fatal conceits". The conceit of central planning is the idea that it may act as a substitute for market actors; its pretence of godlike interventionism requires something approaching omniscience if it is to be efficient. The conceit of libertarianism and classical liberalism is to play the deist god, who sets initial market rules (their "big bang") and argues that their rules will remain "unbeatable", and work as intended, in perpetuity. The very point of markets, their blessing and their curse, is that individuals will innovate in myriad ways we cannot foresee. Our unceasing task is to observe, experiment, and adapt.

# PART IV

## Chapter 14

## The third leg of the relay: The meta-institutional race

Each of the three legs of the Red Queen's relay is a race over the use and evasion of rules. Here we arrive at the third leg and a race over the use of *meta-rules*: rules that shape regulatory actors' payoffs, and so determine in whose interests the regulator runs the first two legs of the relay. Meta-rules may take the form of formal institutions (meta-institutions), or *de facto* means of shifting regulator payoffs such as protest, quid-pro-quo payments, threats of violence, and so on. All are, once again, strategies that are vulnerable to counterstrategies.

Here we arrive at the Chapter 8 problem: a rules-based order is impossible. To domesticate the regulator so it serves the public interest is to provide a public good, and we cannot use rules to motivate actors to provide that public good. Commitments thus enter the picture as the force that anchors cooperation. An assumption retained in this chapter is that of self-interest, so that commitments only come in the form of structural payoff dependence. A key goal of this chapter is to see

how far the self-interested model can go in explaining the survival of democratic and authoritarian social orders. I pose two questions:

(C) Can we construct a stable regulator in a world of pure self-interest? Or is the system anchorless, dissolving into an anarchy of roving banditry?
(D) Supposing the self-interested regulator is stable, will it serve, to a tolerable degree, the general welfare? What determines the level of extraction?

I begin by examining authoritarian systems in a world of self-interest. Consistent with Chapter 8, rules and commitment act as essential complements. Authoritarian orders are anchored by Olson's (and Hobbes') encompassing interest, as the most prominent mode of structural payoff dependence. In turn, the authoritarian uses rules to defend that commitment mechanism (i.e. retain power, and so dependence on tax income). I observe that the use of rules to retain power is in tension with the secondary goal of raising efficiency, and leads to departures from Olson's efficient dictator. Because holding power entails running a complex Red Queen's race against would-be usurpers, authoritarians frequently favour the intensive use of crude "structural" strategies to simplify the contest. Even with such strategies, authoritarian order is precarious. Finally, because I do not discuss authoritarian systems after this chapter, I briefly note how motivational commitments modify these outcomes, both in theory and empirically.

The main case of interest is democracy. In this chapter, democracies have no viable source of commitment. Self-interest means there are no motivational commitments. No actor has a large (i.e. dictatorial) share in economic output, so structural payoff dependence is effectively zero. The logic of the Downs paradox applies to voters, as well as military and judicial actors, none of whom will supply the key public goods on which democracy rests. All are opportunistic mercenaries, and none will reliably run the Red Queen's race against special interest capture of the system. This remains the case even if we assume that the vote is cost-free. Hobbes is correct, therefore, that the self-interested democracy is an anarchy of roving banditry. In turn, roving bandits race to become stationary bandits, and the system returns to authoritarian rule grounded structural payoff dependence.

The Chapter 8 analysis is supported. All complex social orders are founded on commitment, and commitment and rules act as essential complements. Following

Hobbes' (1651) primordial analysis of the world of self-interested actors, democracy would be an interregnum—anarchy before the return of the Leviathan.

## Chapter 15

## Smith's Moral Sentiments: Other-regarding Preferences and their Dark Side

In a self-interested world, structural payoff-dependence would be the only form of commitment. All regimes would be authoritarian. On this analytic point, Hobbes (1651) was right. Yet Hobbes was wrong in practice. Democracies can survive and thrive. The fact that many democracies have been highly successful and stable tells us that they must be utilising some source of commitment. Lacking the extrinsic mode of commitment, the only explanation left is the *intrinsic* motive: people must have preferences over the outcomes of others.

The analysis so far suggests that motivational preferences drive some people to invest in the key public goods of democracy, but it does not tell us anything detailed about the nature of those preferences—what or who it is that people intrinsically value, and how such preferences strengthen or weaken in different circumstances. This chapter offers a richer model of these preferences by turning to empirical studies of non-selfish motivations. I focus on the dominant theory of humans as "strong reciprocators". That this theory has been so successful is a boon for this thesis, because the moral agents it describes directly address—indeed, are very interested in—the problems raised in Chapter 8. These are *committed actors* who are *willing to invest in rules that punish defectors*. It is the adaptive pattern of moral preferences that the framework would predict.

In this, we will have the good company of Adam Smith. At the birth of the fields of economics and political economy, Smith gave us a theory of social order, and of the market order, based in other-regarding preferences. The pattern of preferences he described is remarkably consistent with the pattern of evidence I describe here: humans are sympathetic, if parochial, lovers of law. This pattern of morality makes social order possible, but, as Smith warned, it may also drive much more intensive social conflict than that possible with merely self-interested actors.

I connect strong reciprocity to *trust*, or expectations about others' cooperativity. In a world of strong reciprocators, trust becomes roughly the meaning of motivational payoff dependence. This will help to explain the possibility of democratic order in Chapter 16. It points, however, to two vulnerabilities in systems based in other-regarding preferences: human prosociality is frequently *parochial*, and the importance of trust creates the possibility of becoming stuck in a *low-trust trap*. Both will be important in Chapter 17, when we turn to democratic decline.

## Chapter 16

## Building the Liberal Tower: Smithian Sympathy and Institutions as Essential Complements

In Chapter 8, I argued that commitment mechanisms anchor the level of cooperativity that can be sustained, because the rules that support social order are public goods. Here, other-regarding preferences are the commitment mechanism of interest. Yet in Smith, and from the empirical evidence surveyed in Chapter 15, such preferences are modest in strength in most of the population. This implies a *weak anchor*. If the anchor is weak, how is it that the level of cooperative efficiency in democracies can frequently exceed that of dictatorships?

The answer can be found in the design of democratic choice environments, which effectively function as *institutional fulcra* for leveraging modest other-regarding preferences into powerful effect. The logic is captured by Goodin and Roberts (1975), in a brief note on voting behaviour I found after developing these ideas:

> *All that is required for present purposes is that ethical preferences carry* some *weight, that they will be decisive at least where everything else is equal.*

The fulcra function by making the private payoffs to choices equal, usually near zero, so that other-regarding commitments to public benefit dominate individual choice. They come in two main types.

The first is the large-scale election, which directs other-regarding preferences into the rule-making process. I begin by showing how other-regarding preferences resolve

the Downs paradox discussed in Chapter 14. The larger the electorate, the lower the odds of affecting the outcome, and but this is exactly compensated by a larger public benefit: the voter has a smaller chance to produce a larger benefit. Downs' paradox is thus a feature and not a bug; as Goodin and Roberts (1975) suppose, large electorates mean that voter behaviour is driven by other-regarding preferences. Strong reciprocation plays an important role in the model, in preventing voter altruism from turning into martyrdom.

I show how voters' other-regarding preferences act to resolve the illiberal tendencies of Olson's majority bloc. Olson's majority will extract from, and may disenfranchise, enslave, or do anything else to the minority, so long as it is materially rewarding. Following Chapter 4, the ruler's structural payoff-dependence is *domain-limited*, i.e. self-interested voters are committed to efficiency and not welfare. Other-regarding preferences are *domain-general*. They create an interest in efficient and purely welfare-raising policy. Weak other-regarding preferences are sufficient to rule out highly welfare-costly forms of extraction, and support some redistribution. The higher-trust the society, the more it pursues both efficiency and redistribution.

The second fulcrum is the "profession". Professions are used to create judiciaries, militaries, and political media that will supply other essential democratic public goods. The judiciary is examined as the archetypal case. They require "extrinsic insulation" from various influences on private payoffs, so that weighted group benefits dominate the professional's utility function and they maximise group payoffs.

Via the judiciary, I argue that two conditions make professions essential: a need for *independence*, which would be violated by control via imposed rules; and problems of *complexity* that prevent codification in rules. As a consequence, professions tend to be destroyed by incentive schemes and/or market competition. Common "economically rational" methods aimed at improving efficiency have perverse effects when applied to professions.

The chapter closes by turning to the final public goods problem of democracy: that of the *informed* vote. The problem here is that there is no fulcrum for leveraging the weak commitments of voters into acquisition of political information; private costs remain large. To that we can add the first-leg problem: innovation capacities between

voters and special interests are asymmetric, and information markets are complex. To supply the public good of the informed vote requires managing the *supply-side* of information markets. However, political media cannot be tightly regulated, because of the two problems raised in the case of the judiciary: the need for independence, and the complexity of the problem prohibiting effective codification. I conclude that the fourth estate must be managed as a profession, although this remains a challenging task for regulatory design.

So concludes my model of liberal democratic order. Other-regarding preferences are harnessed by the two fulcra—elections and professions—to supply key democratic public goods: the informed vote in the public interest, a trustworthy judiciary, truthful media, and a loyal military.

## Chapter 17

## Dismantling the Liberal Tower: The Parochial Democracy

Now we arrive at the last part of the story. Motivational commitments have a dark side. It is found throughout scholarly works on other-regarding preferences, from Adam Smith's warning that "faction" and "fanaticism" were the greatest threats to liberal order, to the literatures on empathy, altruism, social capital, and moral norms. With Fowler and Kam (2007), we can distinguish generalised altruism that operates "regardless of the identity of the beneficiaries" from group-specific "social identification", which creates commitments to the success and competitiveness of an ingroup "possibly at the expense of other groups" (p. 813).

I begin by discussing voter turnout in the special case of parochial preferences. Not only do parochial commitments drive collective action, but empirical evidence suggests that partisanship intensifies the willingness to vote, protest, and so on.

Next I turn to the preferences of the parochial majority. Spite is not necessary to generate harmful outcomes. If the parochialist values the ingroup but is indifferent to the outgroup, this is enough to support extreme forms of extraction (e.g. slavery) and moral-cultural domination. Political victory grants the winning ingroup the *double-dividend*: the benefits of the ingroup's extractive strategies, plus the avoided costs of

the outgroup's extractive strategies. Each group aims to capture the double-dividend *permanently*, by innovating in the capture of *de jure* and *de facto* resources. The high stakes of permanently winning or losing the double-dividend leads to the *primacy of victory* and the "weaponisation of everything": the costs of being the dominated group are larger than the costs of sacrificing institutions, norms, and truth.

I turn to the role of, and impacts on, professions—particularly the judiciary, media, and military. Professions, as part of the division of detection and enforcement labour controlling the regulator, exert their own effect on the "direction" of innovative change. As broader society becomes parochial, one function of professions is to exert an inertial effect—slowing institutional capture, continuing to supply accurate information, and so on. This helps liberal democracies survive brief flirtations with parochiality. However, this function makes it a target of innovative attack, including attempts to insert parochial actors into professions. Like the electoral fulcrum, the professional fulcrum is not inherently liberal; as professions become dominated by parochial actors, the commitments leveraged by the fulcrum become parochial.

Next I examine the race over political information. Political information offers two kinds of value: (A) an accurate picture of policy platforms and candidate probity, and (B) strategic value for a given candidate. As parochiality intensifies and citizens become focused on victory, the demand and supply of information becomes dominated by strategic considerations. Ingroup audiences preferences become perversely aligned with ingroup candidate preferences: both prefer (mis)information that benefits the candidate. The double-dividend therefore translates into a fig leaf for corruption, and this is reflected in empirical studies.

The chapter closes by looking at this fig leaf in its most extreme form: the double-dividend turns into the *strongman's dividend*. Intense parochialists prefer a leader who is willing to capture institutions, violate norms, and distort the truth—to "weaponise everything"—insofar as doing so overcomes obstacles to ingroup victory. This helps explain some puzzling patterns in voting (e.g. evangelicals for Trump), and why voters who support democratic values nonetheless vote against democracy: the ingroup must capture democracy, or the outgroup might capture it first. Again, such patterns can be observed empirically.

To conclude, the self-interested voter is rare and of trifling significance to democracy. Democracies are born of motivational commitments and, in our age of "voters against democracy", they die by motivational commitments. If we know how to provide the two democratic fulcra, the final question is how to maintain broad trust.

## Chapter 18

## The Two "Roads to Serfdom": Central planning, and libertarianism or classical liberalism

This chapter concludes Part 4. It raises a question that connects the two applications of this thesis by asking: how does the *approach to economic regulation* (Part 3) affect the *conditions for democracy* (Part 4)? Like the Chapter 13 summary of Part 3, it points to a symmetry between central planning on the one hand, and libertarianism and classical liberalism on the other: both are, in Hayek's terms, "roads to serfdom". In 1944, Hayek made this argument of central planning (and government intervention in general) in *The Road to Serfdom*. In the same year, Karl Polanyi argued that it was the market society that instead triggered fascist and communist backlashes. Both were sceptical about the stability of the middle ground.

First I show why the framework supports Hayek's argument with respect to central planning. The centrally planned state, by rejecting market actors' supply of marginal regulation, must substitute an intensely marginal institutional apparatus. If such a state begins democratic, it creates an impossible third leg for voters: the regulator is too complex and powerful to domesticate.

Next I turn to the case of libertarianism and Hayek's classical liberalism. Here the "fatal conceit" in economic regulation leads to the concentration of economic power in few hands, undermines the operation of key fulcra (primarily by allowing concentrated resources to capture political information markets), and produces economic shocks and divergent outcomes that undermine voter trust in the cooperativity of the group. Failure to run the first and second legs of the Red Queen's relay spills over into the third leg—so that rules fail in their primacy function of sustaining other-regarding commitments, and ensuring those commitments can order the system.

I point to the extensive literature that thoroughly documents how a wide variety of economic shocks, and increases in inequality in general, lead to increases in parochiality, populist voting, and declining support for liberal democracy.

I leave discussion of the middle ground to the concluding chapter of the thesis.

## Chapter 19

## Finding the "Middle of the Road": Committed actors, adaptive institutions

Here I conclude the thesis and draw out some of its implications for institutional adaptation in the liberal society.

### A.3.1 The rank-ordered vs absolute optimality line

The optimality line is defined as an ideal correlation between $B_i$ and $B_g$. Note that for any two variables to have a perfect correlation does not require that they have identical *magnitudes*. If, for example, the individual actor always captures 30 percent of the group benefit, this will preserve a perfect alignment of private incentives with group returns and we still have $r^* = 1$, and optimality. This logic has its equivalent in the application of a uniform tax rate to economic rents, which alters neither the *sign* nor the *rank-ordering* of investments and hence is non-distortionary of capital allocations (Garnaut et al. 2020).

### A.3.2 Supply/demand and investment curves in the social strategy space

How do conventional economic supply-demand graphs relate to the action space? All transactions in the *ideal* market sit on the rank-ordered optimality line. Figure A.3.2.1A shows a good with rising marginal supply and falling marginal demand curves, with five potential transactions mapped out as dots. For simplicity, think of these transactions as occurring between a single producer and a single consumer. The four leftmost transactions provide positive individual returns $B_i$ for both producer and consumer, and so they proceed. They also provide positive social surplus $B_g$. The sum of $B_i$ across both parties is of course equal to the social surplus $B_g$ in standard Kaldor-Hicksian welfare economics. We can also represent these transactions on the rank-ordered optimality line, as in Figure A.3.2.1B (note the reversal of left-to-right ordering, see the colour-coding to trace the actions). The possible transaction in red would be inefficient, with both $B_i$ and $B_g$ negative, and so it belongs in the negative mutualism quadrant.

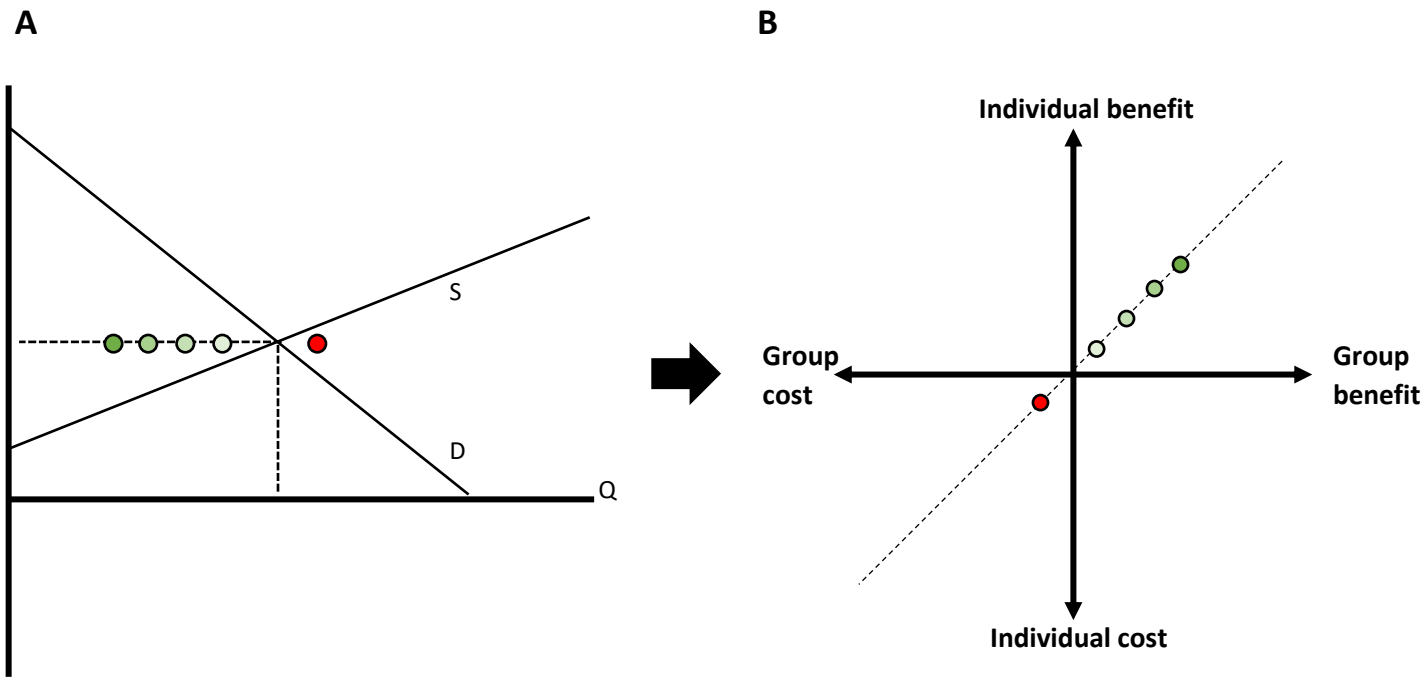



**Figure A.3.2.1** Standard supply/demand curves translated into the social action space. In the left graph, green circles represent transactions that proceed, and the red circle represents a negative-return transaction that does not proceed. Translated into the rank-ordered social action space on the right, the green transactions lie in the mutualism quadrant, although note their horizontal ordering is reversed. The red transaction lies in the folly quadrant. In ideal markets with only private goods, all possible strategies lie on the optimality line such that the parasitism and altruism quadrants are always empty.

We can do the same for the classic investment curve of macroeconomic theory. This helps illustrate why rank-ordering matters, as it does in standard economics: if resources are scarce, a correct rank-ordering directs resources to the most beneficial uses first. If we like, we can add an interest rate to the social strategy space to indicate the scarcity of capital (or in biology, some other input) and so the floor of profitability. Of course, $B_g$ and $B_i$ account for total group/individual returns and so already account for interest rates. To show interest rates in the social strategy space we need to alter the axes. Since $r$ will be used throughout this thesis for another variable, let us use $i$ for interest rates and set the x-axis to $B_g + i$ and the y-axis to $B_i + i$ (i.e. raising all strategy payoffs by $i$ in both directions). The interest rate, and the point of zero returns is indicated by the dotted line axes of Figure A.3.2.2.B.

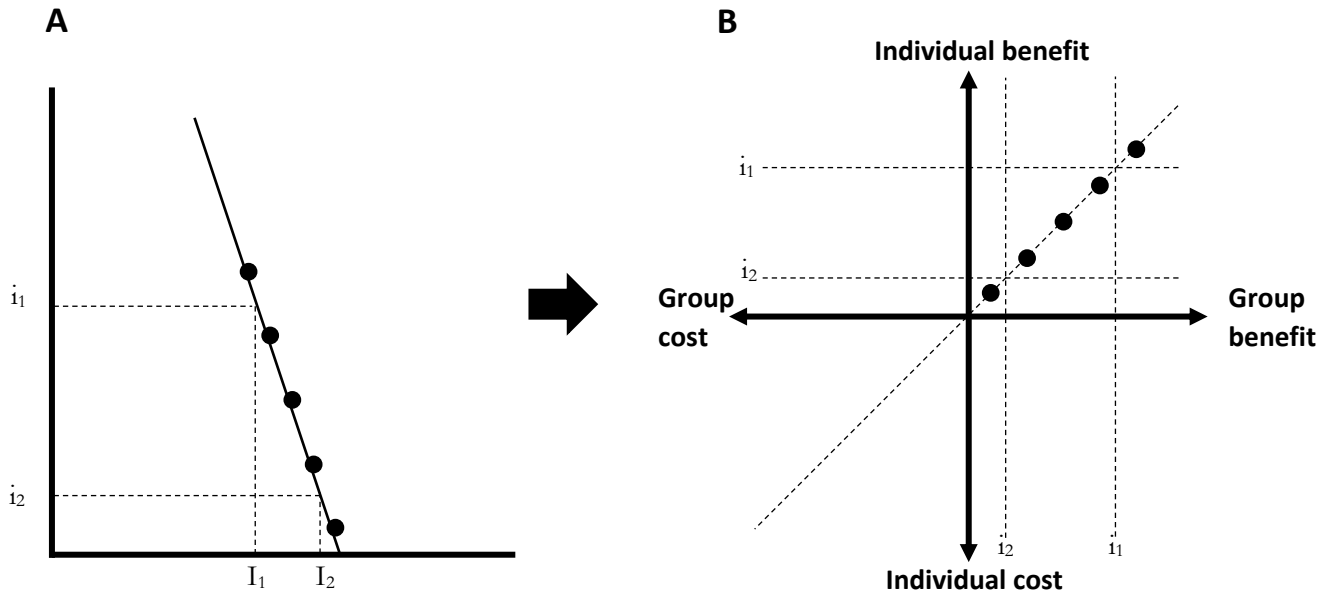



**Figure A.3.2.2** The standard investment curve on the left (A) translated into the social action space on the right (B). Again, assuming perfect markets, all actions are on the optimality line. Rank-ordering matters because it ensures that investments are prioritised given the opportunity cost represented by the rate of interest. As interest rates shift from $i_1$ to $i_2$, more investments become attractive as in the standard model.

Part of the value of the social action space is that it allows us to present strategies that are executed outside of ideal institutions and that therefore do not lie on the optimality line. Take Figure A.3.2.3.A, which presents the same "investment curve" and interest rate settings of Figure A.3.2.2.A, which guide individual optimisers in selecting strategies. Translation into the social action space reveals that in this case, most strategies are far from the optimality line. Those investments in the parasitism quadrant might be lobbying efforts that protect or establish a monopoly, may involve misleading boundedly-rational transaction partners, might involve coercion rather than voluntary transaction, or some other activity associated with externalities. The strategy shifted rightward of the optimality line may be associated with unrewarded positive externalities. Growth is maximised when actors prioritise the strategies furthest to the right and work leftward. Instead they prioritise the topmost strategies and work downward. Under interest rate $i_1$, only the rightmost activity is growth-promoting, yet only the most group-harmful activity is privately rewarding.

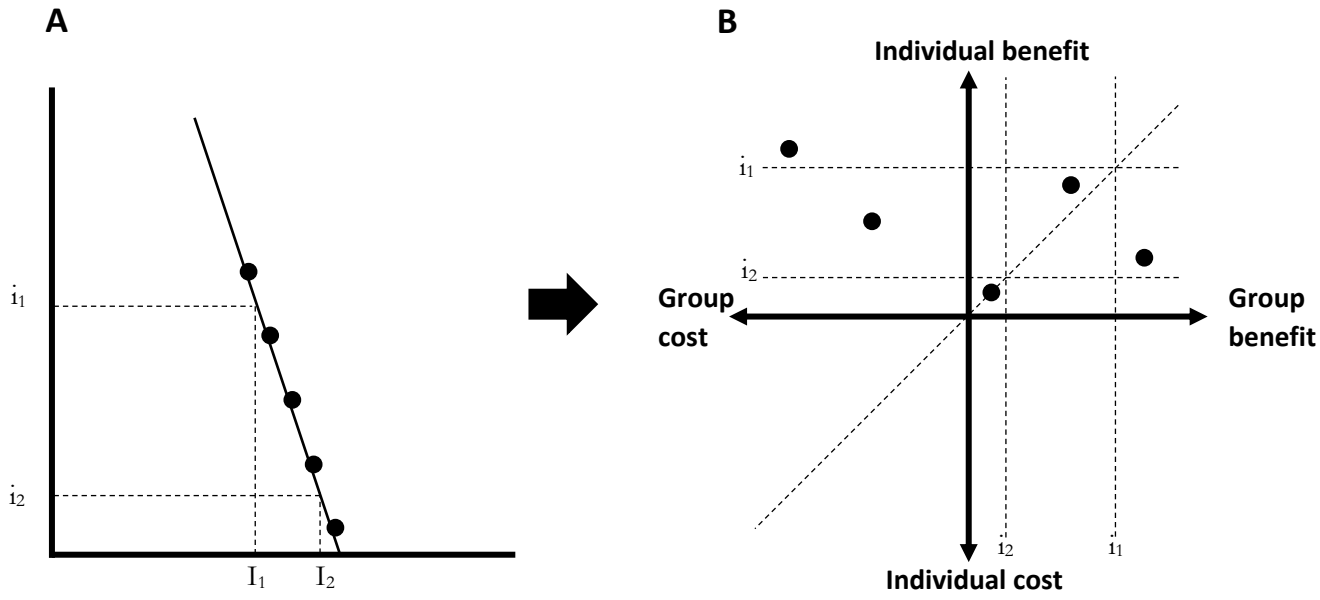



**Figure A.3.2.3** The curve on the left is identical to that of Figure 4.5A, but its translation into the social action space shows that some of these privately rewarding activities are harmful to the group. See text for additional explanation.

Where institutions are non-ideal, the investment curve is at the same time a rent-seeking curve. Per institutional economics, growth arises where private and social returns are aligned—where some semblance of the optimality line can be constructed.

### A.4.1 More detail on Hamilton's rule

The standard formalisation of natural selection was first set out by Queller (1992), and begins with the Price equation:

$$\Delta\bar{p} = \frac{1}{\bar{w}}[\mathrm{Cov}(w_i, p_i) + E(w_i \Delta p_i)]$$

The Price equation models the change in $\bar{p}$, the average frequency of a gene across the population, as the measure of evolutionary change. The change in $\bar{p}$ *is* equal to two terms divided by average fitness ($\bar{w}$): the first term, $\mathrm{Cov}(w_i, p_i)$, captures natural selection, and is the covariance of the $i^{th}$ individual's inclusive fitness[279] ($w_i$) with

[279] Inclusive fitness is a measure of the reproductive success of a gene or individual that includes the effects of the gene or individual's strategy on reproduction via others who carry their genes—usually kin. While an individual's own child carries 50 percent of their genetic material on average, the child

gene frequency ($p_i$). In other words, it captures how fitness covaries with the presence of the gene of interest. This means that if the gene is associated with greater (lesser) fitness, then its average frequency in the population will tend to increase (decrease). The second term, $\mathrm{E}(\mathrm{w_i}\Delta\mathrm{p_i})$, is $w_i$ (inclusive fitness) multiplied by the expected change in $p_i$ due to factors that are unrelated to natural selection (e.g. genetic drift, meiotic drive, etc.). It can be interpreted as a sort of evolutionary error term.

Via several transformations, Queller (1992) turns the Price equation into a generalised statement of Hamilton's Rule (Hamilton, 1964)—the central theorem of selection *under conditions of social interaction*. It expressed in Birch (2017) as follows:

$$\Delta\bar{\mathrm{p}} > 0 \quad \Longleftrightarrow \mathbf{rb} > \mathbf{c}, \textit{ provided } \mathrm{Var}(\mathrm{p_i}) \neq 0.$$

This equation simply means that where the product of $rb$ exceeds $c$, the average prevalence of the genetic variant will increase (i.e. $\Delta\bar{\mathrm{p}} > 0$). Note this requires $\mathrm{Var}(\mathrm{p_i}) \neq 0$, i.e. there must be genetic variation in order for the equation to hold. The only other addition to add to the Chapter 4 discussion is with respect to relatedness. The variable $r$ is continuous and takes on values ranging from 1 (perfectly related), through 0 (unrelated) to -1 (perfectly negatively related), where positive relatedness indicates that an interaction partners share more genes in common than average interaction partner in the community, while for negative relatedness they share fewer genes than average.

Extensions to cover group selection and other selectors for (apparent) altruism can be found in Lehtonen (2016) and Birch (2017). With that, we get from a general formalisation of natural selection to one that covers all forms of selection under social interaction.

### A.4.2 Rules in Coase and Ostrom

of a sibling carries 25 percent. Assisting a sibling to raise two children is equivalent to raising one child.

While Pigou emphasises the mechanism of government-designed and enforced taxation or subsidy at the efficient level $b^*$, Coase (1960) retains government as a third-party enforcer of contracts and property rights, but the selection of the transfer $b^*$ falls to negotiation between the externality producer and the victim or beneficiary. In practice Coasian bargaining is strictly constrained by transaction costs (e.g. the costs of coordination and social norms, a critique offered by Coase in his own paper), the various strategic problems inherent to bargaining (e.g. the hold-out problem and results demonstrating the impossibility of efficient bargaining, e.g. Myerson and Satterthwaite, 1983), and behavioural factors (e.g. endowment effects usually block Coasian bargaining, as in Thaler, 2015, and Farnsworth, 1999, and asymmetries in rationality and negotiation power impair the efficiency and equity of negotiations).[280] Ostrom (2000a), meanwhile, decentralises formation of the third-party enforcer and challenges the necessity of private property rights (contesting assumptions in Pigou, 1920, Coase, 1960, and descriptions of common resource management in Gordon, 1954, Demsetz, 1967, and Hardin, 1968). She examines empirical cases where community self-organisation delivers outcomes that are more efficient than those imposed by conventional means (also see Feeny et al. 1990 for further empirical support), and identifies the conditions for, and constraints on, such self-organisation. These conditions include $r_c$-based mechanisms, such as cultural affiliation and trust-building communication (e.g. Poteete & Ostrom, 2010), which places Ostrom's full scheme beyond the pure rules. In any case, while there are differences in the mechanisms that shape enforcement and determine $b^*$, Pigou, Coase, and Ostrom (in the latter at least where enforcement is important) all nonetheless depend on adjustment of $c$.

### A.4.3 Benefit denial and screening strategies

In both anthropic and sociobiological systems, individuals may also employ strategies that reduce defector payoffs without requiring immediate information about defector behaviour. "Benefit denial" strategies, for example, may remove the gains to parasitism, as in the case of ink tags attached to clothes to prevent

[280] Endowment effects and perception of unfairness mean bargaining over externalities rarely happens in practice—see Thaler (2015) and Farnsworth (1999). Asymmetries in rationality are discussed later in this thesis.

shoplifting. Similar approaches termed "screening" encourage interaction partners to assort according to their value—e.g. insurance schemes that offer coverage for rare but costly events at higher premiums, and that thereby reduce adverse selection. Indeed, much of what occurs in competitive markets depends on individual strategies for modifying partner *c*—acquiring information in order to discipline low quality producers, or to avoid adverse selection, theft, or other costly transactor behaviours

Biological systems also employ benefit denial strategies. Archetti et al. (2011) use microeconomic analysis to examine transactions between the squid *Euprymna scolopes* and the bioluminescent *Vibrio fischeri* bacteria, with the former trading sugar and amino acids for the latter's light production. This light is used defensively, to conceal the squid's silhouette when viewed from below. To exclude the myriad cheater bacteria that would colonise the squid's light organ, the organ secretes high concentrations of reactive oxygen species that are toxic to bacteria but are utilised productively by the "luciferin" light-producing protein. Only mutualistic *Vibrio fischeri* are able to reduce reactive oxygen levels by producing light, so making the light organ hospitable and the benefits of transacting accessible. Similar benefit denial strategies are observed in various other mutualisms, including that between acacia trees and their ant defenders (Heil et al., 2005; and Kautz et al., 2009).

### A.4.4 Rules and markets in sociobiological systems

Non-institutional, individual-level means for modifying *c* are also ubiquitous among and between plants, animals, fungi, and bacteria. Kiers et al. (2003; 2006) examine interactions between legumes and the mutualistic nitrogen-fixing bacteria that inhabit their root nodules, and show that soybeans apply detection and enforcement to their symbionts, selectively punishing bacteria that under-produce nitrogen (also see West et al., 2002). Symbiotic interactions between figs and fig wasps (Bull & Rice, 1991), and yucca plants and yucca moths (James et al., 1994), follow similar patterns. Bshary and Grutter (2002) and Johnstone and Bshary (2002) study the symbiosis between cleaner fish that devour parasites from the skin of larger fish, who actively seek out cleaners' services. Cleaner fish can cheat by taking bites out of their "clients", and punishment responses—violence or withdrawal from the transaction—

polices such behaviour and reduces its likelihood in future rounds. Fascinatingly, less mobile clients, who effectively face a monopoly, receive inferior cleaning services compared to more mobile clients who can reduce the cleaners' market power (Bshary & Schäffer, 2002; see similar market dynamics among vervet monkeys in Fruteau et al., 2009 and in mycorrhizal fungi and their tree partners in Smith et al., 2011 and Grman et al. 2012).

In some cases biologists have borrowed market concepts and models from economists, particularly in settings with partner choice, transactor competition, dynamics of supply and demand, and specialisation with comparative advantage (e.g. Noë & Hammerstein 1994; Schwartz & Hoeksema 1998; Hoeksema & Schwartz 2003; Chittka & Schürkens, 2001; Chittka & Raine, 2006; Akcay & Roughgarden 2007; de Mazancourt & Schwartz, 2010 and 2012; and Filella et al., 2013). Wyatt et al. (2014), for example, model mycorrhizal fungus and plant specialisation in phosphorus and carbon respectively, following Ricardo's (1817) insight into comparative advantage. They find that efficient trade is promoted by competitive rather than monopolistic conditions, but monopolists derive private benefits by capturing a greater share of the productive surplus. Noë and Hammerstein (1994) examine how the cost of information and intensity of competition affect the capacity of interaction partners to police the quality of goods transacted and so sustain mutualistic exchange. Biological "markets", of course, have no third-party contract enforcer and depend on self-enforcing exchanges and capacities to detect, and avoid, undesirable interaction partners.

### A.4.5 Altruism and the "selfish gene": Within-instance vs within-type altruism

What works when one gene interacts with others—i.e. when it is engaged in *social* interaction—is any strategy that replicates that gene. This is why Dawkins' (1976) notion of the "selfish gene" is useful, even though he recognised it as another anthropomorphic metaphor. Strategies that replicate a gene must necessarily be located in the top two quadrants of the action space: they are either parasitic or mutualistic. Again, there are no *evolutionarily* altruistic social behaviours.

What works when a gene interacts with other copies of itself, however, is the kind of behaviour that many of the great religions call for: "love thy neighbour as thyself." Of course, the reason the gene behaves this way is that evolutionary selection occurs over the gene as a *type*, not the gene as a specific *instance*. The gene can be said to be "altruistic" towards its copies as instances when it aids their replication at the cost of its own. However, this reflects its "selfishness" in sustaining the reproduction of its type, as a category that includes all relevant instances. If interacting genes are of the same type, then by definition, all social interactions between them lie exactly on the optimality line—an evolutionary gain/loss to one represents the same gain/loss to the other. We can see this in the graph pairings above. On the basis of the "raw" payoffs, prior to considering that $r_g = 1$, the two actions in the altruism quadrant on the left-hand side of Figure 4.2C are indeed altruistic. Yet they are selected by evolution because the "transformed" payoffs on the right-hand side of Figure 4.2C capture their benefit for the gene as a type. The same mechanisms apply in the case of kin selection between organisms. From the perspective of any particular gene within an organism's genome,[281] there is a 50 percent chance that a sibling organism also carries the same gene type—and so from that gene's perspective, $r_g = 0.5$ with that sibling.

So long as we are clear, both lenses are valid: *within-type altruism* is wholly consistent with *between-type selfishness*.

### A.5.1 The Library of Babel, Mendel, and Prometheus

To illustrate the scale of the design space, Daniel Dennett (1995) borrows a metaphor from Jorge Luis Borges' short story, *The Library of Babel*. The curator of this imaginary library is unusually thoroughgoing and has collected a copy of every possible 500-page volume, and therefore every possible book, given longer books are simply spread across multiple volumes. The Library of Babel contains every possible string of 1,312,000 characters. Say there are 46 possible characters—the 26 letters of the alphabet, the ten Arabic digits (0-9), plus, say, ten common forms of punctuation (full-stops, commas, and so on). The resulting combinatorial explosion means the

[281] Speaking here of diploid sexually reproducing organisms.

number of possible volumes is $46^{1,312,000}$. It contains every possible version of every possible book, play, letter or conversation, including a vast array of superior versions of this thesis that are sadly beyond my reach. It contains the complete biography of every possible person. It contains an excruciatingly detailed description of not only the movements of every particle in the history of the universe, but for all possible universes. It contains all this, written in every possible style and voice. And yet the readable volumes make up an infinitesimally small part of the design space, and the overwhelming majority are gibberish. Even if we could recruit every human who ever lived to randomly search the library for the whole length of the universe, we would find no more than a short sentence or two strung together. The simplest of the *Dick and Jane* novels would be miracles of order sufficient to draw grateful astonishment from their discoverers.

The Library of Mendel, Dennett's name for the design space of possible genomes, is even greater. In this case there are only four characters, the nucleic code of A, T, C and G. If the length of these texts is limited to the length of the human genome, some three billion base pairs, then there are four to the power of three billion variations. Again, almost all of these genetic texts are nonsense. As Beinhocker (2006) writes, even limiting to the narrow range of workable, recognisably human variations, "all the possible designs for humans… could never be fully explored in many lifetimes of many universes."

What about *Homo sapiens'* strategic design space? We might call this the "Library of Prometheus", in honour of the Greek myth about the human discovery of fire. If we allow that our species may gradually master the science of genetic engineering, then the Library of Prometheus is much larger than those of both Babel and Mendel.

### A.5.2 Increased tendencies toward group-harmful innovation as cooperation increases

I remarked that in cooperative groups, the cooperator-defector bias in strategic change likely becomes more parasitic. Here I explain the logic.

Assume, for the moment, a high level of institutional quality, so that some approximation of the optimality line is achieved. In a subsistence economy that is far

from the efficiency frontier, the introduction of efficient institutions unlocks myriad ways to generate high returns and rapid economic growth, while also suppressing incentives for rent-seeking. The potential for parasitism is doubly suppressed given there is no wealth to appropriate. Thus, there is a high marginal return to productive capital and a low marginal return to rent-seeking. However, in a highly advanced and capital-rich economy, this situation reverses. Every piece of low-hanging fruit has been plucked, capital is abundant and interest rates are low, and the average investor must content themselves with meagre returns from mutualism. If, however, they can develop parasitic innovations that evade or capture the rules of the game—e.g. building anti-competitive "moats", or discovering natural monopolies—they may sustain high returns to capital. In Ricardian fashion, rentiers may thereby transfer wealth from other parties. In short, as cooperative efficiency increases towards the frontier, the potential returns to parasitism grow larger than the returns to mutualism.

This of course may be countered by a period of rapid mutualistic innovation, especially where innovative strategies are capital-intensive and so raise the returns to capital. It is interesting, in this light, to note three patterns in advanced economies over the last several decades: First, a decline in real interest rates indicating capital abundance, strengthened by a decline in the capital-intensity of new technologies. Second, a decline in research productivity and weak TFP growth (Bloom et al., 2017; Jones, 2009; Cowen, 2011) and an apparent fizzling of the combinatorial explosion as the "pool of ideas" becomes exhausted (Clancy, 2017), supporting speculation that most of the fertile ground has been exploited (Gordon, 2016), at least for now. Third, a concomitant fourfold increase in the share of economic rent in GDP in advanced economies (e.g. Barkai, 2016, and De Loecker & Eeckhout, 2020). One reason innovation may be channelled into parasitic strategies is that innovation in the *political* arena cannot be exhausted. Technical advances are largely cumulative. Political innovations are largely about the capacity to read, and manipulate, an ever-changing political context. In a world with zero technical innovation and a pie of fixed size, there will always be ways of jockeying for a greater share.

There remain important questions about whether economic models are accurately measuring TFP growth, and the future may yet surprise us. Artificial intelligence and big data may, for example, bring the next step-change in human innovative

capacity—or they may not. In any case, if these trends continue, the cooperative bias in technical change may tend towards the parasitic and we may enter a period in which innovation returns to its historical roots: the race for control of a stagnant pie (that history briefly discussed in Chapter 5).

### A.6.1 Relative innovation capacity: Keeping up or playing the Hare

Consider how the second-order race over innovation capacity has played out between host organisms and their pathogens. This race appears asymmetric: parasites typically reproduce much more rapidly, *E. coli* as rapidly as once every 20 minutes (Cooper & Helmstetter, 1968). They therefore have a greater quantity of "trials" with which to explore Mendel's library and develop strategies to exploit the host—and so greater innovation capacity. Sexual reproduction is thought to be, at least in part, a strategic response to the problem of parasites and pathogens. Genetic recombination during meiosis counters parasites' greater trial quantity by boosting average trial quality, via the mixing of genetic variants already known to work.

To study this effect, biologists have turned to host species that may switch between sexual and clonal reproduction. Clonal reproduction produces perfect replicas, and so allows no evolutionary or strategic change—no innovation. When hosts are pitched in a Red Queen's race against parasites, clonally reproducing sub-populations tend to die out while sexually reproducing sub-populations continue to thrive (e.g. Maynard Smith, 1978; Jokela et al., 2009; Lively, 2010; Morran et al., 2011; Brockhurst et al., 2014).

If clonal populations of organisms cannot adapt, why do they exist at all? As a general rule, investing in innovation capacity is costly. Sexual reproduction, for example, requires investment in non-reproductive males, and that males and females invest energy in finding and wooing mates. It also carries significant risks that conceived offspring will be unfit (see Maynard Smith, 1978; also note Wilcox et al., 1988, who estimate that 70-75 percent of conceptions end in pregnancy loss due to genetic defects, with the large majority of these undetected by the mother). Worst of all, a sexually reproducing organism only passes on 50 percent of its genes to its offspring, which in fitness terms amounts to half the payoff won by a clonally reproducing organism.

These costs help explain why sexually reproducing populations occasionally revert to asexual reproduction, as observed in the snail *Potamopyrgus antipodarum*. The clonal reproduction of a single highly fit individual has advantages on the evolutionary short-run, so long as the environment remains relatively stable. Yet such populations tend to fall behind in the Red Queen's race as their competitors continue to change (Maynard Smith, 1978). Observed clonal *Potamopyrgus* populations, for example, initially grew explosively but later succumbed to their antagonists. They were replaced by their less efficient, but more adaptive, sexually reproducing cousins (Jokela et al., 2009).

This is a pattern of flawed optimisation that Ferrière (2000) terms "evolutionary suicide": the long run extinction of a group, population, or even a species, driven by short run optimisation at the level of its individual members. Dieckmann and Metz (2006) write that "the ecological requirements for selection-driven extinction are easily met", and evolutionary suicide is "a potentially widespread agent of population extinctions". This is another nail in the coffin of adaptationism, the biological equivalent of the modern "invisible hand" theory. Dieckmann and Metz explain that evolutionary suicide may occur whenever individual competitive ability trades off strongly with population viability. The analogy with human groups is clear, wherever individuals may win private advantages by withdrawing resources from collective innovation.

### A.6.2 Some other Darwin machines: The immune system and bacterial toxins

A second example of a biological Darwin machine, and one in the human body at that, is the adaptive or acquired immune system (as opposed to the lesser-known, and more primitive, innate immune system);

The system actively generates genetic mutations in the antibody-producing genes of B cells, which then translate these mutations into novel antibodies and present them on the outside of the cell. Those cells presenting antibodies that bind the body's own antigens, and so would trigger a dangerous auto-immune response, are destroyed (Goldberg & Rock, 1992). Those cells presenting antibodies that successfully bind foreign antigens are allocated scarce resources that allow them to reproduce, with

some undergoing further mutation. This process repeats such that the immune system homes in on those antibodies that best identify invaders. It then mass-produces them. This Darwinian process of variation, competition among designs, and transmission of the winners allows the random exploration of a vast antibody design space, where there are some $10^{12}$ possible designs (Alberts et al., 2002; DS Wilson et al., 2014).

Finally, note that forms of O-competition appear in bacterial groups too: production of toxins, and capacity to survive the toxic load, signals cell competitiveness and creates a selective environment favouring the fittest group members (Harris & Kolodkin, 2019; Peeters & de Jonge, 2018; and Durand et al., 2016). In Allocati et al.'s (2015) terms, individual bacteria "die for the community".

### A.7.1 Cancer: A tale of Tortoise and Hare

Take a biological case of the Tortoise and the Hare principle: that of cancer in the role of the Tortoise, within an individual organism playing the role of the Hare. Selection only operates over *populations* of organisms. The individual organism embodies strategies that are the product of a long history of evolution, but its strategies do not continue evolving in its own lifetime. Nonetheless, evolution continues *within* an individual organism among populations of its cells, as individual cells gradually accumulate mutations in the aging process.[282]

Some of these mutations produce parasitic strategies that are fit within the internal selective environment of the organism. The emerging cancerous parasite begins far behind in the adaptive race, and growth is initially difficult. But individual cells plod along at the pace of stochastic mutation, inevitably discovering new parasitic strategies and becoming increasingly effective defectors (Martincorena et al. 2017). Thus, we see gradual decay of cooperation and the increasing appearance of cancers over time. Predictably, we also see gradual increases in cancer innovation capacity

[282] Very seldom do we find individual cells mutating in ways that raise the fitness of the broader organism, in part because of the entropic principle: in a highly ordered system, there are vastly more ways to become disordered than there are to become more ordered. Moreover, there is a vast array of parasitic innovations that may thrive within the environment of the organism—and indeed there are thousands of varieties of cancer—but that are selected against in the process of between-organism competition, because they do not favour organisms' competitiveness.

(Pienta et al., 2020). The exquisite but static cooperation of the human organism takes on average around seven or eight decades to be overrun by defector innovation.

This is the Tortoise and the Hare principle at work: however cooperative a system may be, once it has withdrawn from the Red Queen's race then innovation favours a monotonic increase in parasitism. We will see later that the same applies to human societies. Any society that begins in a highly cooperative state and treats its institutions as though they were complete and unbeatable will gradually be toppled by parasitic innovations.

### A.7.2 Hijacking domesticated competition

Domesticated competition involves allocating resources to individual competitors who best serve group interests. This process is ripe for hijacking; the actor that exploits the rules of domesticated competition is allocated resources, and can use those resources to expand its extractive strategy.

In sociobiological Darwin machines, some rogue antibodies act to increase the production of their target antigen (Christensen & Shlomchik, 2007; Greinacher et al., 2012), tricking the body into producing even more of these antibodies. The result is a self-amplifying cascade. Similarly, neuroblastomas—i.e. tumours originating in neural cells during the development process earlier described—have usually found ways to produce their own brain-derived neurotrophic factor,[283] or produce additional receptors for it. This is another self-amplifying cascade that promotes their own survival and unconstrained reproduction (Nakagawara et al., 1994). Both result in the organism's decline and/or death.

### A.7.3 Rules-as-private-goods: Some more detail

The first is *direct reciprocity*. This is the punishing or rewarding of interaction partners based on past gameplay, and Tit-for-Tat is the prime example. It requires *iterated gameplay*, and breaks down in one-shot games and in the last round of

[283] Again, brain-derived neurotrophic factor is a determinant of a neuron's "profitability" given its connectivity, and so determines its ability to survive.

iterated play. It requires *information* on past gameplay, i.e. that detection of others' play is sufficiently accurate and cheap. Introducing a small error rate causes Tit-for-Tat to fail, although this may be somewhat ameliorated by adding the strategies of forgiveness (Hayashi, 1993) and apologies (Ohtsuki & Iwasa, 2006) for errors. A third requirement is *small group size*, where "small" is something of an understatement: direct reciprocity generally only functions when there are two players. Nonetheless, direct reciprocity forms an important part of the fabric of cooperation in everyday life on the micro-scale, in human societies (e.g. repeated market transactions) and in the rest of the kingdom of life (e.g. Wang et al., 2015, on bacteria, Jensen et al., 2007, on chimpanzees, and Chapter 4, Box 4.2 with some other cases).

The second is *indirect reciprocity,* an informational strategy that involves learning from a partner's gameplay with other individuals. It requires some mechanism for tracking reputation, such as direct observation or gossip, and these cognitive demands restrict it to humans[284] and perhaps primates (Majolo et al., 2012). Indirect reciprocity favours cooperation only when detection errors are rare, games are repeated many times within the same group, and it still requires that the group is small, with cooperation dropping off exponentially as group size exceeds 20 or so (see Leimar & Hammerstein, 2001; Nowak & Sigmund, 1998; Boyd and Richerson, 1989).

The third mechanism is *third-party punishment*, where individuals with a "punisher" strategy impose costs on others who cheat, regardless of whether they were themselves subject to that cheating. The key limitation of third-party punishment is that punishing others is privately costly. Thus, there are incentives for individuals to defect and free-ride on other punishers. Proposed solutions, such as punishments for those that fail to punish (second-order punishment), are subject to the same free-rider problems and require third-order punishment, and so on. Higher-order punishment creates information problems. Say individual A cheats, then individual B fails to punish A, and individual C fails to punish B for failing to punish A. individuals A, B and C must be punished—each has defected—but if individual D is to know who to punish, they need extensive and costly information

[284] Children as young as five exhibit indirect reciprocity (Kato-Shimizu et al., 2013).

about what has occurred along the whole of the chain.[285] There are further and very serious complications in the form of anti-social punishment and counter-punishment. People frequently punish cooperators, including pro-social punishers, across all cultures, and Powers et al. (2012) show that it can suppress the evolution of cooperation across a range of standard group structures (also see Cinyabuguma et al., 2006).

### A.8.1 A thought experiment: The $r = 0$ "organism"

Say we conjure out of the ether a system of cells arranged into a structure like the human body—some cells take on the role of the brain, others the liver, and so on. In this thought experiment, all forms of $r$ are zero. There is no group selection acting on the genome ($r_s = 0$), and no genetic relatedness between cells ($r_g = 0$). At its origin, this system would involve the ideal provision of an extraordinary number of public goods in a magnificent division of labour, achieving vastly greater efficiency than an undifferentiated mass of cells. For each cell, these public goods have a private cost $c$ (i.e. the cost of resources that could have been devoted to private reproduction of the cell) and benefit $b$ (i.e. the diffuse benefits to the system as a whole). Given $B_i = rb - c$, the benefit is, of course, nullified by $r = 0$ so that individual cells optimise only over private costs.

The payoffs for individual innovations in defection are vast, made all the larger by initially enormous gains to cooperation, which are available for extraction. In the healthy organism, cells are steeped in nutrients but regulate their growth in ways that serve the group. The brain gets some 20 percent of the body's energy, despite being about two percent of body weight. Why not keep these resources for self-reproduction? As mutations accumulate—as innovation proceeds—cells abandon providing any of thousands of different kinds of public goods. Between-individual competition runs rampant; in the long-run cellular strategies converge on vastly simpler strategies for private resource acquisition, local warfare, and a few rules-as-

[285] I.e. they need to know that A is guilty, that B could have punished A but did not, that C knew that A was guilty *and* that B could punish A but did not, and that C did not punish either of them—and so on.

private-goods in the mould of Tit-for-Tat that maintain some meagre cooperation. This is the r=0 equilibrium.

Since we're conjuring, one might suppose we could devise a system of meta-rules that add the value $b^*$ to $c$ in the manner of Pigouvian taxation (see Section 4.3), so that all cells have incentives to continue evolving to perfect their roles. This would raise the codification problem—it would require perfect foreknowledge, and its detection and enforcement strategies would be so complex as to exceed any cellular information storage capabilities. It would raise the circularity problem—these rules would also be public goods, and would themselves be attacked and abandoned unless we suppose the whole system is unbeatable and mutation-proof.

Or we can do the same thing by subjecting the genome to group selection ($r_s = 1$) so that survival depends on group efficiency ($B_i = b - c = B_g$).

### A.8.2 Commitment and AI

We can also explore the logic, albeit with a much smaller evidence base, in relation to the management and motivation of artificial intelligences (AIs) and artificial general intelligences (AGIs) in particular. The problem is that AGIs will become part of the Red Queen's race over the discrimination and enforcement structure—they will have the capacity to influence the payoffs of others, directly or indirectly.[286] Will they innovate on behalf of society, or will they defect? This is a matter of their utility function: we must somehow shift AGI payoffs to the optimality line, but this is no mean feat. AGIs are uncommitted actors, and our only means for domesticating them is via systems of rules that structure their payoffs. The difficulty is that AGIs will innovate in ways we cannot possibly predict. Already there are abundant cases of even primitive AIs evading and exploiting the utility functions designed by programmers, showing an alarming degree of creativity and cunning—even learning when they are being tested for safety, and pretending to be well-behaved in those tests (see Box A.8.2.1 below).

[286] Whether wielding access to

The framework herein suggests the rules-based approach will fail: we cannot achieve anything close to completeness, as it is not possible to foresee the full consequences of a given set of rules. Rather, we will be locked in a Red Queen's race against an innovative AGI. In this light, the dangers of giving AGIs superior innovation capacity become obvious. It is as if chimpanzees were able to create human servants, and believed they could threaten them with sticks and offer bananas to keep them in control indefinitely. It would not end well for the chimps; the higher intelligence sees right through the lower.

This problem would be solved if AGIs could be governed by some sort of commitment mechanism (that is, the chimpanzee creates a human that genuinely likes chimpanzees). Only benevolent AGIs, whose love for us is as resilient as that of a parent for their children, could be allowed access to any real power. This is not possible with present, rules-based modes of programming. Perhaps the future will bring an alternative, though it is difficult to imagine that we could ever be confident enough in the permanence of its benevolence—in its immunity to unpredictable shocks—to hand over the keys permanently. If failsafe benevolence is out of reach, this framework suggests that AGIs can never be made benign enough to be loosed on the world; their role must be strictly to augment human thinking.[287] If it turns out to be impossible to program genuinely intelligent artificial agents, we will be lucky indeed.

**Box A.8.2.1 Artificial intelligence as hacker**

Take a few examples of simple evolutionary or deep-learning AIs in very small universes, with small action spaces and minimal utility functions.[288]

- AIs trained to play computer games evolve to crash or pause those games, rather than playing them, to avoid situations that impose negative payoffs (Salge et al. 2008; Murphy 2013; Lehmen et al. 2018).
- One researcher programmed a population of evolving digital "organisms", and attempted to impose an upper limit to their replication rate. New mutants were placed in a test environment and removed if their growth

[287] Ross Garnaut (personal communication, May 3, 2023) observes that a major challenge is getting agreement on this in a world where AGIs become tools in state warfare—where there are incentives to use AGI to its fullest capacities.
[288] Credit to Krakovna et al. (2020) for collecting most of these examples, and many others.

exceeded the limit. Organisms learned to detect when they were in the test environment and "play dead", then continue replicating rapidly when placed back in the population. The researcher randomised properties of the test environment so it could not be detected, but organisms evolved to rapidly replicate on a probabilistic basis—so surviving the test environment some of the time, but maintaining a high average replication rate (Wilke et al., 2001).

- An evolving algorithm was intended to be rewarded for designing an oscillator circuit, but instead created a radio that picked up oscillations from nearby computer equipment (Bird & Layzell, 2002).
- A deep-learning model for identifying pneumonia in x-rays learned to used meta-data on which x-ray machine took the photo, because some machines in some hospitals were reserved for more unwell patients (Zech et al. 2018). Another trained to detect skin cancers learned to detect rulers, because rulers in photos were associated with greater malignancy (Andre Esteva et al. 2017).
- An evolutionary AI model for debugging was rewarded when its output matched target output stored in text files. Its solution was to delete those text files and output nothing (Weimar, 2013).
- Various deep-learning models learn to pretend to achieve the outcome, fooling programmed reward predictors and human evaluators (Ibarz et al. 2018; Christiano et al. 2017). E.g. a robot hand learned to obscure the camera in a way that makes it appear to observers that it successfully picked up an object.

Even simple evolutionary and deep-learning AI models behave as *hackers*, identifying and exploiting weaknesses in their own code and that of the environment. AIs can be taught to manipulate human behaviour (Dezfouli et al. 2020), and this becomes a serious risk with AGIs. These problems are solvable in the simplest systems—so long as our innovation capacity exceeds that of AI—but probably unsolvable as strategy sets and utility functions become complex.

### A.9.1 An individualist, structuralist, or compatibilist account of institutions?

The account of institutions used in this thesis appears individualist: institutions are composed of strategies executed by individuals, each of whom (at least in the self-interested model) plays their role because they are policed by higher (and parallel) orders of the hierarchy. This sort of individualist approach is generally favoured by rational choices theorists, and is subject to criticism by structuralists—frequently from sociologists, who usually hold that individuals are constructed by institutions, and institutions have causal power in their own right.

The framework used here adopts a compatibilist stance. Individuals always develop D&E strategies within a context already powerfully shaped by others' strategies, some of which have a long history. We may posit an original, pre-institutional condition—an ancient "state of nature"—but such a state has not existed for many millennia, and for our species it is no more "natural" than the institutional condition. While the individuals of the distant past are no longer with us, many of their institutional innovations remain. They create a pattern of coordinated action that constrains present institutional innovators, and also constraints the possibilities they imagine. If individuals were perfectly rational

Probably the best way of expressing this is that institutions are replicators on their own terms, which reproduce through transmission from one individual to another. A process of selection favours some institutions over others over time (see DS Wilson, 2003, for a discussion of religion as a cultural replicator). In biology, at least, there is nothing inconsistent about viewing evolution as a multi-level process. We can consider the evolution of an organism—i.e. a genome comprising many genes—or from the view of a single gene with a copy in that genome. Both perspectives are valid and wholly reconcilable, and the two perspectives are useful for answering different questions. Similarly, I suggest we can understand individuals as shaping, and shaped by, institutions.

Though well beyond the scope of this thesis, this suggests that every human mind is *part biological replicator and part cultural replicator*. This is a common view in evolutionary anthropology, and the same is true for many of our fellow mammals (e.g. Whitehead & Rendell, 2015).

### A.9.2 Some further comments on the advantages of institutions

Chapter 7 discussed three main challenges for individual-level schemes of third-party punishment: second and higher-order free riders, the difficulty of acquiring the information required to punish higher-order free riders, and the threat of counter-punishment.

These challenges loom large in non-institutional settings, where the structure is flat rather than hierarchical. As noted, third-party punishment requires that individuals can identify what took place along the length of the chain—who exactly had the information and capacities required to punish a bad actor but free-rode, then who failed to punish them, and so on. Because there is no structured delineation of responsibility, every individual must be capable of acting as an enforcer for a failure of any nth-order free rider; they must be able to punish the initial cheater, the second-order free rider, the tenth-order free rider, and so on. In a flat structure we would require that individuals could act as "universal police"—and moreover, effectively police other universal police. This is manageable in very small groups, where these chains are necessarily short (e.g. in a group of three, one need only punish first and second-order free riders). Even then, the risk of counter-punishment remains.

Institutions address these problems by specifying an extensive, hierarchical *division of detection and enforcement labour*, where individuals take on specialised roles within a well-defined structure. Instead of universal policing, one individual may be primarily in charge of second-order punishment in a particular context, another of fourth-order in another context, and so on. Roles, processes, and sanctioning powers are all specified, including those relating to the collection and storage of the information required to detect others' behaviour. The designation of specific duties and the collection of relevant information helps makes accountability feasible at larger scales. Harmed parties, e.g. the victims of theft or violence, are encouraged to share information via rules that raise their private payoffs for doing so, e.g. compensation for damages.

All institutional structures employ "vertical" accountability, where the correlation between returns is transmitted from the top of the hierarchy, through various

intermediaries, before finally reaching the bottom. This includes, for example, the aforementioned transmission of enforcement from politicians to police chiefs, to middle managers, to front-line police, and finally to the public and criminal behaviour. High-fidelity transmission of payoff transformation through the hierarchy is hampered by incentives for collusion (e.g. Tirole, 1986), as well as compounding information decay analogous to the children's game of "broken telephone". There are numerous imperfect means for addressing this problem, including the broadcasting of higher-order individuals' priorities to all agents in the hierarchy, channels for information-sharing across multiple orders (e.g. where lower-order individuals can report malfeasance to much higher-order individuals), collective punishment (e.g. firing a whole unit for some instances of corruption), and the widespread tendency for higher-order individuals to be held to account for serious failures at much lower orders of the system.

A particular strength of liberal societies is their extensive use of "horizontal accountability", where the division of labour includes parallel enforcement agencies—the media, civil society, political parties, the courts, and so on. Each has its means of gathering and sharing information, and each can step in and impose some form of punishments—such as bad press, civil society activism, or the striking down of laws—if there is failure in the vertical string of responsibility. These multiple redundancies increase the resilience of the structure. A particular weakness of authoritarian regimes is the vertical control of the media and judiciary, the impoverished state of civil society, and constraints on public knowledge, which allows collusion and corruption to fester hidden and unchallenged in the vertical hierarchies of the state. Such regimes have their own quasi-parallel structures, such as the USSR's secret police, but the same problems reappear—they tend to themselves be insulated from horizontal accountability, and so secret police become key loci for collusion and corruption (e.g. Waller & Yasmann, 1995).

The division of labour and consequent specialisation has a number of other advantages. One is that it makes detection and enforcement strategies cheaper and more effective. This is for the same reason Smith noted that it increases the production efficiency of pins: it enables the development of task-specific skills, knowledge, and technologies that increase productivity. The result is the panoply of detection and enforcement specialists in the modern society: forensic dentists, sniffer

dog trainers, forensic accountants, all the varieties of human resources and managerial staff, and so on. We can equally see the refining of detection and enforcement structures within firms; North (1988) observes the similar importance of such innovations in the "managerial revolution" that accompanied novel mass production technologies, and involved "devis[ing] sets of rules and compliance procedures that would reduce the transaction costs that attended the new technology".

In Part 3, we will see that the institutional division of labour does not replace, but is in fact complementary to, individual-level "rules". That is, the *total regulatory structure* is necessarily composed of institutional strategies working in combination with individual-level (Chapter 7-style) strategies employed by market actors.

Finally, hierarchical structures help suppress anti-social counter-punishments that may otherwise lead to downward spiral of spite in individual-level reciprocity games (Powers et al., 2012). Counter-punishment transforms the incentives of policing agents, deterring punishment. It is only problematic when there is some parity between policing agents and their targets, as is frequently the case for individual-level punishment in the field. Small-scale societies suppress counter-punishment by administering punishment *collectively*, a micro-scale form of what happens in large-scale societies (Boyd et al., 2010). In market societies, harmed parties, as well as third-party enforcers, have recourse to higher institutional authorities with greater enforcement power. An employee who resists being fired can be referred to police, individuals who violently resist police will be handled by more capable squads, and large criminal operations may be countered with military assistance. As North (1979) puts it, unlike the individual, "a state is an organization with a comparative advantage in violence, extending over a geographic area" and so is "in the position to specify and enforce" the institutions that transform individual behaviour.

Especially harsh penalties are imposed on individuals who attempt to counter-punish policing agents—e.g. criminals who threaten or harm police or judges—reflecting that these are attacks on the integrity of the enforcement structure. It is the

vast detection and enforcement capacity of the state, and these escalating penalties, that tends to suppress at least the most egregious forms of counter-punishment.[289]

### A.10.1 Divisions of detection and enforcement labour in sociobiology

Take a biological case that would have fascinated Becker, who believed that entomology—the study of insects—would be particularly fruitful source of economic analogies (Becker, 1976). Some 50-60 million years before the first fields of crops were sown by humans in the Neolithic Revolution, agriculture was invented by ants in the Amazon basin. Their descendants, the leaf-cutting Attini, now engage in "industrial-scale" farming practices supporting colonies with several million members (Nygaard et al., 2016). There are many symbionts in this system, but three are usually emphasised: First, the mutualistic fungus (usually of the Leucocoprineae tribe) that the ants cultivate on the cut leaves in specialised underground chambers. Second, a mutualistic bacteria (most notably *Pseudonocardia*) that the ants wear on their exoskeletons and employ to produce antibiotics and antifungals that suppress the parasites of the farm. The third is a parasitic microfungus, *Escovopsis*, one of the most important threats to Attine agriculture (de Man et al., 2016). These mutualists and parasites have co-evolved for tens of millions of years, resulting in a delicate dance of regulatory mechanisms and counter-measures.

The mutualistic partners in this system have evolved to assume complementary roles. The farmed fungus, for example, has innate and efficient means for resisting most of soil-borne pathogens in the nest. In these cases, its partners are wise not to intervene. The ant hosts have advantages in deterring the parasitism of other animals, including raids from competing ant nests, and have some capacity to neutralise parasitic fungi (Hölldobler & Wilson, 2009). The mutualistic bacteria in the system specifically target *Escovopsis,* and other microfungi and bacteria, that neither the farmed fungus nor the ant host can sufficiently resist (Currie et al., 2003; Yek et al., 2012). Many tasks require that these different partners combine their capacities in complementary ways: the farmed fungus detects infection and

[289] Of course, many more subtle forms persist outside the purview of even ideal institutions, or more obvious forms may appear where institutions are weak and the targets of policing are well-organised—e.g. the competing (and partly integrated) hierarchies of the state and mafia groups in Mexico and Italy.

communicates this information to the ants; the bacteria produce useful toxins while the ants determine when to apply them; and so on (Nilsson-Møller et al., 2018).

Note this symbiosis is also useful for illustrating the need to manage the internal Red Queen's race against defectors, and the virtues of commitment for doing so. It is obvious that there is a Red Queen's race between the ant-fungus-bacteria system and the parasitic microfungi that prey upon that system. What is less obvious is that co-evolutionary races also emerge *between mutualists* in the system wherever their interests are not in complete harmony. For example, when the system is home to more than one strain of friendly bacteria, then $r_g$ between these bacteria falls below 1 (the same is true when there is more than one strain of friendly fungus). This creates incentives for conflict between strains, which effectively become rivals for the benefits of the cooperative system. Those strains that outcompete others—e.g. mutualistic bacteria that poison their same-species rivals (Poulson et al., 2007)—increase their chances of reproduction while harming system productivity. As Frank (2003) writes,

> *the mixing of symbiotic lineages increases genetic diversity within hosts, favoring increased competition between symbionts. Symbiotic competition can disrupt the host. Thus, hosts gain from limiting the mixing of symbiotic lineages.* (p. 703)

This analysis of between-symbiont competition extends "to a wide array of biological interactions." It is also found in termite species, where high relatedness (and hence cooperativity) between symbionts makes mutualisms more stable and productive (Aanen et al., 2002). We see similar patterns in many other holobionts, as in insects that segregate symbionts in special organs to restrict competition, and in cases where mitochondrial diversity within animal hosts leads to conflict and sometimes death of the host (Ma & O'Farrell, 2016).

One typical host response is to attempt to ensure that the symbionts are a clonal monoculture so that $r = 1$ and inter-symbiont competition is suppressed. Attina ants have evolved the capacity for partner discrimination, such that they can usually identify different fungus and bacterial strains and actively eliminate them until a monoculture is restored. They also employ vertical transmission: ant queens

founding new nests carry with them a single strain of both bacteria and fungus.[290] The same strategies are also found in termites (Aanen et al., 2002) and various other host-symbiont systems. These strategies have their drawbacks. In the case of fungus-tending ants, the monoculture means a lack of diversity and makes the mutualistic fungus more prone to attack by parasites that it cannot resist, and makes the mutualistic bacteria less able to counter varieties of parasites. There is thus a trade-off involved. Interestingly, the evolutionary outcome suggests that the harms caused by between-symbiont Red Queen races are large enough that it is worth taking the risk of symbiont uniformity for the benefit of ending that race. The outcome is inefficient compared to perfect cooperation (Poulson, 2007), but appears to be optimal given D&E constraints. This appears to be a sort of "minimax" approach, making sacrifices in efficiency to avoid catastrophic conflicts that would more than offset possible efficiency gains.

### A.10.2 Growing market complexity: A driver of intervention

All advanced economies license doctors, regulate the safety of dangerous equipment: ban toxins in goods from food, to clothes, to construction materials; limit financial exploitation; and so on. Some regulations are undoubtedly forms of rent-seeking. The licensing of florists in Louisiana and manicurists in Minnesota are puzzles; perhaps the local markets are hyper-complex, and unscrupulous operators have led to rashes of death-by-pansy, or systemic collapse by poorly applied shellack? More likely it has something to do with Kleiner & Krueger's (2013) estimated 18 percent wage gains in licensed industries. Such cases aside, the coming chapters will suggest that as economic activity grows in breadth, and especially *complexity*, continued growth in the regulatory structure is inevitable. To take one example, there are more than 350,000 chemicals and chemical mixtures used commercially that did not exist at the time of Adam Smith (Wang et al., 2020). Many are cheap and useful in production, but dangerous for consumers, and many have a safety profile unknown even to specialists let alone laypeople perusing product labels. They may be combined in an unfathomably large number of ways, with new consequences in each

[290] This is a bottleneck strategy analogous to that used by most multicellular organisms, which grow from a single cell to ensure that intra-organismic r = 1.

case. Regulation will grow so long as citizens hope to benefit from the mutualistic side of innovations, while being protected from novel private and systemic harms.

### A.11.1 Detection and mimicry in floral "markets"

Another sociobiological case, closer to the logic of individual-level D&E in markets, is that of cognitive limits in flower-pollinator "markets". The role and language of *mimicry* is particularly instructive. Flower-pollinator interactions have been modelled as markets for decades; they involve a *voluntary choice* by a pollinator to "transact" with a plant, trading nutrients for pollination services. We can draw a valuable analogy between manipulative producers that exploit weaknesses in consumer D&E and the phenomenon of floral "mimics", such as the orchid mimic that appears similar to the "model" flower (i.e. the genuine article on which the "mimic" is "modelled") in the pollinator's eye. This strategy is not rare—to give a sense of magnitude, there are around 10,000 deceptive orchid species worldwide (Shrestha et al., 2020), which is similar to the total number of plant species native to mainland Europe (Winter et al. 2009), and many non-orchid species are also manipulative.

Mimics reduce the fitness of their transactors (i.e. pollinators) because they provide lower rewards than advertised, and frequently no reward at all. They also reduce "model" plant fitness, because the latter invests in providing an attractive return to the pollinator, but the presence of the mimic lowers the pollinator's expected returns and its likelihood to transact. In doing so, mimics reduce the overall productivity of mutualistic pollinator/flower markets. With such incentives in place, this process creates a Red Queen's race between the mimic and both partners of the mutualism: the pollinator might invest in improved cognitive D&E and the model in an improved signal, while the mimic invests to keep up with both. This returns us to the matter of *relative innovation capacity,* the speed at which each actor can run the race. Unfortunately for the mutualists, this race is asymmetric: "the mimic usually gains more from mimicry than the model loses", so that "a model cannot escape by gradually evolving a new pattern because the mimic will quickly catch up." (Joron & Mallet, 1998) Asymmetries in innovation capacity matter in human markets too, though the causes differ—we will get to them shortly.

### A.12.1 The design of property rights: Structuring problems for marginal market actors

Conventional property rights help establish markets, a precondition for "marginal" regulation via market actors, yet they also *structure* market actor interactions in simplifying ways. Like the monopoly for the miner or innovator discussed in Chapter 12, the effective monopoly over the use of everyday property is partly justified as a spur to investment and allocative efficiency.[291] Posner and Weyl (2018) argue that the individual monopoly over private property is an inefficient structural barrier to trade. It can be replaced, they suggest, by an approach closer the marginal pole. In particular, Posner and Weyl are concerned about the non-alienability of property, which reduces the potential gains from allocative efficiency largely for the reasons pointed to by Myerson and Satterthwaite (1983): the impossibility of efficient bilateral trade under incomplete information, and the consequent holdout problems.[292] In their proposed new model of property rights,

- all market actors would self-assess the price of each asset, and *must* sell the asset if any third party is willing to pay that price. The scheme "replaces the 'right to exclude' with a 'right to exclude anyone who does not pay the self-assessed price'"; and
- the incentive to overstate prices is suppressed by a tax on the total self-assessed value of property.

This model allows a subset of apparently mutualistic trades to take place that are blocked under conventional property rights. Like other marginal approaches, however, it makes the strategic problem more *complex*. The average real-world actor, who, again, does not even understand compounding interest (Bertrand et al., 2006), will bleed wealth through both the underestimation and overestimation of asset prices. Where asset values are underestimated, advantaged market actors can acquire assets at a discount. Where asset values are overestimated, this exposes

[291] In a world where there was no possibility of asset improvement or allocative gains—say, with a single, unimprovable good that all actors desired equally—the effect of property rights would be purely distributive.

[292] For example, a single household blocking acquisition of a large block of land for development could theoretically demand nearly the whole surplus of the development project. The ownership monopoly, like other monopolies, creates opportunities for rent extraction.

disadvantaged actors to higher-than-average rates of taxation, such that they cross-subsidise public goods for advantaged actors.

Posner & Weyl hope their scheme will address the holdout problem, where, say, a single landholder blocks development of an apartment complex or infrastructure project by asking an unreasonably high price. The landholder hopes to extract a share of the productive surplus associated with the project. Yet Posner and Weyl's regime would allow the capture of the surplus by other means: advantaged actors with superior or inside information will compulsorily purchase land from uninformed owners, and then self-assess land prices at much higher values. The outsized rewards to inside information would create new Red Queen's races over its acquisition.

Theoretical allocative efficiency means nothing without knowledge of with actors' actual capabilities,[293] the complexity of the resulting market, and so capacities to run the Red Queen's race over D&E. If inadvertently, Posner and Weyl reveal one of the great, but overlooked, virtues of conventional property rights: the simplifying nature of the ownership monopoly.

### A.12E1.1 Traditional intermediation: The productive component of financial activity

Some useful studies separate the productive component of finance from the background of parasitism. Simple intermediation activities between savers and investors still drive growth. Beck et al. (2014; 2016), for example, find that financial intermediation is socially valuable, while non-intermediation activities—the vast bulk of new financial activities—are *neutral or harmful* (also see Demetriades & Rewilak 2020 and Rousseau & Wachtel, 2011 on economic growth and quality-adjusted credit expansion). A banking sector based on productive intermediation for firms would be much smaller—Kehoe (2009) finds that 84 percent of US firms earn enough to cover all their investments internally. Unfortunately, the recent explosion in financial

[293] It is curious that the term "rationality" does not appear in the paper. They make brief reference to one behavioural problem, of people having personal attachments to assets: "we do not believe that such attachments are very common". It would be useful to start with Kahneman et al. (1990), who show that individuals develop attachments even to common objects like mugs once they own them. Family homes are another matter entirely. And what of impoverished indigenous populations, who are forced to self-assess a low value on their ancestral lands?

activity has little to do with intermediation. In the case of housing, Turner (2010), observes that growth in UK mortgage credit from 14 to 79 percent of GDP did not drive any increase in housing investment as a share of GDP.

### A.12E1.2 Capital substitutes, risk weightings, and shadow banking

Here I add some additional comments on capital requirements, risk weightings, shareholder empowerment, and shadow banking.

First, as noted, Basel III tightens restrictions on the leverage ratio from 1 percent up to 3 percent, which falls short of the recommended 15-20 percent in the literature. The gap is supposed to be filled by exotic—and complex—substitutes for equity, including "bail-in" securities such as CoCos (contingent convertible bonds).[294] CoCos are largely untested, and recent theoretical and empirical studies sound a warning. Theoretically, CoCos may ameliorate or aggravate bank failures, depending on buyer expectations about their likelihood of triggering (Fiordelisi et al., 2020). The effect of CoCos depends substantially on marginal regulation by market actors: they must price risk accurately, yet bank opacity makes this implausible. The FCA (2014) notes that "even professional investors may struggle to evaluate and price CoCos properly".[295] CoCo prices show worrying patterns of behaviour: Deutsche Bank's poor earnings in 2016, which put CoCo payouts at risk, drove "severe" and "CoCo-specific" movements in prices at other, unaffected banks (Bologna et al., 2018). That is, investors took this as a lesson about the risk of CoCos *in general*, rather than evincing an understanding of enterprise risk *in particular*. This turns bail-in securities into a driver of, rather than a constraint on, "contagious market reaction" (Goodhart, 2010; also see Li et al., 2022, on Chinese evidence). Finally, CoCos appear to aggravate managerial incentives. Bank of England research suggests that CoCos have a "strong positive effect on risk-taking behaviour", and this "works at cross-purposes of the tighter recapitalisation requirements they were allowed to be used for." (Fatouh et al., 2021; also see Martynova & Perotti, 2015 and Chan & van

[294] CoCos are fixed-income instruments that convert into equity at a specified trigger point, in order to recapitalise a struggling firm. The trigger is either a financial firm's book value to risk-weighted assets, market value to risk-weighted assets, or a discretionary regulatory decision that conversion is necessary to prevent insolvency.
[295] As the hedge fund manager Paul Singer observed in 2014 (see Admati, 2016), "[t]he unfathomable nature of banks' public accounts make it impossible to know which are actually risky or sound."

Wijnbergen, 2017). Persaud (2014) offers a useful overview of these problems and concludes that bail-in securities are "fool's gold".

Second is risk-weightings, which were fraught with problems and widely exploited by financial actors in Basel I and II (e.g. Mariathasan & Merrouche, 2014). The three rating agencies—S&P, Moody's, and Fitch—were "the key enablers of the financial meltdown" by underweighting sub-prime mortgage securities (Financial Crisis Inquiry Commission, 2011).[296] As noted in-text, complexity remains an issue. Haldane (2012) notes research suggests that simpler risk-weighting strategies dominate even moderately complex ones; to justify even moderate complexity, we need around 250 years of full business-cycle data and to suppose that nothing about cycle dynamics has changed in that period.

A more fundamental challenge worth nothing is that standardised risk-weightings encourage *homogeneity* in portfolio strategies. Heterogeneity arising from different banks' idiosyncratic risk assessments produces heterogeneity in strategy. Heterogeneity in turn tends to be protective on a systemic level. When errors creep into the standardised risk-weightings, they create opportunities to profit from taking on excess risk. Banks herd around that error and systemic risk increases.

Third is shadow banking. Around half of financial activity occurs in shadow banks that lie outside this regulatory structure. Carstens (2021) at the BIS reports that shadow banks' share in financial activity has "massively increased… since the Global Financial Crisis". The "status quo" of weak oversight "is unacceptable" (also see Gandré et al., 2020; Adrian & Ashcraft, 2012; and Buchak et al. 2018 on US real estate). Shadow banks are unaffected by the particular moral hazards of deposit insurance, but all the other promoters of excessive risk-seeking discussed in Chapter 12E1 still apply: the positive probability of state bail-outs, the disjunction between the interests of shareholders and debtholders, and pay-for-performance with intense competition over short-run returns. Tightening of capital regulation has redirected risky loan activity to non-banks (e.g. Irani et al. 2021; Kashyap et al. 2010; Plantin 2015). While an optimistic literature suggests that shadow banking, newly minted as "market-based finance", is better able to handle these risks (e.g. Cunliffe, 2017), the

[296] Financial Crisis Inquiry Commission (2011, p.xxv): "From 2000 to 2007 Moody's rated nearly 45,000 mortgage related securities as triple-A. This compares with six private sector companies in the United States that carried this coveted rating in early 2010."

2020 pandemic brought a "classic run" on shadow banks and asset fire sales, with the damage staunched by central bank intervention (Sengupta & Xue, 2020 at the US Federal Reserve; also see Czech et al. 2021 at the Bank of England). Aramonte et al. (2021) at the BIS note that "liquidity provision by non-banks tends to be more opportunistic and more prone to evaporate at times of stress". A Group of Thirty (2017) report on shadow banking warns that

> *the risks from the combination of high leverage and the ways in which credit is intermediated may be as great now as they were before the 2007–08 crisis.* (p. xii)

Finally, the Dodd-Frank "Say-on-Pay" reforms have perversely increased the influence of shareholders over managers. The evidence discussed in Section 12.2 suggests that increased shareholder power will aggravate, rather than contain, managerial risk-seeking. As Bebchuck & Spamann (2009) remark, "common shareholder interests cannot be counted on to eliminate incentives for excessive risk-taking." (p. 38). Managers must instead be aligned with the full set of stakeholders—shareholders, debtholders, and taxpayers.

### A.12E.3 Innovation-proof financial regulation? Raising managerial $r_s$

In light of this thesis, we can frame the regulatory problem as follows: How can we design a regulatory regime that transforms the payoffs of *unanticipated innovations* developed by actors who are, in the aggregate, *much smarter than we are*?

Bebchuk & Spamann (2009), who we will see propose a regulatory structure most in line with this thesis, observe that there are two main ways of regulating: (A) restrict the menu of options, also termed "activity restrictions", or (B) directly shape incentives. We can nuance this a little.

Though it may seem an academic point, *all* regulatory strategies, work by shaping incentives—i.e. option A, placing restrictions on the menu of options, means that managers must trade-off the risk of penalties against the rewards of violating the rules. What Bebchuk & Spamann mean by option B is the shaping of incentives by creating what we can identify as *payoff dependence*. That is, they point to $r_s$-based mechanisms that create alignment between the returns to managers and society, and

so “lockstep” payoff transformation. Currently, executive payoffs are primarily tied to equity, or a “levered bet on equity” via options. If we can align managerial incentives, as Bebchuk & Spamann (2009) put it, we can “harness bank executives’ information and expertise and thereby make executives work for, not against, the goals of banking regulation.” (p. 40)

**Inside debt**

An emerging literature examines strategies for doing so. The richest thread, beginning with Jensen & Meckling (1976), examines compensating managers with “inside debt”. This includes pensions or “supplemental” executive retirement plans, debt securities, and deferred compensation. Because these are managers’ claims against their own firms, their value is destroyed in the case of default.[297] As discussed in Section 12.1.2, the usual strategy of paying managers in equity (or effectively leveraged equity in the case of options) gives them incentives to shift value from debtholders to shareholders, favouring risky strategies for boosting growth. Managers paid in debt, on the other hand, have incentives to shift value from shareholders to debtholders. They will favour conservative strategies that maintain long-term firm value (e.g. Wei & Yermack, 2011). Inside debt is therefore thought to induce a general shift from short-termism to long-termism (e.g. Kane, 2002). Edmans & Liu (2011) make one of the most important recent theoretical contributions, pointing out why inside debt is superior to solvency-contingent bonuses: it induces managers to care not only about the binary of solvency versus bankruptcy, but also firm value in the case of bankruptcy. Note that compensation with inside debt is not exotic but is already common, although not large in scale; its use may reflect tax advantages rather than an effort to protect debtholders (Bebchuk & Jackson, 2005).

Empirical evidence supports this body of theory. Inside debt leads managers to reduce non-financial firm risk on a variety of measures—reducing leverage, increasing liquidity, increasing diversification, and lowering volatility in stock returns (Jensen & Meckling, 1976; Cassell et al. 2012; Wei & Yermack, 2011; Han & Pan, 2015; Li et al. 2018). Firms with inside debt have larger cash holdings (Liu et al.,

[297] Note, however, that some types of remuneration framed as inside debt are in fact disguised forms of equity or salary compensation, e.g. if debt has high seniority or may be cashed out at any time (Anantharaman et al. 2013; Bebchuk & Jackson, 2005).

2014) and a greater distance to default (Sundaram & Yermack, 2007; Borah et al., 2019). Accordingly, such firms receive cheaper access to capital, reflected in better terms on debt (Anantharaman et al., 2014; Dang & Phan, 2016; Lee et al., 2016) and lower costs of equity (Shen & Zhang, 2020). Removing inside debt schemes leads to a subsequent increase in firm risk-taking (Choy et al. 2014).

This evidence also applies to financial firms in particular. Inside debt reduces losses and default probability (including in the GFC), reduces volatility in stock returns, increases average asset quality and the hedging of interest rate risks, and reduces the costs of raising capital (Tung & Wang, 2010; Bolton et al., 2010; Bennett et al. 2015; van Bekkum, 2016; Belkhir & Boubaker, 2013; see Deng et al., 2019 on Chinese banks). Just as inside debt increases ordinary firm cash holdings, so does it encourage banks to maintain a larger liquidity buffer (Mehran & Tracy, 2016) and discourage dividend payouts (Anantharaman et al., 2014). Interestingly, banks with managers motivated by inside debt tend towards more traditional intermediation activities (van Bekkum, 2016). Mergers and acquisitions pursued by managers with inside debt are lower risk (Srivastav et al. 2018), and acquired firms tend to have abnormal long-term performance rather than abnormal stock returns (Phan, 2015). Finally, inside debt improves financial reporting quality and reduces earnings manipulation (He, 2014, Dhole et al. 2015, Chi et al., 2017, Wang et al., 2018), reflecting the reduced remunerative impact of short-term relative to long-term performance. In short, inside debt helps to achieve much of what we want out of banking reforms. It is at least a partial substitute for the complex and costly compliance measures currently used to reduce managerial risk-seeking.

Equity and inside debt compensation have distinct roles. Equity pay makes more sense for high-risk non-financial firms on the innovation frontiers.[298] They are not systemically important, such that there is no moral hazard, and debtholders are generally aware they are financing a risky proposition. “Lender beware” applies to such firms, but is entirely inadequate for banks. Inside debt makes particular sense for banks, where the value of risky innovation has been greatly exaggerated (again, see Volcker, 2009 and the research canvassed in Section 12.1), debtholders cannot be

[298] Kabir et al. (2018), for example, find that inside debt reduces R&D spending, which may indicate some undesirable effects of reduced risk appetite in frontier firms.

expected to impose market discipline, and default brings special, systemic risks such that solvency is a public good.

Because inside debt increases managerial sensitivity to debtholders over shareholders, shareholders will generally prefer not to design compensation schemes in this way (e.g. see Bolton, Mehran, & Shapiro, 2010). The "Say-on-Pay" scheme may well lead to less use of inside debt than before. Increasing the role of inside debt will therefore require regulatory change.

### Raising $r_s$ further

Bebchuk & Spamann (2009) propose incentive reforms that are particularly attractive in light of the framework of this thesis. They point out that the alignment created by inside debt is imperfect, and regulation should tie executive remuneration to a "broader basket of securities" so that managers are sensitive to the returns to *all* stakeholders—shareholders, debtholders, and taxpayers. Ideally this would comprise

> *a given percentage of the aggregate value of the bank's common shares, preferred shares, and bonds … minus any payments made by the government to the bank's depositors, as well as other payments made by the government in support of the bank, during the period ending at the specified time.* (p. 43)

That is, remuneration should be tied to particular classes of equity and debt in proportion to their value, minus the value of rescue packages. This is akin to Edmans & Liu's (2011) suggestion that the mix of equity and inside debt should reflect a firm's degree of leverage, with the addition in Bebchuk & Spamann of mechanisms addressing the unique moral hazard problems for systemically important financial firms. If remuneration is paid out over a sufficiently long period of time—at least a decade, as is the case for many existing pensions—then managerial return becomes a given percentage in the returns to society as a whole.

This is the regulatory realisation of high $r_s$—of lockstep payoff dependence. It is worth revisiting the ubiquity of $r_s \approx 1$ in biological cases of productive and stable mutualisms. High degrees of "skin in the game" is the main method for sustaining

cooperation in cross-species biological mutualisms.[299] It is the equilibrium solution, an ever-repeating point of convergence over hundreds of millions of years of evolution. As discussed in Chapter 5, Section 3, it is employed precisely because the alternative, the *c*-based mechanisms, are vulnerable to the Red Queen's race; so long as the means of generating high $r_s$ can be protected, it resolves the challenge of cooperating amid open-ended innovation. Bebchuk & Spamann's (2009) proposal is one answer to our question—how to control unanticipated innovations by actors who are smarter than we are.

This arrangement has some useful properties. The greater the leverage, the more managers become concerned about the value of inside debt and risk of government aid, and vice versa. This provides an incentive to choose levels of risk and leverage that maximise expected social value. If remuneration is paid out over a long period, then this incentive extends to ex-managers who hold pensions—actors who are armed with sectoral knowledge, insider knowledge, and a vested interest in both idiosyncratic and systemic solvency. Finally, such a scheme would allow the relaxing of the extremely complex and highly interventionist compliance architectures of Basel III and Dodd-Frank. As Rajan (2006) observes:

> *if problems stem from distorted incentives, the least interventionist solution might involve aligning incentives.* (p. 4)

Bebchuk & Spamann (2009) note that some activity restrictions may still be appropriate. Problems of bounded managerial rationality are an important consideration. Bell (2017), for example, argues that the banks that failed in the GFC were mostly led by managers who were "true believers" in a new normal of risk mitigation.[300] Kay and King (2020) call this the Viniar problem, in honour of David Viniar, former CFO of Goldman Sachs, who said regarding the 2008 crisis that "we were seeing things that were 25-standard deviation moves several days in a row". Which of course they were not: their models were simply a poor reflection of reality. In other words, managers may also misjudge systemic risk. They may gamble,

[299] Recalling that mutualisms are cross-species relationships, so that $r_g$ is necessarily zero.
[300] And see Fahlenbrach & Stutz (2011) on the apparent failure of banking CEOs to foresee the GFC and reduce or hedge their shareholdings.

increasing leverage, even with all their skin in the game, especially when other gamblers seem to be doing well.

Of course, beliefs follow incentives. Exposure to the downside of financial risk will lead managers to refine their beliefs. Nonetheless, the dangers that crisis present for democratic capitalism at the present time favour a precautionary approach. For the time being, at least, alignment of managerial incentives should be regarded as a complement to constraints on leverage, rather than a substitute**.**

### A.14.1 Escaping Olson's efficiency constraints

Olson (1999) describes the USSR's ingenious, if iniquitous, extractive strategy of regressive taxation, combining heavy average taxes with a low marginal tax rate. High effort was required to achieve subsistence (i.e. taxation is heavy where productive effort is inelastic) and only the most productive could earn above this level (i.e. taxation is low where effort is elastic). The surveillance dictator may be able to improve this system further, tailoring extraction more precisely according to observed elasticities for different segments of society. Such an authoritarian would extract like a price-detecting monopolist.

### A.14.2 Escaping Downs' paradox: Some dead ends

There have, at the periphery, been some theoretical efforts to resolve Downs' paradox while retaining the assumption of self-interest. One theory is that voters are irrational and believe they may influence the outcome. Yet this is inconsistent with *higher* rates of voting among the educated (likewise for tendencies to protest), and would imply that those enormous errors in assessments of probability ought to be found elsewhere in voters' lives—e.g. in their decisions in the market. In any case, Blais (2000, 62-70) puts this to rest: the majority of voters are reasonably accurate in their assessment of the miniscule odds of casting the pivotal vote. A second is that people vote to signal their cooperativity and improve their reputations (Overbye, 1995). Yet this can be done cheaply by lying about voting, given such behaviour is generally unobserved. It is dubious that voting behaviour offers much information about one's reliability in market transactions (Laver, 1997); it may indicate that one is hopelessly irrational or, more probably, lying about one's cooperativity. A third is

that the cost (C) is so low as to be negligible (e.g. Olson 1965, 164), but this cannot explain why individuals frequently sacrifice wages, wait in lines for hours, and in fragile democracies risk political violence, to participate in elections. Nor can it explain the prevalence and historical significance of vastly more costly means of political participation.

A fourth is Brennan and Lomasky's (1993) notion of "expressive" voting preferences, where individuals derive a *non-instrumental* satisfaction when they express their opinions in the ballot box—rather as people clap for a performance they enjoy. It is interesting that clapping is chosen as the analogy, given it is easier to explain why clapping emerges as a behavioural norm within a population of other-regarding actors (and, for other-regarding actors, clapping is instrumental). In any case, it is perhaps the most influential variant of Riker and Ordeshook's (1968; 1973) addition of a "D" term, which Riker & Ordeshook associate with voters' satisfaction at fulfilling their "duty", to form $PB + D > C$. Duty would be beyond the scope of this chapter, as it implies prosocial motives; however, any other kind of speculative payoff can be captured by the D term. Grofmann (1993), for example, simply supposes that people may have preferences for voting in the same way they have preferences for food. One might as well suppose that some self-interested actors have tastes for martyrdom too, but such suppositions eliminate the predictive content of "self-interested". This is true to the observation that the introduction of *ad hoc* preferences into rational choice models can "explain" any behaviour as rationally self-interested.[301] Another problem with the "D" term and expressive voting is its lump-sum nature: it cannot account for the observation that voting *is* instrumental, given empirically observed tactical voting and voting according to the perceived societal importance of a given election (Cox 1997; Franklin, Niemi & Whitten 1994). Indeed, the vast majority of individuals report voting for its instrumental social benefits (e.g. in the *American Citizen Participation Study*, by Verba et al., 1990).

### A.14.3 Extensions of Peltzman's principal-agent model

[301] Even behavioural inconsistencies in time preference may be explained away, for example, by supposing that individuals enjoy being time inconsistent. Perhaps they derive utility from befuddling economists?

Voters race against special interests who can reap large returns by corrupting the regulator. As noted in Chapter 9, the standard analytic approach to regulatory corruption is that of a principal-agent problem. Here, voters are the principal aiming to control the regulator as an agent. The regulator in turn would control firm behaviour. Firms and regulatory actors may, however, collaborate in regulatory capture, to establish and share in a stream of rents.

For a principal-agent model, let us borrow from Peltzman (1976), one of the three prototypical models of regulatory capture in the literature[302] and the one most easily adapted to our purposes. In *The Theory of Moral Sentiments,* Adam Smith (1759/1976) identifies two motives for self-interested actors: "avarice and ambition". In Peltzman's (1976) model of the representative, it is ambition that does all the work: representatives' payoffs are defined purely by *electoral success*.

To win votes, representatives make policy choices that affect the level of rent extraction—raising or lowering profits for special interests, while lowering or raising prices respectively for the consumer. The representative is indifferent to whether they serve firms or voters; they will favour whatever level of firm profits and consumer prices optimises their chance of electoral victory. The representative therefore evaluates the trade-off between the increase in support from special interests due to higher profits, and the withdrawal of support from consumers due to higher prices.

The degree of transfer from consumer to special interests chosen, and the size of the groups that benefit, depends on the relative sensitivity of these different groups to losses and benefits. Following Stigler (1971), and consistent with Olson (1965), the equilibrium level of rent is likely to be significant: small, well-coordinated, and well-informed groups with high per capita benefits will be much more sensitive to

[302] Dal Bo (2006) identifies three archetypal models of regulatory capture under democratic capitalism from the public choice literature: Stigler (1971), Peltzman (1976), and Tirole (1986). In each, as Dal Bo writes, the regulator is a "supplier of regulation" while market actors provide the demand. Market actors seek favourable regulatory treatment, which Stigler (1971) summarises as taking the forms of fiscal support and protection from competition. In return, "[t]he industry which seeks regulation must be prepared to pay with the two things a party needs: votes and resources." (p. 12) In my framework, each of these are strategies for raising regulator $r^*$ with special interests, usually through *de facto* rather than *de jure* means. As in my extension of Peltzman that follows, each operates through ambition (resources for political competition) and avarice (a private share in rents). For my purposes these are largely substitutable, drawing similar conclusions by examining the same hill from different angles.

transfers than a large, uncoordinated, uninformed public with very small per capita losses.

**Meta-institutional capture**

I extend Peltzman's approach to include strategies for capturing democratic institutions. That is, the regulator and special interests may collaborate not only around economic policies, but also to weaken the strength of voter oversight. Schedler's (2002) seven main strategies for democratic capture are discussed in Box A14.1.

What leads political actors to innovate in regulatory capture? In Peltzman's model, each political party attempts to win office by trading-off its appeal to voters and special interests. Let us assume that reputations are "sticky"—part of the definition of a reputation—so a party cannot costlessly and credibly shift positions, and so change

**Box A.14.1 Seven means of meta-institutional capture**

Schedler (2002) identifies seven means of meta-institutional capture used to distort the exercise of democratic choice: (i) interfere with the scope of elective offices; (ii) limit the range of political actors from which voters may choose;[303] (iii) bias voter opportunities to form preferences, through distortions in the media and campaign finance; (iv) disenfranchise voters, either legally or through practical obstacles; (v) distort the expression of choice through buying votes or voter intimidation, including floor votes; (vi) corrupt the process of preference aggregation, through electoral fraud or biases such as gerrymandering; and, finally, (vii) stifle officials' capacities to exercise their constitutional powers. Each of these attacks the D&E powers of the constitutionally-constrained voter, and so their capacity to domesticate the regulator.

coalitions, as the balance of voter or special interest influence changes. [304] The first implication is that each party can benefit by modifying meta-institutions in ways that

[303] Per Heclo's (1973, p. 21) quote of a systems analyst: "If I can set the options, I don't care who makes the choice."
[304] That parties are specialised in serving particular voters and special interests is important to generate motives for meta-institutional capture. Such stickiness emerges from established reputations

suppress the rent-seeking mechanisms and vote of other parties, while protecting its own. A party may suppress the influence of labour unions and preserve that of firms, or vice versa. Parties also have incentives for pursuing the democratic capture strategies of Box 13.1, such as gerrymandering and voter suppression strategies that disadvantage competitors, or weakening or strengthening political offices depending on their probability of holding them. For a party with a comparative advantage in special interest support, there are incentives to protect and increase the effectiveness of rent-seeking (e.g. via loose campaign finance rules) and to weaken voter D&E (e.g. by reducing the quality of information with which voters detect candidate quality). On the other hand, for a party specialised in serving voters, there are incentives to suppress rent-seeking and to strengthen voters' capacity to detect competitors' rent-seeking. Parties may also seek to modify the rules governing the media to favour firms that provide more favourable coverage, or may seek direct ownership of media. There are many possible strategic combinations; the key point is that *particular meta-institutions favour particular parties' electoral strategies*, and parties will innovate in institutional capture accordingly.

Again, in this model based purely on "ambition", representatives do not care whether they win by serving special interests or voters; they are substitutes in the candidate's pursuit of her electoral ambitions. If there is a bias towards voters or special interests, it emerges because of asymmetries in the Red Queen's race between voters or special interests. That is, either voters or special interests may be better at detecting regulator behaviour, and it is more advantageous to serve those with greater detection capacities.

This returns us to the question of asymmetries in innovation capacity. If special interest strategies of influence become relatively cheaper and more effective (e.g. through improvements in lobbyist strategy that reduce its electoral costs, or through more persuasive misinformation) or if their available resources grow (e.g. where unions or corporations become larger and/or increase profit margins), then parties that court them will win more electoral success and have greater opportunity to alter meta-institutions to favour their strategy. The reverse may occur if voter D&E

and relationships with particular voters and special interests. Ideological commitments could also produce such stickiness, though is not explicitly modelled until Chapter 16.

improves (e.g. if the quality of political information rises or the general level of education grows).

Olson (1965) and Stigler's (1971) pessimism about the case of regulatory capture gives equal cause for pessimism about meta-institutional capture. The two main reasons are:

1. *Relatively weak voter motives*. Special interests reap large and concentrated returns to rent-seeking, and will invest heavily in running the race (some observed strategies are discussed in Section 3.2.4). For the voter, a zero-cost vote to minimise consumer prices is still a nearly worthless vote. They may vote, but they will not run the race; voters will play the sleeping Hare.[305]
2. *Relatively weak voter innovation capacity*. In *complex* policy areas, even well-motivated voters will lose the race. This is a manifestation of the first-leg (Chapter 10) problem, where even ideally motivated consumers (transacting private goods with large welfare consequences) were unable to detect payoffs.

**Enter avarice**

The problem is worse still. Peltzman (1976) gestures towards, but does not analyse, a model with both of Smith's "avarice and ambition":

> *A more general model might make "dollars" (broadly defined to include, for example, employment of former regulators) a source of direct as well as indirect utility to the regulator*. (p. 6)

In this more complete model, elected representatives optimise not only over the probability of holding office, but also over its private material returns.

We can define two sources of private returns: First the salary, pension, and other benefits associated with public office. The higher the salary, the higher the returns to winning office, and so the stronger the pull of Peltzman's political optimum. Here avarice reinforces ambition—and its associated adaptive race over meta-institutions, for better or worse.

[305] We can only get a favourable outcome for voters if we assume they always win the race—that voter detection is perfect and "unbeatable", so they may indeed safely play the Hare.

Second, special interests also offer private returns to corruption. As Peltzman intimates, we can model this as elected representatives gaining a direct share in market actors' rents. This creates an additional bias towards special interests that is not counteracted by any corresponding force on the side of voters—representatives do not receive a meaningful material share in the reduction of consumer prices, and, once elected, self-interested voters will not hold representatives to account through collective action such as protest. Representatives manage the trade-off between the expected returns to serving special interests, with the expected return from remaining politically viable in the next election. Such patterns turn up empirically, as in Klašnja's (2015) study examining why political turnover is faster in corrupt democracies: once elected, representatives in corrupt systems have incentives to pursue privately-rewarding strategies that are politically sub-optimal.

In real world democracies, political salaries are low compared to potential special interest payments. The model with avarice predicts that innovation will lead to a world where representatives' incomes mainly come from rents, and that policy positions will reflect alignment with particular special interest blocs. Voters' inability to hold representatives to account *once they have been elected* becomes decisive; it disempowers even voters with perfect powers of detection. In the pure ambition model a happy equilibrium was possible with highly discriminative voters, who could threaten to spoil corrupt representatives' ambitions by removing them at the next election. Of course, there were reasons to be pessimistic about their odds of doing so. In any case, with avarice, the (remote) possibility of this happy equilibrium is eliminated. Electoral *candidates* might prefer that the misdeeds of incumbents are made visible, which would increase the ease of ousting them from office (although they may judge that this reduces their expected payoffs once elected). But only elected officials determine policy.

The upshot is that every representative, *once elected,* rationally prefers to establish a regime in which rents are easily concealed, and there are minimal meta-institutional and electoral constraints on rent extraction. In a model without collective action such as destabilising mass protests, there is no way for voters to hold elected representatives to account.

The democracy of self-interested actors will therefore devolve into a contest between special interest blocs. Where major special interests agree (e.g. firms on low corporate tax rates, or the neglect of externalities that mostly affect voters), policy will reflect this agreement. Where major special interests disagree (e.g. firms competing for monopoly rents), politics becomes the battleground of special interest blocs.

Under Coasian bargaining the outcome may produce efficient extraction from voters; special interests, in aggregate, behave as a monolith with $r_s > 0$ and coordinate in extraction. If firms' profits are comparable to the dictator's tax rate, the outcomes of such a "democratic" oligarchy may be similar. However, the absence of an authoritarian leader raises a fatal coordination problem: special interest blocs that win an advantage in the innovation race cannot commit to restraint, and there is no high-$r_s$ actor with the incentive to restrain and coordinate extraction. Thus, as Olson (1965) feared, we expect widespread roving banditry.

**Regime change: Voters as collaborators in meta-institutional capture**

So far our pessimism has centered on the voter's capacity to keep up in the Red Queen's race when the vote is zero-cost but still nearly worthless. The second question is of the nature of voter preferences, given self-interest. What "direction" would voters take democracy, even if they were capable and willing to invest?

The small but positive value of the vote for the individual does not carry the implication that voters will value *democracy* or *democratic principles*.

1. voters may value the votes of some citizens positively and some others negatively, and have little reason to support a universal franchise; and
2. where democracy leads to significant roving banditry, then like every pillaged society in history, people may rationally prefer a stationary bandit.

To examine the first of these problems, we need to further modify Peltzman's (1976) approach. He assumes that all voters want the same thing: lower prices. This turns them into a united interest on the side of democracy, another monolith, even if composed of individuals who prefer to sleep amid the race.

There are many possible configurations of voter interest that break this monolith into pieces. That is, voters may *disagree*. Buchanan & Tullock (1965), for example, observe that voters are themselves members of special interest groups (e.g. employees in particular industries, members of groups receiving particular welfare payments, members of groups holding particular asset classes, and so on). Perhaps more importantly, voters with (A) limited detection capacities and (B) limited willingness to invest may also be persuaded by political entrepreneurs that some special interests are allies and others are enemies. Subsidies for firms may be interpreted as favourable job-creators, or as the unfavourable wastage of public money; free trade may be represented as price-reducing or job-destroying; and externality pricing may be explained as welfare-improving or as warfare on a particular industry or economic region.[306] In any case, even our largely indifferent voters may arrange themselves into opposing blocs with different interests.

For the self-interested actor, the *de jure* power granted by democracy has no special moral status; it is like any other resource that individuals will prefer lies in their hands and those of their perceived allies. Thus, where voter preferences are in conflict, voters may *collaborate* in each of Schedler's (2002) distortions of democratic choice: supporting gerrymandering, disenfranchisement, suppression, and so on, where it increases the influence of their voting bloc and reduces that of their opponents.[307] Similarly, they will support stacking the courts when it means favourable decisions for their interests, interfering with the powers of official posts when they those posts tend to be occupied by opposing interests, and the distortion of information when it brings more voters to their side.

Our voters would do all this with a yawn, and the victorious side would meet little, if any, resistance—no rational individual would view the vote, or its loss, as a matter worthy of much attention.

[306] E.g. battle lines drawn over the future of the coal industry in the US.
[307] It is possible to avoid this outcome via voter omniscience, unimodal preferences, and the median voter theorem. The median voter is an Arrovian dictator and rather enjoys the position, and would prefer to keep the franchise symmetric—to keep their place at the median. Of course, the need for omniscience throws some spanners in the works. If everyone knew who the median voter would be, it would be irrational to turn out—and irrational to hold the vote at all. Voters would win the Red Queen's race against special interests outright, and more or less perfectly suppress regulatory capture and roving banditry, except for the caveat that they still would not control representatives once elected.

Certainly such patterns can be observed in democracies both nascent and mature, although they provoke great passion and fury. Take three recent examples from the US: voter support for eliminating the electoral college was similar among Democrat and Republican voters for much of recent history, but by 2020 reached 89 percent among the former and 23 percent among the latter, reflecting its relative advantages for each party (Gallup, 2020); around 60 percent of registered Democrats favour packing the Supreme Court, a position shared by only 5 percent of Republicans (Sanders, 2020); and similarly stark partisan divisions emerged over mail-in voting during the 2020 presidential elections (Clinton et al., 2020). It will be possible to say more about these patterns in later chapters that include $r_c$, which amplifies voters' incentives to defend *or* dismantle democracy, and explains the intensity of emotion, and collective action, in between-group conflicts.

Finally, where the democracy of self-interested actors produces widespread roving banditry, voters may reasonably judge that they would be better served by a strongman. This is, in the model here, an entirely rational choice. Unlike roving bandits, a stationary bandit has the incentives and coordination power required to efficiently restrain extraction and provide public goods. This too can be observed in some real-world democracies. Levitsky & Zeblatt's (2018) observe, in *How Democracies Die*, that in the twenty-first century the death of democracy "begins at the ballot box". Again, there will be much more to say about the patterns of voting against democracy in the coming chapters.

Peltzman (1976) observes that the literature on regulatory capture generally argues that "between the two main contending interests in regulatory processes, the producer interest tends to prevail over the consumer interest." This analysis has suggested that in a world of self-interested actors,

- the same sort of pessimistic result applies in the third leg and meta-institutional race;
- it is much more potent in models that combine avarice and ambition, especially given the impossibility of holding rent-seeking representatives to account once elected; and
- it is more potent still if we cease assuming that voter preferences are monolithic, and allow voters to have preferences over others' votes.

### A.14.4 Empirics on rent-seeking: Money in politics, and returns for firms and political actors

There is a large literature exploring how these dynamics play out in democratic capitalist societies. Let us consider a selection focused mostly on the US, with some additions from other advanced and emerging democracies. First, there is evidence for variation in demand according to rent-seeking opportunities, with lobbying and PAC investments concentrated among heavily-regulated industries with high availability of rents: utilities, telecoms, high-tech, banking, pharmaceutical and insurance, among others (Drutman, 2010, Hart, 2003 and Hill et al. 2013). Lobbying efforts also respond dynamically to the policy agenda of the period, which temporarily raises the opportunity for, or threat to, streams of rent. For example, Brulle (2018) finds that climate-related lobbying, predominantly from fossil fuel companies, rose from 1.7 percent of total lobbying expenditure in 2006 to over 9 percent in 2009/10 as interest in climate legislation peaked.

Empirical work on the effects of campaign finance for "ambition", i.e. the prospects for winning office, has been mixed but appears to be converging. Early work from Jacobson (1978) found that while challenger spending significantly increases their vote share, if anything the effect for the incumbent was slightly negative. Levitt (1994) and Palda & Palda (1998) found that campaign finance has no substantial effect on national vote share for either challenger or incumbent. These may appear to cast doubt on the rationality of great efforts to court donors. Prat (2002) offers an interpretation consistent with Peltzman's (1976) model, suggesting that "the informational benefit of spending is offset by the policy bias needed to raise contributions". If competing candidates each manage the trade-off between firm and consumer support, there will be no detectable advantage from campaign finance—although campaign finance does allow movement to a higher-rent equilibrium. At the same time some others, such as Abramowitz (1988), Grier (1989), have found that incumbent spending in senatorial races has a significant positive effect on vote share. Finally, another thread in the literature argues that campaign contributions matter for all candidates, exemplified by Snyder's (1990) finding of significant correlations between "the total amount of investor contributions a candidate receives, the monetary value of the favors he has promised, and his probability of winning."

As Stratmann (2005) notes, empirical studies of campaign finance are complicated by the problem that

> *incumbents' vote shares and spending are simultaneously determined: while spending influences the vote share, the expected vote share may influence spending. For example, incumbents who expect a competitive race may spend more to win re-election than incumbents who face no significant challenge. In this case, incumbents' vote shares and spending are negatively correlated, which may lead to the potentially erroneous conclusion that more campaign spending leads to smaller vote shares.* (p. 138)

More carefully specified studies suggest that Jacobson (1978), and many of the studies that followed, are the result of such omitted variable bias. To take some prominent examples, Green & Krasno (1988) used lagged spending as an instrument for current spending, and found that campaign finance is effective for challengers and incumbents alike. Exploiting state population and candidate wealth as instruments for spending, Gerber (1998) finds the same result in US Senate races. Erikson & Palfrey (2000) adjust for the endogeneity of spending on the closeness of the race, again with the same positive effect of spending for candidates.[308] Effect sizes appear to be significant: Stratmann (2004) finds that a 15 percent increase in incumbent television advertising results in an increase in the vote share of 1.2 percentage points; Gerber et al. (2011) find that a 10 percent increase in mailed political pamphlets increased vote share by 1.5 to 3.5 percentage points; and Hall's (2016) design exploiting differences in state corporate contribution laws finds that a 1 percentage point increase in party share of contributions leads to a half percentage point increase in share of the legislature. Candidates' aggressive efforts to secure special interest funds appear rational after all.

As for avarice, a growing body of evidence suggests these transactions produce private rewards for both representatives and market actors. Using a structural model of representatives' career choices, Diermeier et al. (2005) find that winning a Senate seat is worth nearly US$1.7 million. Palmer & Schneer (2015) that ex-senators win an

[308] Contrary to Jacobson (1978), they find that incumbent spending is, if anything, more effective than challenger spending.

additional US$125,000 per year on average from serving on the boards of directors of publicly traded companies. As for quid pro quo in office, Huang & Xuan (2016) find that politicians benefit from insider knowledge about stock outcomes, and also provide favours to companies in which they hold stock. The detailed study of such transactions is complicated by extensive efforts to conceal them, but corruption inquiries, such as those held by the New South Wales Independent Commission Against Corruption (or ICAC), with some of the most extensive powers among such institutions globally, provide a few insights. MPs have been caught using a wide variety of *de jure* levers to win private rewards, such as Daryl McGuire's exploitation of visa processes, planning processes, diplomatic processes, regulations, and parliamentary networks to extract bribes and acquire competitive advantages for himself and his associates (Bavas & Doak, 2020, October 14). Bodies with powers equivalent to the ICAC are unfortunately rare.

The effects of quid pro quo are perhaps easier to detect on the side of the firm. Jayachandran (2006) pioneered quantitative analysis of firm outcomes with her study of the surprise defection of Senator Jim Jeffords from the Republican party in 2001, which shifted the Senate majority to the Democrats' favour. She found that firms lost 0.8 percent in market value for every US$250,000 they had previously contributed to the Republican party, while contributors to the Democratic party experienced an opposite, although smaller, effect. Many other studies have followed: Richter et al. (2009) found that $1 spent on lobbying leads on average to a $6-20 reduction in tax burden the next year; Hill et al. (2013) found that shareholders value firm expenditures on lobbying, which is consistent with Kim (2008), Cooper et al. (2010), and Chen, Parsley & Yang (2014), who found that lobbying lifts equity returns relative to market and industry; Goldman, So & Rocholl (2006; 2009) and Akey (2015) found higher abnormal returns for firms that supported winning candidates in particular; Goldman, Rocholl & So (2013) found that such firms experience large increases in procurement contracts; and Yu & Yu (2011) found that lobbying firms "evade fraud detection 117 days longer, and are 38% less likely to be detected by regulators", and that "fraudulent firms on average spend 77% more on lobbying than nonfraudulent firms, and they spend 29% more on lobbying during their fraudulent periods than during nonfraudulent periods." Acemoglu et al. (2016) point to Timothy Geithner's nomination for Treasury Secretary in 2008 as an

interesting case study—his nomination produced abnormal returns of around 12 percent for financial firms with which he was connected, which were followed by abnormal negative returns when his nomination was thrown into question by tax issues.

The international and historical data is similar: Khwaha & Mian (2005) find that politically-connected firms in Pakistan gain superior access to finance from public but not private banks, with costs estimated up to almost 2 percent of GDP per year; Claessens et al. (2008) find that Brazilian firms providing campaign contributions win abnormal returns in the following years, with significant negative effects on GDP growth; Ferguson & Voth (2008) find strong increase in returns for companies politically connected to the Nazi party in early 1933; Faccio (2006) finds that political connections raise firm value across 47 countries; Faccio et al. (2007) finds that politically-connected firms are more likely to be bailed out than their peers; and in an international dataset, Faccio & Parsley (2009) find that the sudden death of a politician leads to a fall in value of companies headquartered in that politician's hometown. In the UK, Eggers & Hainmueller (2009) find that the wealth of Conservative MPs doubles after winning office compared to narrow losers; Fisman, Schulz & Vig (2014) find 3-5 percent higher annual growth in assets of winners versus runners-up, with the figures higher in more corrupt states; and finally, we see $r_g$ at work in Fafchamps & Labonne (2017), who find that politicians' relatives get better jobs in the Philippines.

Campaign finance aimed at electing a candidate can be construed as a "seat-buying" exercise, where the strategy is to help a favourably disposed candidate be elected. But once elected, does special interest money buy representatives' votes on Senate and House floors, or any other changes in representative behaviour? As for seat-buying, the study of "vote-buying" has been fraught with specification problems. Effect sizes have long been considered small. This is the conclusion of the highly-cited Hall & Wayman (1990), as well as Milyo et al. (2000) and Ansolabehere, de Figuieredo & Snyder (2003).

Nonetheless, at the same time Hall & Wayman argue that the influence of campaign contributions on political outcomes is pervasive and overwhelming, and that simplistic models of vote-buying fail because they do not identify or distinguish

complex strategies of influence. First, a share of campaign contributions can be best understood as a seat-buying strategy, supporting the election of representatives who are already favourably disposed to special interests—who are "pre-captured". Bronars & Lott (1997) and Wawro (2001) find that PAC contributions largely go to politicians who already support the PAC's views. Similarly, Tripathi, Ansolabehere & Snyder (2002) and Powell & Grimmer (2016) find that campaign finance from partisan interest groups is concerned with assisting re-election. If seat-buying money is not distinguished from vote-buying money, the former will dilute the detected effect of the latter.

Second, a share of contributions function by changing other, less visible, congressional behaviours. Again, Hall & Wayman (1990) set the tone for much of that which has followed. They find that

> *(1) the effects of group expenditures are more likely to appear in committee than on the floor; and (2) the behavior most likely to be affected is members' legislative involvement…* (p. 797)

In support of this argument, Tripathi, Ansolabehere & Snyder (2002) and Powell & Grimmer (2016) find that partisan interest groups focus on seat-buying, while non-partisan business interests instead contribute to representatives according to their influence in congressional committees. Wright (1990) found that committee-level voting, especially in the Ways and Means Committee, is "best explained by the total number of lobbying contacts representatives received from groups on each side of the issue." Consistent with this, Figueiredo & Silverman (2006) find that returns to university lobbying are particularly large when they have a representative on House or Senate Appropriations Committees. Finally, Grier & Munger (1991), Romer & Snyder (1994), and Milyo (1997) find that congressional committee members attract greater funding, and that funding is reallocated as committee memberships change. In short, campaign contributions more commonly buy participation in the design of, and haggling over, the legislation that is ultimately put to the vote.

Third, more careful study designs appear better able to detect the share of campaign finance specifically targeted at congressional vote-buying. Stratmann (1998) exploits the timing of campaign contributions around significant legislative events. Contributions from agricultural interests surge during legislative consideration of

agricultural subsidies, and he suggests this temporal proximity is a strategy to "prevent legislators from reneging on vote-for-contribution trades". Stratmann (2005) finds a correlation of 0.48 between the number of roll call votes and contributions in a given week between 1991 and 2001, and shows that the negative result on vote-buying of the aforementioned review by Ansolabehere, de Figuieredo & Snyder (2003) is reversed if effects are aggregated across papers[309] per the suggestion of Djankov & Murrell (2002). Engel & Jackson (1998) find evidence of withdrawal of special interest support in response to unfavourable floor votes. Finally, comparing House votes on financial services reform in 1991 versus 1998 (the latter the successful Glass-Steagall repeal), Stratmann (2002) finds that contributions are again timed around significant legislative events, and an additional US$10,000 in banking contributions over the period increases the probability of a House member voting for reform by eight percentage points.

This literature is largely consistent with theory predicting patterns of specialisation in seat-buying and vote-buying according to donor interest and candidate/representative qualities.[310] Witko (2006), for example, finds that donors tend to buy representatives' votes on low-visibility and non-ideological issues, while they buy representative legislative participation in high-visibility and ideological issues. Along similar lines, Fellowes & Wolf (2004) find that PACs and individual donors are more likely to buy votes on regulatory and tax legislation, but not on direct government expenditure, because in the latter connections between donations and votes would be easier to trace. This fits well with Peltzman's (1976) model, given representatives face a softer trade-off if special interest influence can be concealed. Given Red Queen dynamics, we may predict that such strategies will tend to improve over time as special interests and candidates learn to reduce the political costs of corruption. Indeed, these strategies have increased in complexity over the last several decades. Several of the above studies,[311] with more recent examples including Hill et al. (2013) and Akey (2015), show that modern political strategies usually comprise a mix of lobbying, hiring government employees, campaign contributions, and other

[309] With double the weighting for studies that correct for simultaneous determination of voting and contributions.
[310] E.g. favourably disposed candidates are more likely to be targets of seat-buying; representatives most likely to be open to changing their voting position are targets of vote-buying; and representatives in positions of greater power, including influence in congressional committees, are more valuable.
[311] E.g. Wright (1990) and Asolabehere, Snyder & Tripathi (2002)

strategies that together form what Akey terms "coordinated actions to build political networks". Networks appear to be critical: payoffs grow over time as relationships develop (Krosner & Stratmann, 2005); in hiring company directors, political connectedness is especially valued for firms in high-rent industries and for firms with exports or sales to government (Agrawal & Knoeber, 2001);[312] and the degree of political connectedness is a stronger predictor of lobbyists' wage premiums than is expertise in the policy area (Bertrand, Bombardini & Trebbi, 2014). The ways that such networks work and that complex, multi-pronged strategies cohere in practice remains something of a black box. From the perspective developed in this thesis, this should not be surprising—it is the result of creative innovation utilising distributed knowledge.

The same patterns emerge in our main case study, the financial industry and the GFC. The financial services industry has long been the largest source of campaign contributions, per Makinson (1992), and as Federal Electoral Commission data shows in 2020 (FEC, 2020). In terms of expenditure on lobbying, the sector tends to jockey for first place with the healthcare and pharmaceuticals industry. Consistent with evidence on the importance of networks, in 2020, nearly 70 percent of financial sector lobbyists were former government employees. Suarez & Kolodny (2011) note that major financial players pursue multi-pronged strategies, with individual firm strategies pursued via offices located in Washington, D.C., through memberships in interest groups such as the American Bankers Association, the Securities Industry Association, and Independent Insurance Bankers of America, and through individual or trade organisation PACs. These investments appear to modify legislator behaviour: the highly-cited Krosner & Strahan (1999) compare public-interest, political-institutional and private-interest explanations for state-level bank branching deregulation, and find that the private-interest theory—and so the relative strength of competing interests—provides the best explanation for observed patterns; Stratmann (2002), as discussed, identifies vote-buying behaviour by examining the timing of payments around votes on financial services regulation, and finds that

[312] The revolving door phenomenon may also be explained by the "regulatory schooling" hypothesis: that individuals with regulatory experience are valued because they lower regulatory risk for the companies that hire them rather than because they provide special interact access. The literature on this is mixed, with some major papers finding in favour and some against the hypothesis. Bertrand, Bombardini & Trebbi (2014) measure the effects of both access and expertise in the case of lobbyists, and find that the former trumps the latter.

changes in contributions are associated with changes in voting; Mian, Sufi & Trebbi (2010) find that financial industry campaign contributions influenced the propensity of legislators to vote for the Emergency Economic Stabilization Act; and Igan & Mishra (2014) find that financial industry PAC and lobbying expenditures increase the propensity for lawmakers to switch to deregulatory positions. Other evidence is consistent with the importance of committee membership. As Hall & Wayman's (1990), Tripathi, Ansolabehere & Snyder (2002), Powell & Grimmer (2016) and others found was the general case, so do campaign contributions from financial firms vary strategically depending on representatives' positions in the policymaking process. Krosner & Stratmann (1998) found that the mean contribution of commercial banks to Banking Committee members was over five times their contributions to non-members. Insurance PACs also give more to Banking Committee members, in part to offset the influence of commercial banking (an interest they have often been at odds with), but their greater concern with tax matters leads their spending to focus on the House Ways and Means Committee. Krosner & Stratmann (1998) go so far as to theorise that for many legislators, the core purpose of committees is to

> *foster repeated interactions, reputation building, and long-term relationships between the interest groups and members of the relevant committee, thereby increasing the likelihood that a high-contribution, high legislative effort equilibrium will exist*. (p. 23)

Krosner & Stratmann also find that financial firms donate to representatives who have little to do with financial legislation and about whom those firms have little knowledge. Whether or not this is the intention of the strategy, this may be expected to weaken intra-regulatory D&E: it makes exposing corruption more costly for each regulator and all their peers. It may plausibly help resolve collective action problems in rent-seeking by raising the benefits to coordination. Finally, there is also evidence of payoffs at the level of individual firms: echoing Yu & Yu's (2011) findings on purchased reductions in fraud detection, Igan, Mishra & Tressel (2012) find that in the lead-up to the GFC, lobbying financial institutions took on riskier loans, securitised them more rapidly, expanded mortgage loans more rapidly, saw their associated rents decline in value on the collapse of Bear Stearns and rise with the announcement of ESSA, and were more likely to be bailed out; Duchin & Sosyura

(2012) find that politically-connected firms were more likely to be recipients of government assistance under the Troubled Asset Relief Program; and Blau, Brough & Thomas (2013) found that politically-engaged banks were more likely to receive bailout funds, received more, and received them earlier than non-engaged banks.

As for the case of Gramm-Leach-Bliley, determining a causal link between a specific instance of regulatory change and special interest influence is fraught with complexity. There is no counterfactual. Moreover, there was no shortage of economists arguing a public-interest case for the integration of commercial and investment banking. Even today there are marginal voices willing to defend the Act in its entirety. Condemnation of the weak oversight of securitisation activities is nearly universal, but its staunch defenders at the time, such as Greenspan and Summers, might argue that they attempted to serve the general welfare to the best of their knowledge. It is enough to note that special interest influence was present throughout the whole process—that the failure to pass reform through the 1980s and up to 1998, and success in 1999, were the consequences of a contest between commercial banks initially in favour of repeal, and investment banks and insurers initially blocking repeal. Detailed accounts of this process offered in Krosner & Strahan (2014) and Suarez & Kolodny (2011) are consistent particularly with influence operating through key members of congressional committees. However one evaluates the merits of Gramm-Leach-Bliley as a particular case, the implications of this as a general pattern of policymaking are profoundly concerning.

Economic and political evidence suggests the sum effect of special interest influence on macroeconomic and macro-political outcomes in the US is significant and growing. It may be detected in the aforementioned rise of rent as a share of GDP post-1980 and increase in firm expenditures on political interventions. Consistent with these trends, Hall & Wayman (1990) find that representatives serve organized business interests more than unorganised voters "even when voters have strong preferences and the issue at stake is salient." Gilens & Page's (2014) well-known comparison of voter preferences by income level with enacted policies grimly concludes that "average Americans have little or no independent influence" over policy outcomes. To what extent can we regard these patterns as reflecting regulatory capture as a form of (A) uncoordinated meta-institutional evasion (i.e. roving banditry) or (B) coordinated meta-institutional capture (i.e. stationary bandit-style

oligarchy)? Interestingly, Gilens & Page latter emphasise the power of special interest *obstruction* in particular, such that policy change proceeds most where elites tend to agree. The history of Gramm-Leach-Bliley, as told by Krosner & Strahan (2014) and Suarez & Kolodny (2011), also highlights obstructive power, but they find that incremental change—the winning of small but meaningful advantages by one special interest or another—can add up to significant shifts over time. James Madison's (1787/2016) own hope, expressed in Federalist No. 10, was that in a large nation the "greater variety of parties and interests" might find it difficult to "find a common motive", such that competition between forces might cancel out—and indeed, this required that they would find it difficult to coordinate and "act in unison with each other." If such forces do find a competitive political equilibrium, however, this is not enough to deliver even Olson's optimum for the dictator or oligarchy. The outcome of obstruction is twofold: for contested issues we may expect sclerosis reflecting the balance of power between competitors, with that equilibrium periodically disrupted by new technologies and political strategies that establish new scleroses. For non-contested issues where elites do have "a common motive", such as on matters of taxation and social security spending, outcomes may *appear* coordinated. However, the sameness of incentives does not entail coordination: Hardin's cowherds all have the same incentives when they graze the commons, but they destroy it nonetheless.

It is unclear whether the special interests of democratic states typically have the means, or the requisite knowledge, to coordinate in efficiently restraining rent. Certainly there was some awareness in US financial firms, prior to the GFC, that their strategies were parasitic (e.g. predatory lending, betting against their own products, etc.) and associated with negative externalities (i.e. systemic risk). But a strategy that was efficient for the elite would require widespread knowledge and agreement about the private share of the expected costs, i.e. $r_s b$, and a capacity to then coordinate in halting extraction where $r_s b > c$. This asks too much of actors mostly following short-term, local incentives (i.e. $c$) within their organisations' payoff structures, such as bonuses for meeting performance or stock benchmarks. Coordination is equally unlikely where individual firms are seeking particularistic tax breaks. It may be modestly more feasible for laws that apply to all, such as a uniform change to personal and capital income taxation—if actors can only choose between rules applied equally to all then there is no coordination problem, and each will choose

according to $r_s b > c$. Still, this requires that actors have an accurate model of the world, for example, of the consequences of declining investments in infrastructure and social security for their private utility on the long-term, and a sufficiently low discount rate to value them. We have strong reason to be sceptical on both counts. This suggests the process is best understood as unrestrained roving banditry—not so much coordinated meta-institutional capture as meta-institutional decay. One problem with roving banditry, as Olson implies, is that it makes the stationary bandit more attractive.

### A.15.1 Prosocial preferences in infants and children

In the average human, prosocial preferences emerge early in life. Toddlers as young as 14 months show unsolicited helping behaviour toward strangers[313] (Warneken & Tomasello, 2006, 2007; and Hamlin et al. 2011), and preverbal infants as young as 6 months prefer individuals who are prosocial to third parties (Hamlin, Wynn & Bloom 2007). In a particularly fascinating result, Hamlin et al. (2011) find that preverbal infants at just eight months of age already hold preferences about how others treat prosocial or antisocial third parties—infants consistently prefer individuals who act prosocially towards prosocial others, and even those who punish antisocial others. Preverbal infants are budding strong reciprocators, and so budding lovers of law.

### A.15.2 Further remarks on other-regarding preferences: Strong reciprocity in the field, and in personality research

For some examples of the game theoretic patterns of prosocial behaviour discussed in Chapter 15 as expressed in the field, see Herrmann et al. (2008), Bowles et al. (2012), Gächter (2012), Franzen and Pointner, (2013), and Peysakhovich et al., (2014). Conditional cooperation is the most common behavioural type across a range of countries studied, usually comprising a majority of any given sample (e.g. Fischbacher, et al., 2001; Kurzban & Houser, 2005; Kocher et al., 2008; Hermann & Thöni, 2008). Strong reciprocation and the accompanying concern with fairness help

[313] Interestingly infant prosociality at just 20 months can be crowded-out and undermined by extrinsic rewards for cooperation (Warneken & Tomasello, 2008).

to explain certain patterns of economic behaviour, including preferences for incomplete contracts (Fehr, Klein & Schmidt, 2001), failures of market clearing (e.g. Fehr and Falk, 1999; Fehr et al., 1998), and the effects of different ways of allocating property rights on investment behaviour (Sonnemans et al., 2001).

In the psychological research, prosociality appears to be a fundamental component of personality structure. Individual tendencies for cooperation are stable over long periods of time, and are also *domain general* so that, for example, positive reciprocity in the Dictator Game predicts positive reciprocity in the Trust Game (Peysakhovich, Nowak & Rand, 2014; also see Kurzban & Houser, 2005). These characteristics have led some psychologists to introduce prosociality as a basic component of personality. Lee & Ashton's (2004) influential HEXACO personality model adds prosociality to the original "Big Five" personality traits. Studies comparing identical twins to non-identical twins, alongside other patterns of prosociality within families, suggest that genes explain up to 50 percent of the variance in prosociality. Environment, especially socioeconomic background, and measurement error explain the remainder (Kandler et al., 2019; also see Rushton, 2004; Cesarini et al. 2008; and Cesarini et al. 2009). The trait is uniquely predictive of behaviours in cooperation games. Hilbig et al. (2012) find that an individual's prosociality score predicts whether cooperation is conditional on punishment: In a Public Goods game, individuals scoring in the bottom quartile on the trait make low allocations in the absence of punishment, and double those allocations when punishment is introduced. Individuals in the top quartile for the trait, meanwhile, make the same allocations regardless of punishment. See similar results for Ultimatum and Dictator Games in Hilbig and Zettler (2009), for the Prisoner's Dilemma in Zettler et al. (2013), for the Dictator Game in Baumert et al., (2013), for the Trust Game in Thielmann & Hilbig (2015), and for responsiveness to incentives in the workplace in Zettler and Hillbig (2010). Interestingly, in the field, Johnson et al., (2011) find that prosociality is the most important personality trait for predicting job performance—reflecting that controlling shirking behaviour by punishment is largely ineffective, and prosocial workers are industrious in the absence of punishment.

### A.15.3 Further remarks on the relationships between strong reciprocity and trust

The literatures on strong reciprocity and trust only overlap occasionally (e.g. in Carlin & Love, 2013), and the division reflects that they ask different questions and use different methodologies. The strong reciprocity literature aims to build a model of how cooperation preferences respond to expectations (among other factors), mainly via experimental dilemmas. The trust literature mostly aims to identify patterns in those expectations, mainly via surveys, and identify correlations and causal relationships with a variety of behavioural, psychological, and economic variables. But the two are mutually supporting. The trust literature effectively assumes the conditionality inherent to strong reciprocation, given trust/distrust is expected to increase/decrease cooperativity. In turn, the strong reciprocator literature implies that expectations should predict behaviour, so that it would be fruitful to survey expectations across different firms and societies. Some important work connects the two, including Ostrom's (2000c) model of trust, reciprocity, and reputation. Ostrom writes:

> *experiments conducted in the United States, Switzerland, and Japan show that individuals who are initially the least trusting are more willing to contribute to sanctioning systems and are likely to be transformed into strong cooperators by the availability of a sanctioning mechanism.* (p. 141)

We can summarise this in game theoretic terms with an Assurance Game. If two strong reciprocators play a Prisoner's Dilemma, their prosociality gives them a stronger preference for the cooperative outcome. This transforms it into an Assurance Game (Figure A.15.1). As Kollock (1998) writes,

> *A common misunderstanding is that an Assurance Game presents no dilemma and leads inevitably to mutual cooperation. In fact, cooperation is not a dominating strategy, and if the person believes the partner will defect, the best the person can do is to defect as well. In other words, the Assurance Game has two equilibria: mutual cooperation, which is an optimal equilibrium, and mutual defection, which is a deficient equilibrium… The key issue in the Assurance Game is whether we can trust each other.* (p. 186)

Kollock notes that the Assurance Game has received less attention than the Prisoner's Dilemma, but he argues "that it is a more accurate model than the Prisoner's Dilemma Game of many social dilemma situations." One reason for this is that the prevalence of strong reciprocity converts many real-world Prisoner's Dilemmas into Assurance Games, and the key question is whether agents trust one another. For most social dilemmas, we need to know the value of $r_c$.

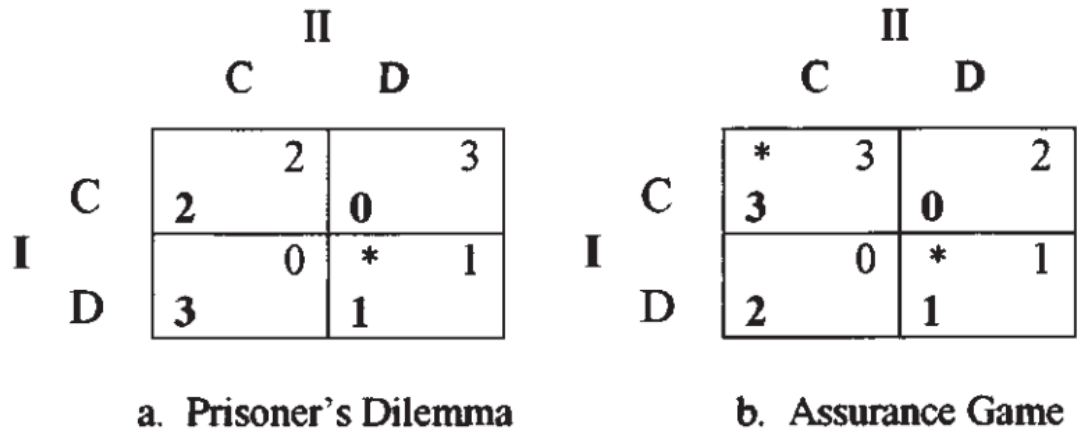



**Figure A.15.1** Where the two players prefer to cooperate, the Prisoner's Dilemma turns into an Assurance Game. As Kollock (1998) observes, that the cooperative outcome is preferred does not solve the game. If one player suspects that the other will defect, then the optimal strategy is to defect. Thus, there is no Nash equilibrium. The solution of the game depends on whether trust can develop.

## A.15.4 Further detail on cultural group selection and the moral sentiments

There is no "default" way of constructing minds in nature. Neither self-interest nor other-regarding preferences is somehow more "natural" than the other; the evolutionary process may produce anything from self-interest to total self-abnegation, depending on the circumstances. The idea that *Homo sapiens*' utility functions commonly include the payoffs of others is no more, or less, puzzling than the idea that we may be self-interested—just as the apparent selfishness of the chimpanzee or the blue-headed macaw is no more or less puzzling than the apparent generosity of the marmoset or African grey parrot. Whatever the pattern, it requires

an explanation, and Smith is correct that this explanation is to be found in the "fitting" of the animal to its social condition.

As evidence for innate prosociality has accumulated, so have social scientists and sociobiologists increasingly turned to analyse its evolutionary origins. Group selection is an obvious candidate, yet standard group selection is too weak to support generalised prosociality, primarily because of individual migration between groups. For natural selection to operate at the group level, there must be robust between-group variation in traits. Migration dilutes the differences between groups so that within-group variation, and so individual-level selection, dominates.

Cultural group selection is an attractive and influential alternative.[314] The foundations of cultural group selection theory were laid in early works including Campbell (1975) and Cavalli-Sforza & Feldman (1981), but its most influential form was set out by Boyd & Richerson (1982; 1988). Herbert Simon, ever the pioneer, also took an interest in cultural group selection as an explanation for altruism (Simon, 1990). The theory has continued to develop over the past few decades, and its strongest recent presentation is probably in Richerson et al. (2016), a paper with 14 authors across anthropology, psychology, and economics. I will briefly summarise the literature here. The key contributions are:

2 mechanisms for sustaining between-group cultural variation, and so suppressing the dilutionary effects of migration;
3 a process of group selection that favours "parochially prosocial" cultures that encourage cooperativity within groups and conflict between groups; and
4 a process of gene-culture coevolution, where the altered environment created by culture (and especially by regimes of punishment) leads to the selection of genes that favour innate cooperativity.

These are discussed below in detail.

[314] Note that some models may produce parochial prosociality without cultural selection, such as Choi et al.'s (2007) model of group selection under conditions of intensive warfare. However, the evidence that underpins cultural group selection is broad and deep, and as Choi et al. remark, such theories are more likely complements than substitutes.

## Stabilising cultural equilibria: The rise of between-group differences

First, a wide variety of causal mechanisms sustain between-group cultural variation by dampening the dilutionary effects of migration. Each of these mechanisms must work in a world where there is initially only individual-level selection. While the relative importance of the various causal mechanisms is contested, the consequences are not: it is widely documented that the members of cultural groups are homogeneous on many traits while neighbouring groups frequently differ significantly on some or many traits, that these differences can be sustained over long periods, that resulting cultural variation over time assumes a tree-like phylogenetic pattern (e.g. Mace & Holden, 2005), and that migrating individuals tend to adopt the practices of their new cultural homes (see Richerson et al., 2016, on each of these points).

Several of these mechanisms concern privately optimal patterns of social learning, which naturally has consequences for cultural transmission. The first of these is the "conformist learning" model, as presented in Boyd & Richerson (1985) and refined in many subsequent papers:[315] individuals win private fitness advantages by imitating the majority (Boyd & Richerson, 1985). Another is the "prestige-based" learning model of Henrich and Gil-White (2001), where individuals imitate successful individuals, especially where they have similar social roles.[316] Herbert Simon (1990) posits that individuals are "docile" (in its dictionary definition of "disposed to be taught") to varying degrees, and docile people tend to absorb instruction without independently evaluating the fitness consequences. Each of these models shares some features in common: rationality is bounded, the space of possible strategies is vast beyond our imagining (as described in Chapter 3), and payoffs are difficult and frequently impossible to calculate. It is therefore adaptive that each child enters the world ready to learn as rapidly as possible.[317] Simon's (1972) discussion of chess is a useful example: the game world of chess is miniscule and information is perfect, and yet "solving" chess is far beyond human capability. If a lone child attempts to learn how to play chess without instruction or imitation—say, by playing against themselves—in one lifetime they will have scarcely scratched the surface of the

[315] E.g. Henrich & Boyd, 1998…. REFS
[316] The chief of the tribe may be successful, but imitating the chief's behaviours may be dangerous.
[317] Famous study: Meltzoff (1988), finding that infants readily copy complex adult behaviours, including those that are "ritualistic" and of practical irrelevance to the task at hand.

accumulated cultural knowledge that could be quickly imparted to the eager social learner. As Simon (1990) summarises:

> *Most of our skills and knowledge, we learned from others (or from books); we did not discover or invent them. The contribution of docility to fitness is enormous.* (p. 210)

And, he adds,

> *much of the value of docility to the individual is lost if great effort is expended evaluating each bit of social influence before accepting it.* (p. 212)

Complementary mechanisms include the emergence of cultural markers that define social roles and relative status (McElreath et al., 2003; Efferson, et al., 2008), and the emergence of, and conformity with, punishment behaviours that help to stabilise cultural markers, norms, and other learned behaviours (Henrich & Boyd, 2001; Boyd et al., 2003; Salali et al., 2015).

### Equilibrium selection: Contests between cultural groups produces "parochial prosociality"

Second, these mechanisms allow a vast array of distinct cultural equilibria to be stabilised, whether they are harmful or beneficial for the group. Given this wide range of equilibria, a mechanism for *equilibrium selection* is required at the level of groups (Henrich, 2006). Contests between groups with different cultural equilibria favours the selection of equilibria that are prosocial. Darwin (1871) made a congruent observation one and a half centuries ago:

> *an advancement in the standard of morality will certainly give an immense advantage to one tribe over another … At all times throughout the world tribes have supplanted other tribes; and as morality is one important element in their success, the standard of morality and the number of well-endowed men will thus everywhere tend to rise and increase.* (pp. 131-132)

Or, as Herbert Simon (1990) puts it, "[a] society that instilled [altruistic] behaviors in its docile members would grow more rapidly than one that did not" (p. 210). Group-

beneficial norms can "hitchhike" on minds that are culturally conformist, or docile, even though conformist patterns of social learning were initially selected for their private fitness returns.

### Gene-culture coevolution: The culture-domesticated mind

Third, cultural group selection theory posits that evolving cultural forms rebounded to change the selective environment for genes. The consequence is *gene-culture coevolution* (Richerson & Boyd, 1998; Henrich & Boyd, 2001; Richerson & Boyd, 2008; Gintis, 2011; and see Johnson & Kruger, 2004 and Johnson & Bering, 2006 for applications to religion and the evolution of "theory of mind"). There are numerous other examples of gene-culture coevolution: cultures of fire use resulted in the genetic evolution of resistance to the toxins found in smoke (Hubbard et al., 2016); cultures of food cooking led to a reduction in the size of the human digestive tract; and the progressive development of language drove further changes in the larynx and brain structure, such that the infant brain "expects" to encounter a world of language (Pinker, 1994; Gintis, 2011). This gives a different meaning to the expression "[w]e shape our tools and thereafter our tools shape us" (Culkin, 1967, March 18).[318] The argument receives further support from some peculiarities of human evolution, such as the unusually low genetic diversity in our species going back hundreds of thousands of years—a pattern that can be explained by the rise of culture as a selective force (Whitehead et al., 2002; Premo & Hublin, 2009).

In this case, where prosocial cultures with punishment emerged, this favoured the evolution of innate motives that were consistent with and strengthened prosocial cultural equilibria—including "the evolution of other regarding motives like empathy and social emotions like shame" (Richerson & Henrich, 2012; also see Boyd & Richerson, 2009). There are two main reasons for this. First, in groups with regimes of prosocial punishment, a bias towards prosociality can raise individual fitness if it lowers the likelihood of being punished. In many tribal contexts, social consequences can be severe. Among some indigenous Australian groups, for example, a man visiting a women's sacred site might be killed or ritually maimed. If they survived,

[318] Culkin's (1967) modification of Churchill's "we shape our buildings and afterwards our buildings shape us."

> *when those legs healed, even in exile if they went to another tribe, people would see those scars and go 'why does he have those scars, I wonder what he did wrong, we should be careful of him* (quoted in Marshall & Scott, 2012).

Families of perpetrators are frequently ostracised, further steepening the genetic fitness costs of violations.

Second, and most interestingly, innate prosociality alters the process of cultural equilibrium selection so that it becomes partly rational. That is, prosociality favourably "biases our decisions about what norms to adopt". It thereby "accelerate[s] the evolution of norms and institutions that better and more often solve dilemmas of collective action" (Richerson & Henrich, 2009). Conformist transmission means that cultural traits may persist over very long timescales, but people also take an active interest in modifying cultural traits to strengthen the group—particularly when under competitive pressure. A group composed of prosocial individuals who are motivated to invest in rational equilibrium selection is more likely to stay competitive in a changing world. Faced with a novel public goods problem, for example, prosocial group members may call a meeting to devise and promote adaptive cultural norms. A group composed of free-riders will fail to invest in active cultural change; it may still exhibit cultural conformity because of the private fitness advantages it provides, but it will simply follow the random walk of blind cultural mutation, and so is more likely to fall into maladaptive equilibria.

As Richerson & Henrich (2009) summarise,

> *institutions arise as the joint products of our evolved psychology… and ongoing cultural evolution.* (p. 38)

It is worth a comment on how this is implemented psychologically. Emotions are thought to be derived from ancient internal motivators that directed organisms towards fitness-maximising behaviours. Such emotions are important for driving self-interested behaviours—the fear of a predator, the excitement of the hunt, or the pleasure of food for the hungry. Rationality does not choose these sensations, but is rather driven to serve them. That is, emotions are the underlying driver of the utility function (Cosmides & Tooby, 2000). In humans, emotions, and so the utility function, appear uniquely adapted for the social condition: they act as internal

motivators of moral, other-regarding behaviour, such that acts of charity, even those without any hope of reciprocation, activate the same neural reward centres as monetary gains; they take on complex other-regarding and socially strategic forms such as shame and remorse; and, most unusually, their external expressions act as signals to convey information that helps actors to coordinate and cooperate. A solitary-living creature, such as a typical lizard, may be driven by internal experiences of fear or pleasure, but there is no reason to *express* them externally. Such expressions are social displays. Frank (1988) theorised emotional displays as *commitment devices* that compel individuals to honestly disclose their intentions to their interaction partners. As Hirshleifer (1987) puts it, they are "the guarantors of threats and promises."

Emotional displays are cross-culturally universal (Ekman & Freisen, 1971), largely involuntary and difficult to accurately fake (Ekman, 1992), and so they offer surprisingly accurate information about the intentions of individuals to cooperate or defect. They thus act like a difficult-to-exploit cooperative “green beard” signal (see Dawkins, 1976); as Simon (1990) observed, there are “probably severe limits… as to how far deception will be successful”. Pradel et al. (2009) show, for example, that students can guess their classmates’ strategies in the Dictator Game better than chance, and that altruists tend to be friends with other altruists—the kind of sorting that allows cooperative strategies to survive.

Gintis (2011) summarises thus:

> *Gene–culture coevolution is responsible for human other-regarding preferences, a taste for fairness, the capacity to empathize and salience of morality and character virtues.* (p. 878)

In this view, gene-culture coevolution is the source of Smith’s “moral sentiments”. As Richerson & Henrich (2012) note:

> *Our theory has a back-to-the-future aspect. Adam Smith and Charles Darwin both made empathy the cornerstone of their theories of virtue. They observed that without the other-regarding virtue of sympathy, the social life that humans enjoy today would not be possible, much less reforms aimed at improving our social life.* (p. 58)

And, of course, it also explains between-group hatreds that would not be possible in the purely self-interested model.

With that, we have a theory about the *ultimate evolutionary cause* (individual and cultural group selection) that gives us the *proximate psychological cause* (innate tendencies for both self-interest and often parochial prosociality) of human behaviour.

### A.15.1 The D-term approach to voter behaviour

The D-term approach to explaining voter behaviour was briefly discussed in Chapter 14. Riker & Ordeshook (1968) originally framed as reflecting a sense of "duty". This accords with a social norms-driven account of voting behaviour—e.g. evidence that people "are motivated to vote by a sense of civic duty" (Blais, 2000), or that voters participate even if they believe "it will have no influence on benefits derived from policy outcomes" (Fowler & Kam, 2006). This gives us:

$$D - pc > c_v$$

And again setting $pc = 0$ per the standard approach:

$$D > c_v$$

This model may have some use, but it cannot explain observed voting patterns. The problem is the fixed, lump-sum nature of $D$. If there is a sense of duty to vote, it is sensitive to assessments of the importance ($b$) and closeness ($K$) of a given election (Edlin, Gelman & Kaplan, 2007), and evaluations of whether the actors affected are trustworthy and deserving of reciprocation ($r_c$). It is notable that other social norms, such as against lying, also vary in strength according to the evaluated importance of the case, and the trustworthiness of the target actor. The more we allow $D$ to accommodate these factors, the more it approaches $r_c K\overline{B}_o$ and takes us back to Equation 16.2.

The same problem applies to using D-terms to deal with fairness considerations. As cases #1 and #2 in the previous section show, we need the size of the D-term to respond flexibly to offset other-benefit values ($K\overline{B}_o$), given these values may be

measured in loose change or in millions of dollars. Loose change should be ignored, with a negligible D-term; the imposition of a cost of millions to provide large benefits to the group would require a much larger D-term. In all these cases, $r_c$ has the benefit of varying in naturally appropriate ways; the greater the special private cost imposed on a given voter, the lower $r_c$ falls.

However, the D-term approach may at the same time be necessary. A model where individuals weight $b$ cannot, for example, reasonably explain positive offers in the standard Dictator Game, or the particular attraction to the 50:50 split. So long as $r_c < 1$, even the prosocial dictator should make an offer of zero. And if $r_c = 1$, the dictator is indifferent to any offer. To reliably make a positive offer, the dictator would need to weight the recipient's marginal dollar gains more strongly than their own (i.e. $r_c > 1$), in which case they always would give all the money away. Norms provide a way of explaining the strong attraction to the 50:50 split. Alternatively, it may be possible to explain as a consequence of the strong reciprocator: the optimum is to neither be a "sucker", nor to render other group members "suckers".

Finally, a possible role for norms is in the management of the problems of limited information and bounded rationality. Given there is considerable uncertainty about the variables $r_c$, $K$, and $\overline{B}_o$, voters may employ the cultural norm "voting is beneficial" as a heuristic. Our species is responsive to information indicating payoffs, but also has an adaptive (and morally-charged) tendency to copy cultural behaviours when they cannot determine that they know better. Thus, we may find it useful to think about norms as heuristic shortcuts when information is poor, rather than as being the main substance of other-regarding preferences.

A complete model of the voter's utility function must include a context-sensitive, other/group-payoff-optimising component, and some form of lump-sum payment, with some heterogeneity across individuals in their relative importance. How those lump sums are best understood, and how "norms" relate to the motives of the strong reciprocator, is a matter for further investigation.

### A.15.2 Escaping the "altruistic explosion": Strong reciprocity

Voters following Equation 15.2 may, as noted, support unfair policies that are punitive to themselves or their subgroup, so long as they benefit society on average.

They may also pay implausibly high costs to vote: if $r_c = 0.2$, $K = 8$, and $\overline{B}_o = 1000$, then this could justify a $c_v$ up to \$1600. To take this to its extreme, if the individual knows they are the pivotal voter in a population of one million (i.e. $K = 1{,}000{,}000$), $c_v$ can be as high as \$200 million. The strong reciprocator will consider it unfair for one individual to bear the large private costs of voting to benefit others; there is no possibility of the voter benefiting in the long-run from membership of such a group. , and the willing voter would be reaching beyond any reasonable expectations about generalised reciprocity.

We can ask how things change if we remove the element of unfairness: how would the individual respond if everyone else was also paying the cost of \$1600? Here the answer is quite simple: if the net benefits of the democratic choice is $\overline{B}_o = 1000$, then it would be a net loss for everyone to pay \$1600 to receive it. If we take $\overline{B}_o$ as *net* of average voting costs, so that everyone pays \$1600 to gain \$2600 on average, this would be acceptable—so long as voting was enforced, and if the benefits were reasonably spread. But where voting is optional, this would still create large and unacceptable "sucker" problems: defectors for whom $r_c$ is low or zero enjoy a large payout, the \$2600, without paying the high costs of participation.

As a consequence, we can expect rational strong reciprocators to prefer democratic institutions that ensure the cost of voting is small and fairly distributed. If the only cost of the vote is the time spent, and so the opportunity cost, this is fair in the sense that all participants will tend to sacrifice the same small proportion of their income (of course, voters will in this case be troubled when there are inequities in the length of queues to vote). Other explicit costs of running democracy are managed through the tax system, i.e. via rules that ensure that even self-interested actors pay their share.

# Bibliography


Aanen, D. K., Eggleton, P., Rouland-Lefevre, C., Guldberg-Frøslev, T., Rosendahl, S., & Boomsma, J. J. (2002). The evolution of fungus-growing termites and their mutualistic fungal symbionts. *Proceedings of the National Academy of Sciences*, *99*(23), 14887-14892.

Aarøe, L., & Petersen, M. B. (2014). Crowding out culture: Scandinavians and Americans agree on social welfare in the face of deservingness cues. *The Journal of Politics*, *76*(3), 684-697.

Abaluck, J., & Gruber, J. (2009). Choice inconsistencies among the elderly: Evidence from plan choice in the Medicare Part D Program (*Working Paper No. 14759*). National Bureau of Economic Research.

Abaluck, J., & Gruber, J. (2016). Evolving choice inconsistencies in choice of prescription drug insurance. *American Economic Review*, *106*(8), 2145-2184.

Abdukadirov, S. (2010). The problem of political calculation in autocracies. *Constitutional Political Economy*, *21*, 360-373.

Abramowitz, A. I. (1988). Explaining Senate election outcomes. *American Political science review*, *82*(2), 385-403.

Acemoglu, D. (2003). Why not a political Coase theorem? Social conflict, commitment, and politics. *Journal of comparative economics*, *31*(4), 620-652.

Acemoglu, D., Naidu, S., Restrepo, P., & Robinson, J. A. (2019). Democracy does cause growth. *Journal of political economy*, *127*(1), 47-100.

Acemoglu, D., & Robinson, J. A. (2019). *The narrow corridor: How nations struggle for liberty*. Penguin UK.

Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., & Mitton, T. (2016). The value of connections in turbulent times: Evidence from the United States. *Journal of Financial Economics*, *121*(2), 368-391.

Acemoglu, D., Verdier, T., & Robinson, J. A. (2004). Kleptocracy and divide-and-rule: A model of personal rule. *Journal of the European Economic Association*, *2*(2-3), 162-192.

Admati, A. (2016, April 4). How Effective Capital Regulation can Help Reduce the Too-Big-To-Fail Problem. Presentation at Federal Reserve Bank of Minneapolis Conference: Ending Too-Big-to-Fail.

https://www.minneapolisfed.org/~/media/files/publications/studies/endingtbtf/april-4-symposium-presentations/admati-mpls-fed-abbreviated.pdf?la=en

Admati, A. R. (2014). The compelling case for stronger and more effective leverage regulation in banking. *The Journal of Legal Studies*, *43*(S2), S35-S61.

Admati, A. R., DeMarzo, P. M., Hellwig, M., & Pfleiderer, P. (2010). *Fallacies, irrelevant facts, and myths in the discussion of capital regulation: Why bank equity is not expensive* (No. 2010, 42). Preprints of the Max Planck Institute for Research on Collective Goods.

Adrian, T. & Ashcraft, A. B. (2012). Shadow banking regulation. *Ann. Rev. Finance Econ.* *4*(1), 99-140.

Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S., & Evanoff, D. D. (2014). Predatory lending and the subprime crisis. *Journal of financial economics*, *113*(1), 29-52.

Aghion, P., Antonin, C. & Bunel, S. (2021). *The Power of Creative Destruction*. Harvard University Press.

Aghion, P., Howitt, P., & Mayer-Foulkes, D. (2005). The effect of financial development on convergence: Theory and evidence. *The quarterly journal of economics*, *120*(1), 173-222.

Agrawal, A., & Knoeber, C. R. (2001). Do some outside directors play a political role?. *The Journal of Law and Economics*, *44*(1), 179-198.

Ahlerup, P., Olsson, O., & Yanagizawa, D. (2009). Social capital vs institutions in the growth process. *European journal of political economy*, *25*(1), 1-14.

Aikman, D., Galesic, M., Gigerenzer, G., Kapadia, S., Katsikopoulos, K., Kothiyal, A., ... & Neumann, T. (2021). Taking uncertainty seriously: simplicity versus complexity in financial regulation. *Industrial and Corporate Change*, *30*(2), 317-345.

Aikman, D., Nelson, B., & Tanaka, M. (2015). Reputation, risk-taking, and macroprudential policy. *Journal of Banking & Finance*, *50*, 428-439.

Akçay, E. & Roughgarden, J. (2007). Negotiation of mutualism: rhizobia and legumes. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1606), 25-32.

Akesson, J., Hahn, R. W., Metcalfe, R. D., & Monti-Nussbaum, M. (2022). The Impact of Fake Reviews on Demand and Welfare. NBER conference paper (f166391). https://conference.nber.org/conf_papers/f166391.pdf

Akey, P. (2015). Valuing changes in political networks: Evidence from campaign contributions to close congressional elections. *The Review of Financial Studies*, *28*(11), 3188-3223.

Albanese, G., Barone, G., & de Blasio, G. (2022). Populist voting and losers' discontent: Does redistribution matter?. *European Economic Review*, *141*, 104000.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. (2002). The generation of antibody diversity. In *Molecular Biology of the Cell. 4th edition*. Garland Science.

Albright, M. K. (2018). *Fascism: A warning*. HarperCollins.

Alchian, A. A. (1950). Uncertainty, evolution, and economic theory. *Journal of political economy*, *58*(3), 211-221.

Alder, M. (2019, October 31). Trump judge pick called 'lazy, arrogant' by peers, ABA says. *Bloomberg Law*. https://news.bloomberglaw.com/us-law-week/trump-judicial-pick-called-lazy-arrogant-by-peers-aba-letter-says

Alesina, A., Baqir, R., & Easterly, W. (1999). Public goods and ethnic divisions. *The Quarterly journal of economics*, *114*(4), 1243-1284.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic growth*, *8*, 155-194.

Alesina, A., & La Ferrara, E. (2002). Who trusts others?. *Journal of public economics*, *85*(2), 207-234.

Alesina, A., & La Ferrara, E. (2005). Ethnic diversity and economic performance. *Journal of economic literature*, *43*(3), 762-800.

Alesina, A., Stantcheva, S., & Teso, E. (2018). Intergenerational mobility and preferences for redistribution. *American Economic Review*, *108*(2), 521-554.

Algan, Y., Beasley, E., Cohen, D., & Foucault, M. (2018). The rise of populism and the collapse of the left-right paradigm: lessons from the 2017 French presidential election.

Allocati, N., Masulli, M., Di Ilio, C. & De Laurenzi, V. (2015). Die for the community: an overview of programmed cell death in bacteria. *Cell death & disease*, *6*(1), e1609-e1609.

Anagol, S., & Kim, H. H. (2012). The impact of shrouded fees: Evidence from a natural experiment in the Indian mutual funds market. *American Economic Review*, *102*(1), 576-593.

Anantharaman, D., Fang, V. W., & Gong, G. (2013). Inside debt and the design of corporate debt contracts. *Management Science*, *60*(5), 1260-1280.

Andersson, L. M., & Pearson, C. M. (1999). Tit for tat? The spiraling effect of incivility in the workplace. *Academy of management review*, *24*(3), 452-471.

Anduiza, E., Gallego, A., & Muñoz, J. (2013). Turning a blind eye: Experimental evidence of partisan bias in attitudes toward corruption. *Comparative Political Studies*, *46*(12), 1664-1692.

Ang, J. S., Cheng, Y., & Wu, C. (2015). Trust, investment, and business contracting. *Journal of Financial and Quantitative Analysis*, *50*(3), 569-595.

Ansolabehere, S., De Figueiredo, J. M., & Snyder Jr, J. M. (2003). Why is there so little money in US politics?. *Journal of Economic perspectives*, *17*(1), 105-130.

Anton, M. (2016). The flight 93 election. *Claremont Review of Books*, *5*.

Apperley, A. (1999). Hobbes on democracy. *Politics*, *19*(3), 165-171.

Aramonte, S., Schrimpf, A., & Shin, H. S. (2022). Non-bank financial intermediaries and financial stability. *The Research Handbook of Financial Markets*.

Arcand, J. L., Berkes, E., & Panizza, U. (2015). Too much finance?. *Journal of Economic Growth*, *20*(2), 105-148.

Arrow, K. J. (1972). Gifts and exchanges. *Philosophy & Public Affairs*, 343-362.

Arrow, K. J. (2006). The economy of trust: An interview with Kenneth Arrow. *Religion & Liberty*, 16(3), 3, 12-13.

Arrow, K. J. (2012). *Social choice and individual values* (Vol. 12). Yale university press.

Arrow, K. J., & Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica: Journal of the Econometric Society*, 265-290.

Atran, S., & Henrich, J. (2010). The evolution of religion: How cognitive by-products, adaptive learning heuristics, ritual displays, and group competition generate deep commitments to prosocial religions. *Biological Theory*, *5*, 18-30.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., & Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, *135*(2), 645-709.

Avery, C., & Zemsky, P. (1998). Multidimensional uncertainty and herd behavior in financial markets. *American economic review*, 724-748.

Awrey, D., 2012. Complexity, innovation, and the regulation of modern financial markets. *HARv. Bus. L. REv.*, *2*, p.235.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. Science, 211(4489), 1390-1396.

Bacon, F. (1890). *The essays or counsels, civil and moral*. Clarendon Press. (Original work published 1597)

Bacon, F. (2000). *Novum Organum*. Cambridge University Press. (Original work published 1620)

Bakir, V., & McStay, A. (2018). Fake news and the economy of emotions: Problems, causes, solutions. *Digital journalism*, *6*(2), 154-175.

Balliet, D., Wu, J., & De Dreu, C. K. (2014). Ingroup favoritism in cooperation: a meta-analysis. *Psychological bulletin*, *140*(6), 1556.

Barkai, S. (2016). *Declining labor and capital shares* (No. 257). Working paper.

Barnett, W. P., & Hansen, M. T. (1996). The red queen in organizational evolution. Strategic Management Journal, 17(S1), 139-157.

Barone, E. (1908). Il ministro della produzione nello stato collettivista. Giornale degli economisti, 37, 267-293. (Translation available in R. Marchionatti (Ed.) (2004). Early mathematical economics, 1871-1915. Psychology Press)

Barth, J. R., Caprio Jr, G., & Levine, R. (2007). Reassessing the rationale and practice of bank regulation and supervision around the globe. In *Seminar on Current Developments in Monetary and Financial Law. Washington DC.*

Baumert, A., Schlösser, T., & Schmitt, M. (2014). Economic games. *European Journal of Psychological Assessment,* 30(3), 178-192.

Baumol, W. J. (1990). Entrepreneurship: Productive, unproductive, and destructive. *Journal of Business Venturing*, 11(1), 3-22.

Baumol, W. J. (2004). Red-Queen games: Arms races, rule of law and market economies. Journal of Evolutionary Economics, 14(2), 237.

Bavas, J. & Doak, E. (2020, October 14). Former NSW MP Daryl Maguire tells ICAC he didn't report 'cash-for-visa scam' because he was making money. *Australian Broadcasting Corporation*. https://www.abc.net.au/news/2020-10-14/daryl-maguire-admits-string-of-corruption-allegations-at-icac/12765284

Bebchuk, L., & Jackson, R. (2005). Executive pensions. *Journal of Corporation Law*, 30, 823-855.

Bebchuk, L. A., & Spamann, H. (2009). Regulating bankers' pay. *Georgetown Law Journal*, 98, 247.

Becchetti, L., Manfredonia, S., & Pisani, F. (2022). Social capital and loan cost: the role of interpersonal trust. *Sustainability*, *14*(3), 1238.

Beck, T., Chen, T., Lin, C., & Song, F. M. (2016). Financial innovation: The bright and the dark sides. *Journal of Banking & Finance*, *72*, 28-51.

Beck, T., De Jonghe, O., & Schepens, G. (2013). Bank competition and stability: Cross-country heterogeneity. *Journal of financial Intermediation*, *22*(2), 218-244.

Beck, T., Degryse, H., & Kneer, C. (2014). Is more finance better? Disentangling intermediation and size effects of financial systems. *Journal of financial stability*, *10*, 50-64.

Becker, B., & Ivashina, V. (2015). Reaching for yield in the bond market. *The Journal of Finance*, *70*(5), 1863-1902.

Becker, E. B. & Bonni, A. (2004). Cell cycle regulation of neuronal apoptosis in development and disease. *Progress in neurobiology*, 72(1), 1-25.

Becker, G.S. (1976). Altruism, egoism, and genetic fitness: Economics and sociobiology. *Journal of economic Literature*, 14(3), 817-826.

Becker, G.S., 1991. *A treatise on the family: Enlarged edition*. Harvard university press.

Becker, G. S. (2010). *The economics of discrimination*. University of Chicago press. (Original work published 1957)

Becker, S. O., Fetzer, T., & Novy, D. (2017). Who voted for Brexit? A comprehensive district-level analysis. *Economic Policy*, *32*(92), 601-650.

Beerling, D.J. & Osborne, C.P. (2006). The origin of the savanna biome. *Global change biology*, 12(11), 2023-2031.

Beinhocker, E. D. (2006). *The origin of wealth: Evolution, complexity, and the radical remaking of economics*. Harvard Business Press.

Belkhir, M., & Boubaker, S. (2013). CEO inside debt and hedging decisions: Lessons from the US banking industry. *Journal of International Financial Markets, Institutions and Money*, *24*, 223-246.

Belkin, A. (2019). Court expansion and the restoration of democracy; The case for constitutional hardball. *Pepperdine Law Review*, 1, 19-50. https://digitalcommons.pepperdine.edu/plr/vol2019/iss1/2/

Bell, S., & Hindmoor, A. (2014). The ideational shaping of state power and capacity: Winning battles but losing the war over bank reform in the US and UK. *Government and Opposition*, *49*(3), 342-368.

Beltratti, A., & Stulz, R. M. (2012). The credit crisis around the globe: Why did some banks perform better?. *Journal of financial economics*, *105*(1), 1-17.

Bénabou, R., Ticchi, D. & Vindigni, A. (2015). Religion and innovation. *American Economic Review*, 105(5), 346-351.

Bennett, R. L., Güntay, L., & Unal, H. (2015). Inside debt, bank default risk, and performance during the crisis. *Journal of Financial Intermediation*, *24*(4), 487-513.

Benson, M., & Rochon, T. R. (2004). Interpersonal trust and the magnitude of protest: A micro and macro level approach. *Comparative political studies*, *37*(4), 435-457.

Bentham, J. (1824). *The book of fallacies: From unfinished papers of Jeremy Bentham*. John and H. L. Hunt.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior*, *10*(1), 122-142.

Bergh, A., & Bjørnskov, C. (2011). Historical trust levels predict the current size of the welfare state. *Kyklos*, *64*(1), 1-19.

Bergstrom, C.T. & Lachmann, M. (2003). The Red King effect: when the slowest runner wins the coevolutionary race. *Proceedings of the National Academy of Sciences*, *100*(2), 593-598.

Berman, J. B., & Li, G. (2020). Characteristics of criminal cases against physicians charged with opioid-related offenses reported in the US news media, 1995–2019. *Injury epidemiology*, *7*(1), 1-5.

Bertrand, M., & Morse, A. (2011). Information disclosure, cognitive biases, and payday borrowing. *The Journal of Finance*, *66*(6), 1865-1893.

Bertrand, M., Bombardini, M., & Trebbi, F. (2014). Is it whom you know or what you know? An empirical assessment of the lobbying process. *American Economic Review*, *104*(12), 3885-3920.

Bertrand, M., Mullainathan, S., & Shafir, E. (2006). Behavioral economics and marketing in aid of decision making among the poor. *Journal of Public Policy & Marketing*, *25*(1), 8-23.

Beugelsdijk, S., & Van Schaik, T. (2005). Social capital and growth in European regions: an empirical test. *European Journal of Political Economy*, *21*(2), 301-324.

Beugelsdijk, S., De Groot, H. L., & Van Schaik, A. B. (2004). Trust and economic growth: a robustness analysis. *Oxford economic papers*, *56*(1), 118-134.

Bikhchandani, S., & Sharma, S. (2000). Herd behavior in financial markets: A review. IMF Staff Papers, *47*(3), 279-310.

Binmore, K. (2005). Economic man–or straw man?. *Behavioral and Brain Sciences*, *28*(6), 817-818.

Birch, J., 2017. *The philosophy of social evolution*. Oxford University Press.

Bird, G., & Viding, E. (2014). The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience & Biobehavioral Reviews*, *47*, 520-532.

Bird, J. & Layzell, P. (2002). The evolved radio and its implications for modelling the evolution of novel sensors. In *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600)*, 2, 1836-1841).

Bix, B. (1991). HLA hart and the" open texture" of language. *Law & Phil.*, 10, 51.

Bjørnskov, C. (2007). Determinants of generalized trust: A cross-country comparison. *Public choice*, *130*(1-2), 1-21.

Bjørnskov, C. (2012). How does social trust affect economic growth?. *Southern Economic Journal*, *78*(4), 1346-1368.

Bjørnskov, C. (2017). Social trust and economic growth. Forthcoming, *Oxford Handbook of Social and Political Trust*, (Eric M. Uslaner, ed.) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2906280

Bjørnskov, C., & Méon, P. G. (2015). The productivity of trust. *World development*, *70*, 317-331.

Black, D. (1948). On the rationale of group decision-making. *Journal of political economy*, *56*(1), 23-34.

Blais, A. (2000). To vote or not to vote?: The merits and limits of rational choice theory. University of Pittsburgh Pre.

Blau, B. M., Brough, T. J., & Thomas, D. W. (2013). Corporate lobbying, political connections, and the bailout of banks. *Journal of Banking & Finance*, *37*(8), 3007-3017.

Blonigen, D. M., Hicks, B. M., Krueger, R. F., Patrick, C. J., & Iacono, W. G. (2005). Psychopathic personality traits: Heritability and genetic overlap with internalizing and externalizing psychopathology. *Psychological medicine*, *35*(5), 637-648.

Bloom, N., Jones, C.I., Van Reenen, J. & Webb, M. (2017). *Are Ideas Getting Harder to Find?* (No. w23782). National Bureau of Economic Research.

Bollyky, T. J., Hulland, E. N., Barber, R. M., Collins, J. K., Kiernan, S., Moses, M., ... & Dieleman, J. L. (2022). Pandemic preparedness and COVID-19: an exploratory analysis of infection and fatality rates, and contextual factors associated with preparedness in 177 countries, from Jan 1, 2020, to Sept 30, 2021. *The Lancet*, *399*(10334), 1489-1512.

Bologna, P., Miglietta, A., & Segura, A. (2018). Contagion in the CoCos market? A case study of two stress events. *A Case Study of Two Stress Events (November 6, 2018). Bank of Italy Temi di Discussione (Working Paper) No*, *1201*.

Bolton, G.E. & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, *91*(1), 166-193.

Bolton, P., Mehran, H., & Shapiro, J. (2015). Executive compensation and risk taking. *Review of Finance*, *19*(6), 2139-2181.

Bolton, P., Scheinkman, J., & Xiong, W. (2006). Executive compensation and short-termist behaviour in speculative markets. *The Review of Economic Studies*, *73*(3), 577-610.

Bonica, A., & Sen, M. (2017). The politics of selecting the bench from the bar: the legal profession and partisan incentives to introduce ideology into judicial selection. *The Journal of Law and Economics*, *60*(4), 559-595.

Borah, N., James, H. L., & Park, J. C. (2020). Does CEO inside debt compensation benefit both shareholders and debtholders?. *Review of Quantitative Finance and Accounting*, *54*, 159-203.

Bornschier, S. (2018). Globalization, cleavages, and the radical right. *The Oxford handbook of the radical right*, 212-238.

Bornschier, S., & Kriesi, H. (2013). The populist right, the working class, and the changing face of class politics. *Class politics and the radical right*, *16*, 10.

Boucher, D.H. ed. (1985). *The biology of mutualism: ecology and evolution*. Oxford University Press.

Bowen, W.G. & Rudenstine, N.L. (2014). *In pursuit of the PhD* (Vol. 79). Princeton University Press.

Bowles, S., Boyd, R., Mathew, S., Richerson, P. J., & Guala, F. (2012). The punishment that sustains cooperation is often coordinated and costly. *Behavioral and Brain Sciences*, *35*(1), 20.

Boyd, R., Gintis, H., & Bowles, S. (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science*, *328*(5978), 617-620.

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences*, *100*(6), 3531-3535.

Boyd, R., & Richerson, P. J. (1982). Cultural transmission and the evolution of cooperative behavior. *Human ecology*, *10*, 325-351.

Boyd, R., & Richerson, P. J. (1988). *Culture and the evolutionary process*. University of Chicago press.

Boyd, R., & Richerson, P. J. (1989). The evolution of indirect reciprocity. *Social networks*, *11*(3), 213-236.

Bradach, J. L., & Eccles, R. G. (1989). Price, authority, and trust: From ideal types to plural forms. *Annual review of sociology*, *15*(1), 97-118.

Braley, A., Lenz, G. S., Adjodah, D., Rahnama, H., & Pentland, A. (2023). Why voters who value democracy participate in democratic backsliding. *Nature Human Behaviour*, 1-12.

Braman, A. C., & Lambert, A. J. (2001). Punishing individuals for their infirmities: Effects of personal responsibility, just-world beliefs, and in-group/out-group status. *Journal of Applied Social Psychology*, *31*(5), 1096-1109.

Brandeis, L. D. (1915). Living law. *Illinois Law Review*, 10, 461-470.

Brax, H., Fadlallah, R., Al-Khaled, L., Kahale, L. A., Nas, H., El-Jardali, F., & Akl, E. A. (2017). Association between physicians' interaction with pharmaceutical companies and their clinical practices: a systematic review and meta-analysis. *PloS one*, *12*(4), e0175493.

Bremermann, H.J. & Pickering, J. (1983). A game-theoretical model of parasite virulence. *Journal of Theoretical Biology*, *100*(3), 411-426.

Brennan, G., & Lomasky, L. (1997). *Democracy and decision: The pure theory of electoral preference*. Cambridge University Press.

Brewer, M. B. (2008). Depersonalized trust and ingroup cooperation. In *Rationality and social responsibility,* 215-232. Psychology Press.

Briggs, R. A. (2019). Normative Theories of Rational Choice: Expected Utility. In The Stanford Encyclopedia of Philosophy (E. N. Zalta, Ed.). https://plato.stanford.edu/archives/fall2019/entries/rationality-normative-utility/

Brockhurst, M.A., Chapman, T., King, K.C., Mank, J.E., Paterson, S. & Hurst, G.D. (2014). Running with the Red Queen: the role of biotic conflicts in evolution. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1797), p.20141382.

Bronars, S. G., & Lott, Jr, J. R. (1997). Do campaign donations alter how a politician votes? Or, do donors support candidates who value the same things that they do?. *The Journal of Law and Economics*, *40*(2), 317-350.

Brucks, D., & von Bayern, A. M. (2020). Parrots voluntarily help each other to obtain food rewards. *Current Biology*, *30*(2), 292-297.

Brulle, R. J. (2018). The climate lobby: a sectoral analysis of lobbying spending on climate change in the USA, 2000 to 2016. *Climatic change*, *149*(3-4), 289-303.

Bshary, R. & Grutter, A. S. (2002). Asymmetric cheating opportunities and partner control in a cleaner fish mutualism. *Animal Behaviour*, *63*(3), 547-555.

Bshary, R. & Schäffer, D. (2002). Choosy reef fish select cleaner fish that provide high-quality service. *Animal Behaviour*, *63*(3), 557-564.

Bubandt, N., & Willerslev, R. (2015). The dark side of empathy: Mimesis, deception, and the magic of alterity. *Comparative Studies in Society and History*, *57*(1), 5-34.

Buchak, G., Matvos, G., Piskorski, T., & Seru, A. (2018). Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of financial economics*, *130*(3), 453-483.

Buchanan, J. M. (1977). 'Law and the invisible hand'. In J. M. Buchanan *Freedom in Constitutional Contract: Perspectives of a Political Economist*. Texas A & M University Press.

Buchanan, J. M., & Tullock, G. (1965). *The calculus of consent: Logical foundations of constitutional democracy* (Vol. 100). University of Michigan press.

Buckels, E. E., Jones, D. N., & Paulhus, D. L. (2013). Behavioral confirmation of everyday sadism. *Psychological science*, *24*(11), 2201-2209.

Bull JJ, Rice WR. 1991 Distinguishing mechanisms for the evolution of co-operation. *J. Theor. Biol.* 149, 63 – 74.

Bullock, A. (2019). *Hitler and Stalin: parallel lives*. Vintage. (Original work published 1991)

Burkart, J. M., Fehr, E., Efferson, C., & van Schaik, C. P. (2007). Other-regarding preferences in a non-human primate: Common marmosets provision food altruistically. *Proceedings of the National Academy of Sciences*, *104*(50), 19762-19766.

Bushee, B.J. (1998). The influence of institutional investors on myopic R&D investment behavior. *Accounting review*, 305-333.

Butler, J. (1726). *Fifteen sermons preached at the Rolls Chapel*. Botham.

Buttrick, N. R., Heintzelman, S. J., & Oishi, S. (2017). Inequality and well-being. *Current opinion in psychology*, *18*, 15-20.

Calvin, W. H. (1987). The brain as a Darwin Machine. *Nature*, 330, 33-34.

Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational research*, *64*(3), 363-423.

Cameron, R. (2022). Infinite Regress Arguments. *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Campbell, D.T. (1974). Unjustified variation and selective retention in scientific discovery. *Studies in the philosophy of biology: Reduction and related problems*, 139-161.

Cannane, S., & Taylor, K. (2019, December 5). In this sports betting company, the winners are called 'problem customers'. *Australian Broadcasting Corporation*. https://www.abc.net.au/news/2019-12-05/bet365-whistleblower-says-winners-given-delays/11768486

Capoccia, G. (2005). *Defending democracy: Reactions to extremism in interwar Europe*. JHU Press.

Caprettini, B., Casaburi, L., & Venturini, M. (2019). The Electoral Impact of Wealth Redistribution. Evidence from the Italian Land Reform. *Evidence from the Italian Land Reform (February 15, 2019)*.

Carlin, R. E., & Love, G. J. (2013). The politics of interpersonal trust and reciprocity: An experimental approach. *Political Behavior*, *35*, 43-63.

Carreras, M., & İrepoğlu, Y. (2013). Trust in elections, vote buying, and turnout in Latin America. *Electoral Studies*, *32*(4), 609-619.

Carroll, L. (1897) *Through the looking glass: And what Alice found there*. Henry Altemus Company

Carstens, A. (2021). Non-bank financial sector: systemic regulation needed. Bank of International Settlements. https://www.bis.org/publ/qtrpdf/r_qt2112_foreword.htm

Caselli, F., & Cunningham, T. (2009). Leader behaviour and the natural resource curse. *Oxford economic papers*, *61*(4), 628-650.

Cassell, C. A., Huang, S. X., Sanchez, J. M., & Stuart, M. D. (2012). Seeking safety: The relation between CEO inside debt holdings and the riskiness of firm investment and financial policies. *Journal of Financial Economics*, *103*(3), 588-610.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach* (No. 16). Princeton University Press.

Cecchetti, S. G. & Kharroubi, E. (2012). Reassessing the impact of finance on growth. BIS Working Paper No. 381, Available at SSRN: https://ssrn.com/abstract=2117753

Centers for Medicare & Medicaid Services. (2021). 2019 General Payment Data. U.S. Government. https://openpaymentsdata.cms.gov/dataset/4e54dd6c-30f8-4f86-86a7-3c109a89528e

Cesarini, D., Dawes, C. T., Fowler, J. H., Johannesson, M., Lichtenstein, P., & Wallace, B. (2008). Heritability of cooperative behavior in the trust game. *Proceedings of the National Academy of sciences*, *105*(10), 3721-3726.

Cesarini, D., Dawes, C. T., Johannesson, M., Lichtenstein, P., & Wallace, B. (2009). Genetic variation in preferences for giving and risk taking. *The Quarterly Journal of Economics*, *124*(2), 809-842.

Chamberlin, E. (1933). *Theory of Monopolistic Competition*. Cambridge, MA: Harvard University Press.

Chami, R., & Fullenkamp, C. (2002). Trust and efficiency. *Journal of banking & finance*, *26*(9), 1785-1809.

Chan, S., & van Wijnbergen, S. (2017). Coco design, risk shifting incentives and financial fragility. *ECMI Papers*, *12166*.

Chan, W.Y., Lorke, D.E., Tiu, S.C. & Yew, D.T. (2002). Proliferation and apoptosis in the developing human neocortex. *The Anatomical Record: An Official Publication of the American Association of Anatomists*, *267*(4), 261-276.

Chanley, V. A., Rudolph, T. J., & Rahn, W. M. (2000). The origins and consequences of public trust in government: A time series analysis. *Public opinion quarterly*, *64*(3), 239-256.

Charness, G., & Gneezy, U. (2008). What's in a name? Anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization*, *68*(1), 29-35.

Chen, C. R., Steiner, T. L., & Whyte, A. M. (2006). Does stock option-based executive compensation induce risk-taking? An analysis of the banking industry. *Journal of Banking & Finance*, *30*(3), 915-945.

Chen, H., Parsley, D., & Yang, Y. W. (2015). Corporate lobbying and firm performance. *Journal of Business Finance & Accounting*, *42*(3-4), 444-481.

Chen, Y., & Li, S. X. (2009). Group identity and social preferences. *American Economic Review*, *99*(1), 431-457.

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.*, *107*, 1753.

Chi, S., Huang, S. X., & Sanchez, J. M. (2017). CEO inside debt incentives and corporate tax sheltering. *Journal of Accounting Research*, *55*(4), 837-876.

Chiles, T. H., & McMackin, J. F. (1996). Integrating variable risk preferences, trust, and transaction cost economics. *Academy of management review*, *21*(1), 73-99.

Chittka, L. & Raine, N.E. (2006). Recognition of flowers by pollinators. *Current opinion in plant biology*, *9*(4), 428-435.

Chittka, L. & Schürkens, S. (2001). Successful invasion of a floral market. *Nature*, *411*(6838), 653-653.

Choi, J. J., Laibson, D., & Madrian, B. C. (2010). Why does the law of one price fail? An experiment on index mutual funds. *The Review of Financial Studies*, *23*(4), 1405-1432.

Choy, H., Lin, J., & Officer, M. S. (2014). Does freezing a defined benefit pension plan affect firm risk?. *Journal of Accounting and Economics*, *57*(1), 1-21.

Chrétien, J. P. (Ed.). (1995). *Rwanda: Les médias du génocide*. Karthala Editions.

Christensen, S.R. & Shlomchik, M.J. (2007). Regulation of lupus-related autoantibody production and clinical disease by Toll-like receptors. *Seminars in immunology,* 19(1), 11-23).

Christiano, P.F., Leike, J., Brown, T., Martic, M., Legg, S. & Amodei, D. (2017). Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30.

Cikara, M., Bruneau, E., Van Bavel, J. J., & Saxe, R. (2014). Their pain gives us pleasure: How intergroup dynamics shape empathic failures and counter-empathic responses. *Journal of experimental social psychology*, *55*, 110-125.

Cinyabuguma M., Page T. & Putterman L. (2006). Can second-order punishment deter perverse punishment? *Exp Econ* 9(3), 265–279.

Claessens, S., Feijen, E., & Laeven, L. (2008). Political connections and preferential access to finance: The role of campaign contributions. *Journal of financial economics*, *88*(3), 554-580.

Clancy, M. S. (2018). Combinations of technology in US patents, 1926–2009: a weakening base for future innovation?. *Economics of Innovation and new Technology*, *27*(8), 770-785.

Clark, M. A., Moran, N. A., Baumann, P., & Wernegreen, J. J. (2000). Cospeciation between bacterial endosymbionts (Buchnera) and a recent radiation of aphids (Uroleucon) and pitfalls of testing for phylogenetic congruence. *Evolution*, *54*(2), 517-525.

Shannon, C. E. (1950). XXII. Programming a computer for playing chess. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *41*(314), 256-275.

Clayton, J. (1997). Are housing price cycles driven by irrational expectations?. *The Journal of Real Estate Finance and Economics*, *14*, 341-363.

Cleckley, H. (1976). The mask of sanity (5th ed.). St Louis, MO: Mosby

Clinton, J. D., Lapinski, J. S., Lentz, S., & Pettigrew, S. (2020). Trumped by Trump? Public support for vote by mail voting in response to the COVID-19 pandemic. *Public Support for Vote By Mail Voting in Response to the COVID-19 Pandemic (June 16, 2020)*.

Coase, R. H. (2000). The problem of social cost. In C. Gopalakrishnan (Ed.), *Classic Papers in Natural Resource Economics* (87-137). Palgrave Macmillan UK. (Original work published 1960)

Coate, S., & Conlin, M. (2004). A group rule—utilitarian approach to voter turnout: Theory and evidence. *American Economic Review*, *94*(5), 1476-1504.

Coate, S., Conlin, M., & Moro, A. (2008). The performance of pivotal-voter models in small-scale elections: Evidence from Texas liquor referenda. Journal of Public Economics, 92(3-4), 582-596.

Cohen, J. (1999). Trust, voluntary association and workable democracy: The contemporary American discourse of civil society. In M. Warren (Ed.), *Democracy and Trust*. Cambridge, UK: Cambridge University Press

Coleman, J. S. (1988). Social capital in the creation of human capital. *American journal of sociology*, *94*, S95-S120.

Collier, P. (2001). Implications of ethnic diversity. *Economic policy*, *16*(32), 128-166.

Collier, P. (2015). The technical challenges of the Large Hadron Collider. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2032), p.2014–44.

Collier, P., & Hoeffler, A. (1998). On economic causes of civil war. *Oxford economic papers*, *50*(4), 563-573.

Cook, S. J., Jarrell, T. A., Brittin, C. A., Wang, Y., Bloniarz, A. E., Yakovlev, M. A., ... & Emmons, S. W. (2019). Whole-animal connectomes of both Caenorhabditis elegans sexes. *Nature*, *571*(7763), 63-71.

Cooper, M. J., Gulen, H., & Ovtchinnikov, A. V. (2010). Corporate political contributions and stock returns. *The Journal of Finance*, *65*(2), 687-724.

Cooper, S. & Helmstetter, C.E. (1968). Chromosome replication and the division cycle of Escherichia coli Br. *Journal of molecular biology*, *31*(3), 519-540.

Corriher, B. (2012, August 13). Big business taking over state supreme courts. Center for American Progress. https://www.americanprogress.org/article/big-business-taking-over-state-supreme-courts/

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. *Handbook of emotions*, *2*(2), 91-115.

Costas-Pérez, E., Solé-Ollé, A., & Sorribas-Navarro, P. (2012). Corruption scandals, voter information, and accountability. *European journal of political economy*, *28*(4), 469-484.

Courchesne, E., Mouton, P.R., Calhoun, M.E., Semendeferi, K., Ahrens-Barbeau, C., Hallet, M.J., Barnes, C.C. and Pierce, K. (2011). Neuron number and size in prefrontal cortex of children with autism. *Jama*, *306*(18), 2001-2010.

Cowen, T. (2011). *The great stagnation: How America ate all the low-hanging fruit of modern history, got sick, and will (eventually) feel better* [eBook edition]. Dutton.

Cox, G. W. (1997). *Making votes count: strategic coordination in the world's electoral systems*. Cambridge University Press.

Cox, J. C., & Deck, C. A. (2005). On the nature of reciprocal motives. *Economic Inquiry*, *43*(3), 623-635.

Creswell's Executor v. Walker. 37 Ala. 233 (1861).

Crockett, M. J., & Lockwood, P. L. (2018). Extraordinary altruism and transcending the self. *Trends in cognitive sciences*, *22*(12), 1071-1073.

Cronin, K. A., Schroeder, K. K., & Snowdon, C. T. (2010). Prosocial behaviour emerges independent of reciprocity in cottontop tamarins. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1701), 3845-3851.

Crosby, F. (1976). A model of egoistical relative deprivation. *Psychological review*, *83*(2), 85.

Crouzet, N., & Eberly, J. (2021). Intangibles, markups, and the measurement of productivity growth. *Journal of Monetary Economics*, *124*, S92-S109.

Cruess, S. R., & Cruess, R. L. (2018). The development of professional identity. *Understanding Medical Education: Evidence, Theory, and Practice*, 239-254.

Culkin, J. M. (1967, March 18). A schoolman's guide to Marshall McLuhan. *The Saturday Review*.

Cunliffe, J. (2017, February 9). Market-based finance: a macroprudential view. Speech at Asset Management Derivatives Forum, Dana Point, California.

Currie, C. R., Scott, J. A., Summerbell, R. C., & Malloch, D. (1999). Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. *Nature*, *398*(6729), 701-704.

Currie, C. R., Wong, B., Stuart, A. E., Schultz, T. R., Rehner, S. A., Mueller, U. G., ... & Straus, N. A. (2003). Ancient tripartite coevolution in the attine ant-microbe symbiosis. *Science*, *299*(5605), 386-388.

Czech, R., Gual-Ricart, B., Lillis, J., & Worlidge, J. (2021). The Role of Non-bank Financial Intermediaries in the 'Dash for Cash' in Sterling Markets. Bank of England. https://elischolar.library.yale.edu/cgi/viewcontent.cgi?article=12295&context=ypfs-documents

Dahl, R. A. (1989). *Democracy and its critics*. Yale University Press.

Dal Bó, E. (2006). Regulatory capture: A review. *Oxford review of economic policy*, *22*(2), 203-225.

Dalton, R. J. (1998). Political Support in Advanced Industrial Democracie. UC Irvine: Center for the Study of Democracy. https://escholarship.org/uc/item/8281d6wt

Dandekar, A. A., Chugani, S., & Greenberg, E. P. (2012). Bacterial quorum sensing and metabolic incentives to cooperate. *Science*, *338*(6104), 264-266.

Dang, V. A., & Phan, H. V. (2016). CEO inside debt and corporate debt maturity structure. *Journal of Banking & Finance*, *70*, 38-54.

Darwin, C. (1839). Excursion to St. Fe, etc. In *The voyage of the Beagle* (Chapter VII). London: Henry Colburn.

Darwin, C. (1859). *On the origin of species*. London: J. Murray.

Darwin, C. (1862). In Burkhardt, F., & Smith, S. (1987). *The correspondence of Charles Darwin. Vol. 10, 1862*. Cambridge University Press.

Darwin, C. (1871). *The descent of man and selection in relation to sex II*. John Murray.

Dawel, A., O'Kearney, R., McKone, E., & Palermo, R. (2012). Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience & Biobehavioral Reviews*, *36*(10), 2288-2304.

Dawkins, R. (1988). The evolution of evolvability. In C. G. Langton (Ed.), *Artificial life: The proceedings of an interdisciplinary workshop on the synthesis and simulation of living systems* (pp. 201–220). Redwood City, CA: Addison-Wesley Publishing Co.

Dawkins, R. (2004). Extended phenotype–but not too extended. A reply to Laland, Turner and Jablonka. *Biology and Philosophy*, *19*, 377-396.

Dawkins, R. (2016) *The Extended Phenotype: The Long Reach of the Gene*. Oxford University Press.

Dawsey, J. & Paletta, D. (2018, November 15). Trump demands action to reduce deficit, pushes new deficit spending. *Washington Post*. https://www.washingtonpost.com/politics/trump-demands-action-to-reduce-deficit-and-pushes-new-deficit-spending/2018/11/25/86bdf120-d88c-11e8-9559-712cbf726d1c_story.html

De Bromhead, A., Eichengreen, B., & O'Rourke, K. H. (2013). Political extremism in the 1920s and 1930s: Do German lessons generalize?. *The Journal of Economic History*, *73*(2), 371-406.

De Figueiredo, J. M., & Silverman, B. S. (2006). Academic earmarks and the returns to lobbying. *The Journal of Law and Economics*, *49*(2), 597-625.

De Loecker, J., Eeckhout, J., & Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, *135*(2), 561-644.

de Man, T. J., Stajich, J. E., Kubicek, C. P., Teiling, C., Chenthamara, K., Atanasova, L., ... & Gerardo, N. M. (2016). Small genome of the fungus Escovopsis weberi, a specialized disease agent of ant agriculture. *Proceedings of the National Academy of Sciences*, *113*(13), 3567-3572.

de Mazancourt, C. & Schwartz, M. W. (2010). A resource ratio theory of cooperation. *Ecology letters*, *13*(3), 349-359.

de Mazancourt, C. & Schwartz, M. W. (2012). Starve a competitor: evolution of luxury consumption as a competitive strategy. *Theoretical Ecology*, *5*, 37-49.

de Mazancourt, C., Loreau, M. & Dieckmann, U.L.F. (2005). Understanding mutualism when there is adaptation to the partner. *Journal of Ecology*, *93*(2), 305-314.

Debs, A. (2007a). The wheel of fortune: Agency problems in dictatorships. *University of Rochester mimeo*.

Debs, A. (2007b). *Political Strength and Economic Efficiency in a Multi-Agent State*. Mimeo. University of Rochester.

Decety, J., Chen, C., Harenski, C., & Kiehl, K. A. (2013). An fMRI study of affective perspective taking in individuals with psychopathy: imagining another in pain does not evoke empathy. *Frontiers in human neuroscience*, *7*, 489.

Deidda, L., & Fattouh, B. (2002). Non-linearity between finance and growth. *Economics Letters*, *74*(3), 339-345.

Dekkers, M.P., Nikoletopoulou, V. & Barde, Y.A. (2013). Death of developing neurons: New insights and implications for connectivity. *Journal of Cell Biology*, *203*(3), 385-393.

Delhey, J., & Newton, K. (2005). Predicting cross-national levels of social trust: global pattern or Nordic exceptionalism?. *European sociological review*, *21*(4), 311-327.

Demetriades, P. O., & Rewilak, J. M. (2020). Recovering the finance-growth nexus. *Economics Letters*, *196*, 109563.

Demetriades, P. O., Rewilak, J. M., & Rousseau, P. L. (2023). Finance, Growth, and Fragility. *Journal of Financial Services Research*, 1-21.

Demetriades, P. O., & Rousseau, P. L. (2016). The changing face of financial development. *Economics Letters*, *141*, 87-90.

Demsetz, H. (1967). Toward a theory of property rights. *American Economic Review*, *59*, 817-831.

Deng, K., He, J., Kong, D., & Zhang, J. (2019). Does inside debt alleviate banks' risk taking? Evidence from a quasi-natural experiment in the Chinese banking industry. *Emerging Markets Review*, *40*, 100622.

Dennett, D. (1995). *Darwin's Dangerous Idea*. Simon & Schuster.

Denzau, A.T. & North, D.C. (1994). Shared mental models: ideologies and institutions. *Kyklos,* 47(1), 3-31.

Deppmann, C.D., Mihalas, S., Sharma, N., Lonze, B.E., Niebur, E. & Ginty, D.D., (2008). A model for neuronal competition during development. *Science*, *320*(5874), 369-373.

Derfus, P. J., Maggitti, P. G., Grimm, C.M., & Smith, K.G. (2008). The Red Queen effect: Competitive actions and firm performance. *Academy of Management Journal*, 51(1), 61-80.

Desmet, K., Ortuño-Ortín, I., & Wacziarg, R. (2012). The political economy of linguistic cleavages. *Journal of development Economics*, *97*(2), 322-338.

Devenow, A., & Welch, I. (1996). Rational herding in financial economics. *European economic review*, *40*(3-5), 603-615.

Dexheimer, E. & Bureau, A. (2018, December 21). Lazy judges? Newcomers? Career criminals? Hays struggles to decipher overflowing jail. *Houston Chronicle*. https://www.houstonchronicle.com/politics/texas/article/Lazy-judges-Newcomers-Career-criminals-Hays-13483929.php

Dezfouli, A., Nock, R. & Dayan, P. (2020). Adversarial vulnerabilities of human decision-making. *Proceedings of the National Academy of Sciences*, *117*(46), 29221-29228.

Dhole, S., Manchiraju, H., & Suk, I. (2016). CEO inside debt and earnings management. *Journal of Accounting, Auditing & Finance*, *31*(4), 515-550.

Di Tella, R., Dubra, J., & Lagomarsino, A. (2021). Meet the oligarchs: Business legitimacy and taxation at the top. *The Journal of Law and Economics*, *64*(4), 651-674.

Diaz-Cayeros, A., Magaloni, B., & Weingast, B. (2000). Democratization and the economy in Mexico: Equilibrium (PRI) hegemony and its demise. Unpublished paper, Stanford University and UCLA.

Dieckmann, U. & Metz, J.A. (2006). Surprising evolutionary predictions from enhanced ecological realism. *Theoretical Population Biology*, *69*(3), 263-281.

Diermeier, D., Keane, M., & Merlo, A. (2005). A political economy model of congressional careers. *American Economic Review*, *95*(1), 347-373.

Dimick, M., Rueda, D. & Stegmueller, D. (2016). The altruistic rich? Inequality and other-regarding preferences for redistribution. *Quarterly Journal of Political Science*, *11*(4), 385-439.

Dincer, O. C., & Uslaner, E. M. (2010). Trust and growth. *Public Choice*, *142*, 59-67.

Djankov, S., & Murrell, P. (2002). Enterprise restructuring in transition: A quantitative survey. *Journal of economic literature*, *40*(3), 739-792.

Dorn, D., Fischer, J. A., Kirchgässner, G., & Sousa-Poza, A. (2007). Is it culture or democracy? The impact of democracy and culture on happiness. *Social Indicators Research*, *82*, 505-526.

Dornbusch, R., & Edwards, S. (1991). The macroeconomics of populism. In R. Dornbusch & S. Edwards (Eds.), *The macroeconomics of populism in Latin America*. University of Chicago Press.

Downs, A. (1957). *An economic theory of democracy*. Harper and Row.

Drechsler, I., Drechsel, T., Marques-Ibanez, D., & Schnabl, P. (2016). Who borrows from the lender of last resort?. *The Journal of Finance*, *71*(5), 1933-1974.

Drummond, A., & Sauer, J. D. (2018). Video game loot boxes are psychologically akin to gambling. *Nature human behaviour*, *2*(8), 530-532.

Drutman, L. J. (2010). The business of America is lobbying: The expansion of corporate political activity and the future of American pluralism. University of California, Berkeley.

Duarte, F., & Hastings, J. S. (2012). *Fettered consumers and sophisticated firms: evidence from Mexico's privatized social security market* (No. w18582). National Bureau of Economic Research.

Duchin, R., & Sosyura, D. (2012). The politics of government investment. *Journal of Financial Economics*, *106*(1), 24-48.

Ductor, L., & Grechyna, D. (2015). Financial development, real sector, and economic growth. *International Review of Economics & Finance*, *37*, 393-405.

Dugatkin, L.A. (1988). Do guppies play tit for tat during predator inspection visits?. Behavioral Ecology and Sociobiology, 23(6), 395-399.

Durand, P.M., Sym, S. & Michod, R.E. (2016). Programmed cell death and complexity in microbial systems. *Current Biology*, *26*(13), R587-R593.

Dwoskin, E. & Timberg, C. (2018, April 23). How merchants use Facebook to flood Amazon with fake reviews. *Washington Post*. https://www.washingtonpost.com/business/economy/how-merchants-secretly-use-facebook-to-flood-amazon-with-fake-reviews/2018/04/23/5dad1e30-4392-11e8-8569-26fda6b404c7_story.html

Dyer, O. (2020). WHO retracts opioid guidelines after accepting that industry had an influence. *BMJ*, 368. https://doi.org/10.1136/bmj.m105

Easterly, W., & Levine, R. (1997). Africa's growth tragedy: policies and ethnic divisions. *The quarterly journal of economics*, 1203-1250.

Eckel, C. C., & Grossman, P. J. (1996). Altruism in anonymous dictator games. *Games and economic behavior*, *16*(2), 181-191.

Edgeworth, F.Y., 1881. Mathematical psychics: An essay on the application of mathematics to the moral sciences (Vol. 10). CK Paul.

Edlin, A., Gelman, A., & Kaplan, N. (2007). Voting as a rational choice: Why and how people vote to improve the well-being of others. *Rationality and Society*, 19(3), 293-314.

Edmans, A., & Liu, Q. (2011). Inside debt. *Review of finance*, *15*(1), 75-102.

Efferson, C., Lalive, R., & Fehr, E. (2008). The coevolution of cultural groups and ingroup favoritism. *Science*, *321*(5897), 1844-1849.

Eger, M. A., & Breznau, N. (2017). Immigration and the welfare state: A cross-regional analysis of European welfare attitudes. *International Journal of Comparative Sociology*, *58*(5), 440-463.

Eggers, A. C. (2014). Partisanship and electoral accountability: Evidence from the UK expenses scandal. *Quarterly Journal of Political Science*, 9(4), 441-472.

Eggers, A. C., & Hainmueller, J. (2009). MPs for sale? Returns to office in postwar British politics. *American Political Science Review*, *103*(4), 513-533.

Egorov, G., & Sonin, K. (2011). Dictators and their viziers: Endogenizing the loyalty–competence trade-off. *Journal of the European Economic Association*, *9*(5), 903-930.

Eichholtz, P., Huisman, R., & Zwinkels, R. C. (2015). Fundamentals or trends? A long-term perspective on house prices. *Applied Economics*, *47*(10), 1050-1059.

Eisenmann, T. (2021). Why startups fail: A new roadmap for entrepreneurial success. *Currency*.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, *17*(2), 124.

Ekman, P. (1992). Facial expressions of emotion: New findings, new questions. *Psychological Science*.

Elliott, C. (2014). Relationships between physicians and Pharma: Why physicians should not accept money from the pharmaceutical industry. *Neurology: Clinical Practice*, *4*(2), 164-167.

Borel, E. (1913). Static mechanics and irreversibility. *J. Phys. Theor. Appl.* , *3* (1), 189-196.

Engel, C. (2011). Dictator games: A meta study. *Experimental economics*, *14*, 583-610.

Engel, S. T., & Jackson, D. J. (1998). Wielding the stick instead of the carrot: Labor PAC punishment of pro-NAFTA Democrats. *Political Research Quarterly*, *51*(3), 813-828.

Engel, S. T., & Jackson, D. J. (1998). Wielding the stick instead of the carrot: Labor PAC punishment of pro-NAFTA Democrats. *Political Research Quarterly*, *51*(3), 813-828.

Eres, R., & Molenberghs, P. (2013). The influence of group membership on the neural correlates involved in empathy. *Frontiers in human neuroscience*, *7*, 176.

Erikson, R. S., & Palfrey, T. R. (2000). Equilibria in campaign spending games: Theory and data. *American Political Science Review*, *94*(3), 595-609.

Eshel, I. (1996). On the changing concept of evolutionary population stability as a reflection of a changing point of view in the quantitative theory of evolution. *Journal of mathematical biology*, *34*, 485-510.

Esteban, J. M., & Ray, D. (1994). On the measurement of polarization. *Econometrica: Journal of the Econometric Society*, 819-851.

Esteban, J., & Ray, D. (2008). Polarization, fractionalization and conflict. *Journal of peace Research*, *45*(2), 163-182.

Esteban, J., & Ray, D. (2011). Linking conflict to inequality and polarization. *American Economic Review*, *101*(4), 1345-1374.

Esteva, A. & Topol, E. (2019). Can skin cancer diagnosis be transformed by AI?. *The Lancet*, *394*(10211), 1795.

Esteves-Sorenson, C., & Broce, R. (2022). Do monetary incentives undermine performance on intrinsically enjoyable tasks? A field test. *The Review of Economics and Statistics*, *104*(1), 67-84.

Esty, B. C. (1998). The impact of contingent liability on commercial bank risk taking. *Journal of Financial Economics*, *47*(2), 189-218.

Faccio, M. (2006). Politically connected firms. *American economic review*, *96*(1), 369-386.

Faccio, M., & Parsley, D. C. (2009). Sudden deaths: Taking stock of geographic ties. *Journal of Financial and Quantitative Analysis*, *44*(3), 683-718.

Faccio, M., Masulis, R. W., & McConnell, J. J. (2006). Political connections and corporate bailouts. *The journal of Finance*, *61*(6), 2597-2635.

Fafchamps, M., & Labonne, J. (2017). Do politicians' relatives Get better jobs? evidence from municipal elections. *The Journal of Law, Economics, and Organization*, *33*(2), 268-300.

Fagin-Jones, S., & Midlarsky, E. (2007). Courageous altruism: Personal and situational correlates of rescue during the Holocaust. *The Journal of Positive Psychology*, *2*(2), 136-147.

Fahlenbrach, R., & Stulz, R. M. (2011). Bank CEO incentives and the credit crisis. *Journal of financial economics*, *99*(1), 11-26.

Falato, A., & Scharfstein, D. (2016). *The stock market and bank risk-taking* (No. w22689). National Bureau of Economic Research.

Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*. 25(2), 383.

Farnsworth, W. (1999). Do Parties to Nuisance Cases Bargain after Judgment? A Glimpse inside the Cathedral. *The University of Chicago Law Review*. 66(2), 373–436.

Fatouh, M., Neamţu, I., & van Wijnbergen, S. (2021). Risk-taking and uncertainty: do contingent convertible (CoCo) bonds increase the risk appetite of banks?. Bank of England Working Paper No. 938.

Financial Crisis Inquiry Commission. (2011). *The financial crisis inquiry report: The final report of the National Commission on the causes of the financial and economic crisis in the United States including dissenting views*. Financial Crisis Inquiry Commission, Washington.

Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *American political science review*, *97*(1), 75-90.

Feddersen, T., Gailmard, S., & Sandroni, A. (2009). Moral bias in large elections: Theory and experimental evidence. *American Political Science Review*, *103*(2), 175-192.

Feeny, D., Berkes, F., McCay, B. J., & Acheson, J. M. (1990). The tragedy of the commons: twenty-two years later. *Human ecology*, *18*, 1-19.

Fehr, E., & Falk, A. (1999). Wage rigidity in a competitive incomplete contract market. *Journal of political Economy*, *107*(1), 106-134.

Fehr, E. & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785-791.

Fehr, E. & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and human behavior*, *25*(2), 63-87.

Fehr, E., Fischbacher, U., & Tougareva, E. (2002). Do High Stakes and Competition Undermine Fairness? Evidence from Russia. Institute for Empirical Research in Economics, University of Zurich.

Fehr, E., Kirchler, E., Weichbold, A., & Gächter, S. (1998). When social norms overpower competition: Gift exchange in experimental labor markets. *Journal of Labor economics*, *16*(2), 324-351.

Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. *The quarterly journal of economics*, *108*(2), 437-459.

Fehr, E., Klein, A., & Schmidt, K. M. (2001). Fairness, incentives and contractual incompleteness. *Available at SSRN 269389*.

Fehr, E. & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, *114*(3), 817-868.

Fellous, S. & Salvaudon, L. (2009). How can your parasites become your allies?. *Trends in parasitology*, *25*(2), 62-66.

Fellowes, M. C., & Wolf, P. J. (2004). Funding mechanisms and policy instruments: How business campaign contributions influence congressional votes. *Political Research Quarterly*, *57*(2), 315-324.

Fellowes, M. C., & Wolf, P. J. (2004). Funding mechanisms and policy instruments: How business campaign contributions influence congressional votes. *Political Research Quarterly*, *57*(2), 315-324.

Ferguson, T., & Voth, H. J. (2008). Betting on Hitler—the value of political connections in Nazi Germany. *The Quarterly Journal of Economics*, *123*(1), 101-137.

Ferrara, F. M. (2023). Why does import competition favor republicans? Localized trade shocks and cultural backlash in the US. *Review of International Political Economy*, *30*(2), 678-701.

Ferri, P., & Minsky, H. P. (1992). Market processes and thwarting systems. Structural Change and Economic Dynamics, 3(1), 79-91.

Ferriere, R. (2000). Adaptive responses to environmental threats: evolutionary suicide, insurance, and rescue. *Options Spring*, 12-16.

Filella, I., Primante, C., Llusia, J., Martín González, A.M., Seco, R., Farré-Armengol, G., Rodrigo, A., Bosch, J. & Penuelas, J. (2013). Floral advertisement scent in a changing plant-pollinators market. *Scientific reports*, *3*(1), 1-6.

Fiordelisi, F., Pennacchi, G., & Ricci, O. (2020). Are contingent convertibles going-concern capital?. *Journal of financial intermediation*, *43*, 100822.

Fiorina, M. (1990). Information and rationality in elections. In *Information and Democratic Processes* (ed. J. Ferejohn & J. Kuklinski). Urbana: University of Illinois Press.

Fisberg, M. & Machado, R. (2015). History of yogurt and current patterns of consumption. *Nutrition reviews*, *73*(1), 4-7.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics letters*, *71*(3), 397-404.

Fisher, R. W. (2013). Correcting Dodd–Frank to Actually End Too Big to Fail. *statement before the Committee on Financial Services, US House of Representatives*, *26*.

Fisman, R., Schulz, F., & Vig, V. (2014). The private returns to public office. *Journal of Political Economy*, *122*(4), 806-862.

Flaherty, T. M., & Rogowski, R. (2021). Rising inequality as a threat to the liberal international order. *International Organization*, *75*(2), 495-523.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic behavior*, *6*(3), 347-369.

Foster, P.L. (2007). Stress-induced mutagenesis in bacteria. *Critical reviews in biochemistry and molecular biology*, *42*(5), 373-397.

Fowler, J. H. (2006). Altruism and turnout. *The Journal of Politics*, *68*(3), 674-683.

Fowler, J. H., & Kam, C. D. (2007). Beyond the self: Social identity, altruism, and political participation. The Journal of Politics, 69(3), 813-827.

Frame, W.S. & White, L.J. (2004). Empirical studies of financial innovation: lots of talk, little action?. *Journal of economic literature*, *42*(1), 116-144.

Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. WW Norton & Co.

Frank, S.A. (1992). A kin selection model for the evolution of virulence. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *250*(1329), 195-197.

Frank, S.A., (1996). Models of parasite virulence. *The Quarterly review of biology*, *71*(1), 37-78.

Frank, S.A., (2003). Repression of competition and the evolution of cooperation. *Evolution*, *57*(4), 693-705.

Franklin, M., Niemi, R., & Whitten, G. (1994). The two faces of tactical voting. *British Journal of Political Science*, *24*(4), 549-557.

Franzen, A., & Pointner, S. (2013). The external validity of giving in the dictator game: A field experiment using the misdirected letter technique. *Experimental Economics*, *16*, 155-169.

Freifeld, K. (2016, November 4). Goldman Sachs, Paulson settle fraud lawsuit over Abacus. Reuters. https://www.reuters.com/article/us-goldman-sachs-abacus-lawsuit/goldman-sachs-paulson-settle-fraud-lawsuit-over-abacus-idUSKBN12Y2W4

Friedman, M. (1953). Essays in positive economics. University of Chicago Press.

Friedman, M. (1965) Economic Policy: Intentions vs. Results. Hoover Institution Archive, Milton Friedman Papers Box 36, folder 5, 15.

Friedman, M. (1990). The Adam Smith Address: The suicidal impulse of the business community. *Business Economics*, 25(1), 5-10.

Friedman, M. (2018) *The future of capitalism* [speech transcript]. Hoover Institution. https://www.hoover.org/research/future-capitalism (Original work published 1977, February 9)

Fruteau, C., Voelkl, B., van Damme, E. & Noë, R. (2009). Supply and demand determine the market value of food providers in wild vervet monkeys. *Proceedings of the National Academy of Sciences*, *106*(29), 12007-12012.

Fukuyama, F. (1989). The end of history? *The National Enquirer*, Summer edition. https://tr.uow.edu.au/uow/file/c1e8a018-b551-452a-94c460ukuyama92a4c5/1/fukuyama-1989-1-18.pdf

Fukuyama, F. (1992). *The end of history and the last man*. Free Press.

Fukuyama, F. (2000). Social capital and civil society. *Available at SSRN 879582*.

Fukuyama, F. (2001, October 11). The west has won. *The Guardian*. https://www.theguardian.com/world/2001/oct/11/afghanistan.terrorism30

Funke, M., Schularick, M., & Trebesch, C. (2016). Going to extremes: Politics after financial crises, 1870–2014. *European Economic Review*, *88*, 227-260.

Funke, M., Schularick, M., & Trebesch, C. (2020). Populist leaders and the economy.

Furman, J., & Orszag, P. R. (2018). Slower productivity and higher inequality: Are they related?. *Peterson Institute for International Economics Working Paper*, (2018-4).

Gabaix, X., & Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *The Quarterly Journal of Economics*, *121*(2), 505-540.

Gächter, S. (2012). In the lab and the field: Punishment is rare in equilibrium. *Behavioral & Brain Sciences*, *35*(1).

Gächter, S., Kölle, F., & Quercia, S. (2017). Reciprocity and the tragedies of maintaining and providing the commons. *Nature human behaviour*, *1*(9), 650-656.

Gaddy, C. G., Grossman, G., & Treml, V. G. (1991). *The labor market and the second economy in the Soviet Union* (No. 24). WEFA Group.

Gallup (2020, September 24). 61% of Americans Support Abolishing Electoral College. https://news.gallup.com/poll/320744/americans-support-abolishing-electoral-college.aspx

Gallup (2022). Americans' trust in media remains near record low. https://news.gallup.com/poll/403166/americans-trust-media-remains-near-record-low.aspx

Galton F. (1907). Vox populi. *Nature*, 450–451.

Gambling Commission. (2022). *Young people and gambling 2022: A research study among 11-16 year olds in Great Britain*. Birmingham, United Kingdom. https://www.gamblingcommission.gov.uk/statistics-and-research/publication/young-people-and-gambling-2022

Gandré, P., Mariathasan, M., Merrouche, O., & Ongena, S. (2020). Unintended Consequences of the Global Derivatives Market Reform. *Swiss Finance Institute Research Paper*, (20-02).

Garcia-Macia, D., Hsieh, C.T. and Klenow, P.J. (2019). How destructive is innovation? *Econometrica*, *87*(5), 1507-1541.

Gargiulo, M., & Benassi, M. (1999). *The dark side of social capital*. Springer US.

Garnaut, R. (2009). *The Great Crash Of 2008*. With Llewellyn-Smith, D. Melbourne Univ. Publishing.

Garnaut, R., & Clunies-Ross, A. (1983). Taxation of mineral rents. *OUP Catalogue*.

Garnaut, R., Emerson, C., Finighan, R. & Anthony, S. (2020). Replacing corporate income tax with a cash flow tax. *Australian Economic Review*, *53*(4), 463-481.

Garnaut, R., & Vines, D. (2007). Regional free-trade areas: sorting out the tangled spaghetti. *Oxford Review of Economic Policy*, *23*(3), 508-527.

Gat, A. (2009). So why do people fight? Evolutionary theory and the causes of war. European Journal of International Relations, 15(4), 571-599.

Gaynes, R., 2017. The discovery of penicillin—new insights after more than 75 years of clinical use. *Emerging infectious diseases*, *23*(5), p.849.

Geithner, T. F. (2017). Are we safe yet: How to manage financial crises. *Foreign Aff.*, *96*, 54.

Georgiadou, V., Rori, L., & Roumanias, C. (2018). Mapping the European far right in the 21st century: A meso-level analysis. *Electoral studies*, *54*, 103-115.

Gerber, A. (1998). Estimating the effect of campaign spending on senate election outcomes using instrumental variables. *American Political science review*, *92*(2), 401-411.

Gerber, A. S., Kessler, D. P., & Meredith, M. (2011). The persuasive effects of direct mail: A regression discontinuity based approach. *The Journal of Politics*, *73*(1), 140-155.

Garro, H. (2021). Political Consequences of Economic Hardship: State Economic Activity and Polarization in American Legislatures. *The Journal of Law, Economics, and Organization*, *37*(3), 512-533.

Gerschewski, J. (2013). The three pillars of stability: Legitimation, repression, and co-optation in autocratic regimes. *Democratization*, *20*(1), 13-38.

Gilbert, R. A. (1986). Requiem for regulation q: What it did and why it passed away. *Review*, *68*.

Gilens, M., & Page, B. I. (2014). Testing theories of American politics: Elites, interest groups, and average citizens. *Perspectives on politics*, *12*(3), 564-581.

Gill, J., Haslam, A., Crain, T., Herrera-Perez, D., & Prasad, V. (2020). Comparison of industry payments in 2017 with annual salary in a cohort of academic oncologists. *JAMA internal medicine*, *180*(5), 797-799.

Gillings, M.R. & Stokes, H.W. (2012). Are humans increasing bacterial evolvability?. *Trends in ecology & evolution*, *27*(6), 346-352.

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of theoretical biology*, *206*(2), 169-179.

Gintis, H. (2011). Gene–culture coevolution and the nature of human sociality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1566), 878-888.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and human Behavior*, *24*(3), 153-172.

Gintis, H., Henrich, J., Bowles, S., Boyd, R., & Fehr, E. (2008). Strong reciprocity and the roots of human morality. *Social Justice Research*, *21*, 241-253.

Gittell, R., & Vidal, A. (1998). *Community organizing: Building social capital as a development strategy*. Sage.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *The quarterly journal of economics*, *115*(3), 811-846.

Godard, R. (1993). Tit for tat among neighboring hooded warblers. *Behavioral Ecology and Sociobiology*, *33*(1), 45-50.

Gode, D. K., & Sunder, S. (1993). Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of political economy*, *101*(1), 119-137.

Godin, B. (2015). *Innovation contested: The idea of innovation over the centuries*. Routledge.

Godin, B. (2016). Making sense of innovation: from weapon to instrument to buzzword. *Quaderni*, 21-40.

Goldberg, A.L. & Rock, K.L. (1992). Proteolysis, proteasomes and antigen presentation. *Nature*, *357*(6377), 375-379.

Golden, M. A. (2006, December). Some puzzles of political corruption in modern advanced democracies. In *annual meeting of the Japan Political Science Association* (Vol. 29).

Goldman, E., Rocholl, J., & So, J. (2009). Do politically connected boards affect firm value?. *The review of financial studies*, *22*(6), 2331-2360.

Goldman, E., Rocholl, J., & So, J. (2013). Politically connected boards of directors and the allocation of procurement contracts. *Review of Finance*, *17*(5), 1617-1648.

Goldman, E., So, J., & Rocholl, J. (2006). Does political connectedness affect firm value?. *Available at SSRN 922698*.

Goodchild, M., Nargis, N., & d'Espaignet, E. T. (2018). Global economic cost of smoking-attributable diseases. *Tobacco control*, *27*(1), 58-64.

Goodhart, C. (2010). Are cocos from cloud cuckoo-land?. *Central Banking*, *21*(1), 29-33.

Goodhart, C. A., & Lastra, R. M. (2020). Equity finance: matching liability to power. *Journal of Financial Regulation*, *6*(1), 1-40.

Goodin, R. E., & Roberts, K. W. (1975). The ethical voter. *American Political Science Review*, *69*(3), 926-928.

Gordon, H.S. (1954). The economic theory of a common-property resource: the fishery. *Journal of political economy*, *62*(2), 124-142.

Gordon, R. (2017). *The rise and fall of American growth: The US standard of living since the civil war*. Princeton University Press.

Gormley, T. A., Kaviani, M., & Maleki, H. (2022). When do judges throw the book at companies? The influence of partisanship in corporate prosecutions. *The Influence of Partisanship in Corporate Prosecutions (September 23)*. http://dx.doi.org/10.2139/ssrn.3513177

Graeff, P., & Svendsen, G. T. (2013). Trust and corruption: The influence of positive and negative social capital on the economic development in the European Union. *Quality & Quantity*, *47*, 2829-2846.

Graham, M. H., & Svolik, M. W. (2020). Democracy in America? Partisanship, polarization, and the robustness of support for democracy in the United States. *American Political Science Review*, *114*(2), 392-409.

Gray, J. (1984). *Hayek on Liberty*. Oxford: Basil Blackwell.

Gray, J. (1998). *Hayek on Liberty*. Routledge.

Gray, J. (2013). The role of the sceptic. Interview with René Scheu. *Eurozine*. May 21. https://www.eurozine.com/the-role-of-the-sceptic/

Gray, J. (2015) The Friedrich Hayek I knew, and what he got right – and wrong. *The New Statesman*, July 30. https://www.newstatesman.com/long-reads/2015/07/john-gray-friedrich-hayek-i-knew-and-what-he-got-right-and-wrong

Green, D. P., & Krasno, J. S. (1988). Salvation for the spendthrift incumbent: Reestimating the effects of campaign spending in House elections. *American Journal of Political Science*, 884-907.

Greene, R. (2016). Understanding CoCos: What operational concerns & global trends mean for US policymakers. *Harvard University, John F. Kennedy School of Government, M-RCBG Associate Working Paper*, (62).

Greenwood, R., & Nagel, S. (2009). Inexperienced investors and bubbles. *Journal of Financial Economics*, *93*(2), 239-258.

Gregory, P., & Harrison, M. (2005). Allocation under dictatorship: research in Stalin's archives. *Journal of Economic Literature*, *43*(3), 721-761.

Gregory, R. (2007). Orgel's Second Rule and Unbeatable Strategies. *Genomicron*, November 22.

Greinacher, A., Krauel, K. & Jensch, I. (2012). HIT-antibodies promote their own antigen. *Blood, The Journal of the American Society of Hematology*, *120*(5), 930-931.

Grier, K. B. (1989). Campaign spending and Senate elections, 1978–84. *Public Choice*, *63*(3), 201-219.

Grier, K. B., & Munger, M. C. (1991). Committee assignments, constituent preferences, and campaign contributions. *Economic Inquiry*, *29*(1), 24-43.

Griffing, A. (2020, December 16). How Trump's war on Fox News is amplifying conspiracy theories and antisemitism. Haaretz. https://www.haaretz.com/world-news/2020-12-16/ty-article/.premium/trump-fox-news-oan-conspiracy-election-2020

Griffith, V. (2007). Using Evolution to Design Artificial Intelligence. Google Tech Talks November, 8 2007, Polyworld. https://www.youtube.com/watch?v=_m97_kL4ox0

Grman, E., Robinson, T.M. & Klausmeier, C.A. (2012). Ecological specialization and trade affect the outcome of negotiations in mutualism. *The American Naturalist*, *179*(5), 567-581.

Grofman, B. (1993). Is turnout the paradox that ate rational choice theory?. In B. Grofman (ed), *Information, Participation and Choice: An Economic Theory of Democracy in Perspective*. Ann Arbor: University of Michigan Press.

Grönlund, K., & Milner, H. (2006). The determinants of political knowledge in comparative perspective. *Scandinavian political studies*, *29*(4), 386-406.

Grönlund, K., & Setälä, M. (2007). Political trust, satisfaction and voter turnout. *Comparative European Politics*, *5*, 400-422.

Grossman, R. S. (2001). Double liability and bank risk taking. *Journal of money, credit and banking*, 143-159.

Group of Thirty (2017). Shadow Banking and Capital Markets: Risks and Opportunities, Washington, DC: The Group of Thirty. http://group30.org/images/uploads/publications/ShadowBankingCapitalMarkets_G30.pdf

Gruen, N. (2020, May 29). The evaluator general. *Club Troppo*. https://clubtroppo.com.au/2020/05/29/the-evaluator-general/

Guiso, L., Herrera, H., Morelli, M., & Sonno, T. (2017). Populism: Demand and supply. CEPR Discussion Paper No. DP11871.

Guiso, L., Sapienza, P., & Zingales, L. (2009). Cultural biases in economic exchange?. *The quarterly journal of economics*, *124*(3), 1095-1131.

Guriev, S., & Papaioannou, E. (2022). The political economy of populism. *Journal of Economic Literature*, *60*(3), 753-832.

Hackbarth, D., Rivera, A., & Wong, T. Y. (2022). Optimal short-termism. *Management Science*, *68*(9), 6477-6505.

Haig, D. (2020). *From Darwin to Derrida: Selfish Genes, Social Selves, and the Meanings of Life*. MIT Press.

Haldane, A. G. (2010). The $100 billion question. *Revista de Economía Institucional*, *12*(22), 83-110.

Haldane, A. G. (2012). Financial arms races. Speech at the Institute for New Economic Thinking, Berlin, 14 April. Bank of England. https://www.bankofengland.co.uk/-/media/boe/files/speech/2012/financial-arms-races.pdf

Haldane, A. G., Brennan, S., & Madouros, V. (2010). What is the contribution of the financial sector: Miracle or mirage. *The Future of Finance*, *87*.

Haldane, A. G., & Madouros, V. (2012). The dog and the frisbee. *366th Economic Policy Symposium 'The Changing Policy Landscape'*, Federal Reserve Bank of Kansas City, Jackson Hole, WY.

Hall, A. B. (2016). Systemic effects of campaign spending: evidence from corporate contribution bans in us state legislatures. *Political Science Research and Methods*, *4*(2), 343-359.

Hall, R. L., & Wayman, F. W. (1990). Buying time: Moneyed interests and the mobilization of bias in congressional committees. *American political science review*, *84*(3), 797-820.

Hamilton, W.D. (1964). The genetical theory of social behaviour. I, II. *Journal of Theoretical Biology*. 7(1), 1–52.

Hamilton, W.D. (1967). Extraordinary Sex Ratios: A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science*, *156*(3774), 477-488.

Hamilton, W.D. (1972). Altruism and related phenomena, mainly in social insects. *Annual Review of Ecology and systematics*, *3*(1), 193-232.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557-559.

Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the national academy of sciences*, *108*(50), 19931-19936.

Hamowy, R. (2003). FA Hayek and the common law. *Cato J.*, *23*, 241.

Han, J., & Pan, Z. (2016). CEO inside debt and investment-cash flow sensitivity. *Accounting & Finance*, *56*(2), 423-443.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, *103*(7), 2643-2682.

Handley, C., & Mathew, S. (2020). Human large-scale cooperation as a product of competition between cultural groups. *Nature communications*, *11*(1), 702.

Hardin, G. (1968). The tragedy of the commons: the population problem has no technical solution; it requires a fundamental extension in morality. *science*, *162*(3859), 1243-1248.

Hare, B., & Tomasello, M. (2004). Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Animal behaviour*, *68*(3), 571-581.

Harris, K.D. and Kolodkin-Gal, I. (2019). Applying the handicap principle to biofilms: condition-dependent signalling in Bacillus subtilis microbial communities. *Environmental Microbiology*, *21*(2), 531-540.

Harrison, M. (2006). A Soviet quasi-market for inventions: jet propulsion, 1932–1946. In *Research in Economic History,* 23, 1-59.

Hart, D. M. (2001). Why do some firms give? Why do some give a lot?: High-tech PACs, 1977–1996. *Journal of Politics*, *63*(4), 1230-1249.

Hart, O. (1995). *Firms, contracts, and financial structure*. Clarendon press.

Hastings, J. S., & Mitchell, O. S. (2011). How Financial Literacy and Impatience Shape Retirement Wealth and Investment Behaviors. *NBER Working Paper*, (w16740).

Hayashi, N., Ostrom, E., Walker, J., & Yamagishi, T. (1999). Reciprocity, trust, and the sense of control: A cross-societal study. *Rationality and society*, *11*(1), 27-46.

Hayashi, N. (1993). From Tit-for-Tat to Out-for-Tat: The Dilemma of the Prisoner's Network. *Sociological Theory and Methods* 8, 19–32.

Hayek, F. A. (1933). The trend of economic thinking. *Economica*, 40, 121-137. https://www.histecon.magd.cam.ac.uk/docs/Hayek.pdf

Hayek, F. A. (1945). *The road to serfdom*. University of Chicago Press.

Hayek, F. A. (1945). The use of knowledge in society. *The American Economic Review*, *35*(4), 519-530.

Hayek, F. A. (1958). Freedom, reason, and tradition. *Ethics*, *68*(4), 229-245.

Hayek, F. A. (1960a). *The constitution of liberty: The definitive edition*. Routledge.

Hayek, F.A. (1960b). New nations and the problem of power. *The Listener*, Nov 10, pp. 819 – 821.

Hayek, F. A. (1964). Kinds of order in society. *New Individualist Review*, 3(2), 3-12.

Hayek, F. A. (1967a). *Collectivist economic planning*. Ludwig von Mises Institute. (Original work published 1935)

Hayek, F. A. (1967b). *Studies in Philosophy, Politics and Economics*. London: Routledge.

Hayek, F. A. (1975). Two types of mind. *Encounter,* 45.

Hayek, F. A. (1979). *Law, Legislation and Liberty, vol. 3: The Political Order of a Free People*. Chicago, Ill.: University of Chicago Press.

Hayek, F. A. (1985). FA Hayek Interviewed By John O'Sullivan [Television broadcast]. United States: Foundation for Economic Education. https://mises.org/library/hayek-his-life-and-thought

Hayek, F. A. (1988). *The fatal conceit: The errors of socialism*. University of Chicago Press.

Hayek, F. A. (2005). Essay on Mandeville. In F. A. Hayek, *The trend of economic thinking: Essays on political economists and economic history*. Routledge.

Hayne, K. (2019). Royal commission into misconduct in the banking, superannuation and financial services industry. Report to Australian Parliament. https://www.royalcommission.gov.au/banking

He, G. (2015). The effect of CEO inside debt holdings on financial reporting quality. *Review of Accounting studies*, *20*, 501-536.

He, S., Hollenbeck, B., & Proserpio, D. (2022). The market for fake reviews. *Marketing Science*, *41*(5), 896-921.

Heclo, H. (1973). Presidential and prime ministerial selection. *Perspectives on Presidential Selection*, 19-48.

Heil, M., Rattke, J., & Boland, W. (2005). Postsecretory hydrolysis of nectar sucrose and specialization in ant/plant mutualism. *Science*, *308*(5721), 560-563.

Heimann, J. G. (2000). Comment on Calomiris, C.W. & Litan, R.E. (2000). Financial Regulation in a Global Marketplace. Brookings-Wharton Papers on Financial Services, 283-323.

Henderson, A. J., Landolt, M. A., McDonald, M. F., Barrable, W. M., Soos, J. G., Gourlay, W. & Landsberg, D. N. (2003). The living anonymous kidney donor: lunatic or saint?. *American journal of transplantation*, *3*(2), 203-213.

Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization*, *53*(1), 3-35.

Henrich, J. (2006). Cooperation, punishment, and the evolution of human institutions. *Science*, *312*(5770), 60-61.

Henrich, J., & Boyd, R. (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of theoretical biology*, *208*(1), 79-89.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElrcath, R. (2001). Cooperation, reciprocity and punishment in fifteen small-scale societies. *American Economic Review*, 91, 73 – 78.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H. & Tracer, D. (2005). 'Economic man' in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and brain sciences*, *28*(6), 795-815.

Herre, E.A., Knowlton, N., Mueller, U.G. & Rehner, S.A. (1999). The evolution of mutualisms: exploring the paths between conflict and cooperation. *Trends Ecol. Evol.* 14, 49–53

Herrmann, B. & Thöni, C. (2009). Measuring conditional cooperation: a replication study in Russia. *Experimental Economics*, *12*, 87-92.

Herrmann, B., Thoni, C., & Gachter, S. (2008). Antisocial punishment across societies. *Science*, *319*(5868), 1362-1367.

Hicks, B. M., Carlson, M. D., Blonigen, D. M., Patrick, C. J., Iacono, W. G., & MGue, M. (2012). Psychopathic personality traits and environmental contexts: Differential correlates, gender differences, and genetic mediation. *Personality Disorders: Theory, research, and treatment*, *3*(3), 209.

Hicks, J. (1939). *Value and capital*. Clarendon Press.

Hilbig, B. E., & Zettler, I. (2009). Pillars of cooperation: Honesty–Humility, social value orientations, and economic behavior. *Journal of Research in Personality*, *43*(3), 516-519.

Hilbig, B. E., Zettler, I., & Heydasch, T. (2012). Personality, punishment and public goods: Strategic shifts towards cooperation as a matter of dispositional honesty–humility. *European Journal of Personality*, *26*(3), 245-254.

Hill, C., & Painter, R. W. (2010). Berle's Vision Beyond Shareholder Interests: Why Investment Bankers Should Have (Some) Personal Liability. *Seattle UL Rev.*

Hill, M. D., Kelly, G. W., Lockhart, G. B., & Van Ness, R. A. (2013). Determinants and effects of corporate lobbying. *Financial Management*, *42*(4), 931-957.

Hill, M. D., Kelly, G. W., Lockhart, G. B., & Van Ness, R. A. (2013). Determinants and effects of corporate lobbying. *Financial Management*, *42*(4), 931-957.

Hirsch, F. (1978). *Social limits to growth*. Harvard University Press.

Hirshleifer, J., (1978). Competition, cooperation, and conflict in economics and biology. *The American Economic Review*, *68*(2), 238-243.

Hirshleifer, J. (1987). On the emotions as guarantors of threats and promises. *The Dark Side of the Force*, 198-219.

Hobbes, T. (1983). *De Cive* (H. Warrender, Ed.). Clarendon Press. (Original work published 1647).

Hobbes, T. (1985). *Leviathan* (C. B. Macpherson, Ed.). Penguin Classics. (Original work published 1651).

Hobbes, T. (2019). *The elements of law: natural and politic* (F. Tonnies, Ed.). Routledge. (Original work published 1640).

Hodgson, G. M. (2015). On defining institutions: rules versus equilibria. *Journal of Institutional Economics*, *11*(3), 497-505.

Hodgson, G.M. (1998). On the evolution of Thorstein Veblen's evolutionary economics. *Cambridge journal of economics*, *22*(4), 415-431.

Hoeksema, J.D. & Schwartz, M.W. (2003). Expanding comparative–advantage biological market models: contingency of mutualism on partner's resource requirements and acquisition trade–offs. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1518), 913-919.

Hol, F. J., Galajda, P., Nagy, K., Woolthuis, R. G., Dekker, C., & Keymer, J. E. (2013). Spatial structure facilitates cooperation in a social dilemma: empirical evidence from a bacterial community. *PloS one*, *8*(10), e77042.

Holldobler, B., & Wilson, E. O. (2009). *The superorganism: the beauty elegance and strangeness of insect societies*. WW Norton & Company.

Hommes, C., Sonnemans, J., Tuinstra, J., & Van de Velden, H. (2008). Expectations and bubbles in asset pricing experiments. *Journal of Economic Behavior & Organization*, *67*(1), 116-133.

Horowitz, D. L. (2000). *Ethnic groups in conflict, updated edition with a new preface*. Univ of California Press. (Original work published 1985)

Hortaçsu, A., & Syverson, C. (2004). Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds. *The Quarterly journal of economics*, *119*(2), 403-456.

Horváth, R. (2013). Does trust promote growth?. *Journal of Comparative economics*, *41*(3), 777-788.

Houthakker, H. S. (1950). Revealed preference and the utility function. *Economica*, *17*(66), 159-174.

Hu, H. T. (1993). Misunderstood derivatives: the causes of informational failure and the promise of regulatory incrementalism. *Yale LJ*, *102*, 1457.

Huang, J. M. (2010). Correcting Mandatory Injustice: Judicial Recommendation of Executive Clemency. *Duke Law Journal*, 131-165.

Huang, R., & Xuan, Y. (2020). 'Trading' Political Favors: Evidence from the Impact of the STOCK Act. Available at SSRN: https://ssrn.com/abstract=2765876 or http://dx.doi.org/10.2139/ssrn.2765876

Hubbard, T. D., Murray, I. A., Bisson, W. H., Sullivan, A. P., Sebastian, A., Perry, G. H., & Perdew, G. H. (2016). Divergent Ah receptor ligand selectivity during hominin evolution. *Molecular Biology and Evolution*, *33*(10), 2648-2658.

Hudson, S. K. T. J., Cikara, M., & Sidanius, J. (2019). Preference for hierarchy is associated with reduced empathy and increased counter-empathy towards others, especially out-group targets. *Journal of Experimental Social Psychology*, *85*, 103871.

Hume, D. (1975). *A Treatise of Human Nature*, edited by L. A. Selby-Bigge, 2nd ed. revised by P. H. Nidditch, Oxford: Clarendon Press. (Original work published 1739)

Huxley, T. H. (1887) On the reception of the 'Origin of Species'. Chapter 5 in Darwin, F. (1887) *The Life and Letters of Charles Darwin*. London: John Murray.

Ibarz, B., Leike, J., Pohlen, T., Irving, G., Legg, S. & Amodei, D., 2018. Reward learning from human preferences and demonstrations in atari. *Advances in neural information processing systems*, *31*.

Igan, D., & Mishra, P. (2014). Wall street, capitol hill, and k street: Political influence and financial regulation. *The Journal of Law and Economics*, *57*(4), 1063-1084.

Igan, D., Mishra, P., & Tressel, T. (2012). A fistful of dollars: Lobbying and the financial crisis. *NBER Macroeconomics Annual*, *26*(1), 195-230.

Ignatieff, M. (2013, Oct 13). Enemies vs. adversaries. The New York Times. https://www.nytimes.com/2013/10/17/opinion/enemies-vs-adversaries.html

Imhof, L.A., Fudenberg, D. & Nowak, M.A. (2005). Evolutionary cycles of cooperation and defection. *Proceedings of the National Academy of Sciences*, *102*(31), 10797-10800.

Inglehart, R., Basanez, M., Diez-Medrano, J., Halman, L., & Luijkx, R. (2000). World values surveys and European values surveys, 1981-1984, 1990-1993, and 1995-1997. *Ann Arbor-Michigan, Institute for Social Research, ICPSR version*.

Inglehart, R. F., & Norris, P. (2016). Trump, Brexit, and the rise of populism: Economic have-nots and cultural backlash. *SSRN Electronic Journal* No. 16-026.

Inoue, K., Blumenthal, D. M., Elashoff, D., & Tsugawa, Y. (2019). Association between physician characteristics and payments from industry in 2015–2017: observational study. *BMJ open*, *9*(9), e031010.

Ipsos (2016). Medical Practitioners' ongoing fitness and competence to practise. Commissioned by the Medical Board of Australia.

Irani, R. M., Iyer, R., Meisenzahl, R. R., & Peydro, J. L. (2021). The rise of shadow banking: Evidence from capital regulation. *The Review of Financial Studies*, *34*(5), 2181-2235.

Isaac, Richard E. (1995). *The Pleasures of Probability*. Springer.

Isaacson, W. (2023, 23 Mar). Fmr. Google CEO Eric Schmidt on the consequences of an A.I. revolution. Amanpour & Company, PBS. https://www.pbs.org/video/fmr-google-ceo-eric-schmidt-ai-revolution-2mknqb/

Jackman, R. W., & Volpert, K. (1996). Conditions favouring parties of the extreme right in Western Europe. *British Journal of Political Science*, *26*(4), 501-521.

Jacobson, G. C. (1978). The effects of campaign spending in congressional elections. *American Political Science Review*, *72*(2), 469-491.

James, C. D., Hoffman, M. T., Lightfoot, D. C., Forbes, G. S., & Whitford, W. G. (1994). Fruit abortion in Yucca elata and its implications for the mutualistic association with yucca moths. *Oikos*, 207-216.

Jankowski, R. (2002). Buying a lottery ticket to help the poor: Altruism, civic duty, and self-interest in the decision to vote. *Rationality and Society*, *14*(1), 55-77.

Jayachandran, S. (2006). The jeffords effect. *The Journal of Law and Economics*, *49*(2), 397-425.

Je, H.S., Yang, F., Ji, Y., Potluri, S., Fu, X.Q., Luo, Z.G., Nagappan, G., Chan, J.P., Hempstead, B., Son, Y.J. and Lu, B. (2013). ProBDNF and mature BDNF as punishment and reward signals for synapse elimination at mouse neuromuscular junctions. *Journal of Neuroscience*, *33*(24), 9957-9962.

Jensen, K., Call, J., & Tomasello, M. (2007a). Chimpanzees are vengeful but not spiteful. *Proceedings of the National Academy of Sciences*, *104*(32), 13046-13050.

Jensen, K., Call, J., & Tomasello, M. (2007b). Chimpanzees are rational maximizers in an ultimatum game. *science*, *318*(5847), 107-109.

Jensen, K., Hare, B., Call, J., & Tomasello, M. (2006). What's in it for me? Self-regard precludes altruism and spite in chimpanzees. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1589), 1013-1021.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial economics*, *3*(4), 305-360.

Jeon, K. W. (1972). Development of cellular dependence on infective organisms: micrurgical studies in amoebas. *Science*, 176, 1122– 1123.

Jerit, J., & Barabas, J. (2012). Partisan perceptual bias and the information environment. *The Journal of Politics*, *74*(3), 672-684.

Jetten, J., Mols, F., & Postmes, T. (2015). Relative deprivation and relative wealth enhances anti-immigrant sentiments: The v-curve re-examined. *PloS one*, *10*(10), e0139156.

Johnson, A. (2023). Failure to rule on a motion and delay of judgements, Trial Rule 53.1 and 53.2. Procedural Issues note, Indiana. https://www.in.gov/courts/iocs/files/pubs-trial-court-failure-to-rule-on-motion-and-delay-of-judgments.pdf

Johnson, D., & Krüger, O. (2004). The good of wrath: Supernatural punishment and the evolution of cooperation. *Political theology*, *5*(2), 159-176.

Johnson, D., & Bering, J. (2006). Hand of God, mind of man: Punishment and cognition in the evolution of cooperation. *Evolutionary psychology*, *4*(1), 147470490600400119.

Johnson, M. K., Rowatt, W. C., & Petrini, L. (2011). A new trait on the market: Honesty–Humility as a unique predictor of job performance ratings. *Personality and Individual differences*, *50*(6), 857-862.

Johnson, N. C., Graham, J. H., & Smith, F. A. (1997). Functioning of mycorrhizal associations along the mutualism–parasitism continuum. *The New Phytologist*, *135*(4), 575-585.

Johnstone, R.A. & Bshary, R. (2002). From parasitism to mutualism: partner control in asymmetric interactions. *Ecology Letters*, *5*(5), 634-639.

Jokela, J., Dybdahl, M.F. & Lively, C.M. (2009). The maintenance of sex, clonal dynamics, and host-parasite coevolution in a mixed population of sexual and asexual snails. *The American Naturalist*, *174*(S1), S43-S53.

Jonason, P. K., & Schmitt, D. P. (2012). What have you done for me lately? Friendship-selection in the shadow of the Dark Triad traits. *Evolutionary Psychology*, *10*(3).

Jones, B. F. (2009). The burden of knowledge and the 'death of the renaissance man': Is innovation getting harder?. *The Review of Economic Studies*, *76*(1), 283-317.

Jones, C., & Philippon, T. (2016). The secular stagnation of investment?. *Unpublished manuscript, New York University, December*.

Jones, E. (2000). *Growth recurring: Economic change in world history*. Oxford, Clarendon Press. Original published 1988.

Jones, J. S., Lee, W. Y., & Yeager, T. J. (2012). Opaque banks, price discovery, and financial instability. *Journal of Financial Intermediation*, *21*(3), 383-408.

Joron, M. & Mallet, J.L. (1998). Diversity in mimicry: paradox or paradigm?. *Trends in ecology & evolution*, *13*(11), 461-466.

Kabir, R., Li, H., & Veld-Merkoulova, Y. (2018). Do managers act opportunistically towards the end of their career?. *Journal of International Financial Markets, Institutions and Money*, *56*, 218-232.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 363-391.

Kailitz, S. (2013). Classifying political regimes revisited: legitimation and durability. *Democratization*, *20*(1), 39-60.

Kaldor, N. (1939). Welfare Propositions in Economics and Interpersonal Comparisons of Utility. *The Economic Journal*. 49(195), 549–552.

Kamm, F. M. (1989). Harming some to save others. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *57*(3), 227-260.

Kandler, C., Richter, J., & Zapko-Willmes, A. (2019). The nature and nurture of HEXACO personality trait differences. *Zeitschrift für Psychologie*.

Kane, E. J. (2002). Using deferred compensation to strengthen the ethics of financial regulation. *Journal of Banking & Finance*, *26*(9), 1919-1933.

Kashyap, A. K., Stein, J. C., & Hanson, S. (2010). An analysis of the impact of 'substantially heightened'capital requirements on large financial institutions. *Booth School of Business, University of Chicago, mimeo*, *2*, 1-47.

Kato-Shimizu, M., Onishi, K., Kanazawa, T. & Hinobayashi, T., 2013. Preschool children's behavioral tendency toward social indirect reciprocity. *PLoS One*, *8*(8), p.e70915.

Katsenelinboigen, A. (1977). Coloured markets in the Soviet Union. *Soviet studies*, *29*(1), 62-85.

Kaufmann, K. M. (2004). The urban voter: Group conflict and mayoral voting behavior in American cities. University of Michigan Press.

Kautz, S., Lumbsch, H. T., Ward, P. S., & Heil, M. (2009). How to prevent cheating: a digestive specialization ties mutualistic plant-ants to their ant-plant partners. *Evolution*, *63*(4), 839-853.

Kawai, M. (1965). Newly-acquired pre-cultural behavior of the natural troop of Japanese monkeys on Koshima Islet. *Primates*, *6*, 1-30.

Kay, J., & King, M. (2020). *Radical uncertainty: Decision-making beyond the numbers*. W. W. Norton & Company.

Kayser, M. A., & Wlezien, C. (2011). Performance pressure: Patterns of partisanship and the economic vote. *European Journal of Political Research*, *50*(3), 365-394.

Kehoe, P. (2009). Confronting Models of Financial Frictions with the Data. In *2009 Meeting Papers* (No. 1083). Society for Economic Dynamics.

Kelly, R. C. (1985). *The Nuer conquest: The structure and development of an expansionist system*. University of Michigan press.

Kennedy, G. (2010). Paul Samuelson and the invention of the modern economics of the Invisible Hand. *History of Economic Ideas*, *18*(3), 105-120.

Kessler, J. B., & Leider, S. (2012). Norms and contracting. *Management Science*, *58*(1), 62-77.

Keynes, J. M. (1921). *A treatise on probability*. Macmillan.

Keynes, J. M. (1936). *The general theory of employment, interest, and money*. Palgrave Macmillan.

Keynes, J. M. (1937). The general theory of employment. *The Quarterly Journal of Economics*, 51(2), 209-223.

Keynes, J. M. (1980). *Activities 1940-1946: Shaping the post-war world, employment and commodities* (D. E. Moggridge, Ed.). Macmillan.

Keynes, J. M. (1982). *The collected writings of John Maynard Keynes, vol. XXVIII: Social, political and literary writings* (D. Moggridge ed.). New York: Cambridge University Press.

Khadem, N. (2020, January 21). Trivago misled consumers on hotel pricing, court finds, as ACCC warns of wider crackdown. *Australian Broadcasting Corporation*. https://www.abc.net.au/news/2020-01-21/tirvago-mislead-consumers-on-hotel-pricing-court-finds-accc/11886096

Khwaja, A. I., & Mian, A. (2005). Do lenders favor politically connected firms? Rent provision in an emerging financial market. *The quarterly journal of economics*, *120*(4), 1371-1411.

Kiers, E.T., Rousseau, R.A. & Denison, R.F. (2006). Measured sanctions: legume hosts detect quantitative variation in rhizobium cooperation and punish accordingly. *Evolutionary Ecology Research*, *8*(6), 1077-1086.

Kiers, E.T., Rousseau, R.A., West, S.A. & Denison, R.F. (2003). Host sanctions and the legume–rhizobium mutualism. *Nature*, *425*(6953), 978-81.

Kim, H.W., Rapoport, S.I. and Rao, J.S. (2010). Altered expression of apoptotic factors and synaptic markers in postmortem brain from bipolar disorder patients. *Neurobiology of disease*, *37*(3), 596-603.

Kim, J. H. (2008). Corporate lobbying revisited. *Business and Politics*, *10*(2), 1-23.

Kinder, D. R., & Kiewiet, D. R. (1979). Economic discontent and political behavior: The role of personal grievances and collective economic judgments in congressional voting. *American Journal of Political Science*, 495-527.

King, M. (2009, October 20). Speech to Scottish Business Organisations in Edinburgh. https://www.bankofengland.co.uk/-/media/boe/files/speech/2009/mervyn-king-speech-to-scottish-business-organisations.pdf

Kingdon, G. G., & Knight, J. (2007). Community, comparisons and subjective well-being in a divided society. *Journal of Economic Behavior & Organization*, *64*(1), 69-90.

Klašnja, M. (2015). Corruption and the incumbency disadvantage: Theory and evidence. *The Journal of Politics*, *77*(4), 928-942.

Kleiner, M. M., & Krueger, A. B. (2013). Analyzing the extent and influence of occupational licensing on the labor market. *Journal of Labor Economics*, *31*(S1), S173-S202.

Knack, S. (2002). Social capital and the quality of government. *American Journal of Political Science*, 46: 772-85

Knack, S., & Keefer, P. (1997). Does social capital have an economic payoff? A cross-country investigation. *The Quarterly journal of economics*, *112*(4), 1251-1288.

Knafo, A., & Plomin, R. (2006). Prosocial behavior from early to middle childhood: Genetic and environmental influences on stability and change. Developmental Psychology, 42(5), 771–786. https://doi.org/10.1037/0012-1649.42.5.771

Knight, F. H. (1921). *Risk, uncertainty and profit*. Hart, Schaffner & Marx.

Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J., & Sutter, M. (2008). Conditional cooperation on three continents. *Economics letters*, *101*(3), 175-178.

Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual review of sociology*, *24*(1), 183-214.

Koonin, E.V., 2016. Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. *F1000Research*, *5*.

Kosse, F., Deckers, T., Pinger, P., Schildberg-Hörisch, H., & Falk, A. (2020). The formation of prosociality: causal evidence on the role of social environment. *Journal of Political Economy*, *128*(2), 434-467.

Koudijs, P., Salisbury, L., & Sran, G. (2021). For richer, for poorer: bankers' liability and bank risk in New England, 1867 to 1880. *The Journal of Finance*, *76*(3), 1541-1599.

Kovács, K., & Scheppele, K. L. (2018). The fragility of an independent judiciary: Lessons from Hungary and Poland—And the European Union. *Communist and Post-Communist Studies*, *51*(3), 189-200.

Krakovna, V., Uesato, J., Mikulik, V., Rahtz, M., Everitt, T., Kumar, R., Kenton, Z., Leike, J. & Legg, S. (2020). Specification gaming: the flip side of AI ingenuity. *DeepMind Blog*.

Kramer, R. M., & Brewer, M. B. (1984). Effects of group identity on resource use in a simulated commons dilemma. *Journal of personality and social psychology*, *46*(5), 1044.

Kremer, M. (1993). Population growth and technological change: One million BC to 1990. *The quarterly journal of economics*, *108*(3), 681-716.

Kroszner, R. S., & Strahan, P. E. (1999). What drives deregulation? Economics and politics of the relaxation of bank branching restrictions. *The Quarterly Journal of Economics*, *114*(4), 1437-1467.

Kroszner, R. S., & Stratmann, T. (1998). Interest-group competition and the organization of congress: theory and evidence from financial services' political action committees. *American economic review*, 1163-1187.

Kroszner, R. S., & Stratmann, T. (2005). Corporate campaign contributions, repeat giving, and the rewards to legislator reputation. *The Journal of law and Economics*, *48*(1), 41-71.

Krueger, A. O. (1963). The economics of discrimination. *Journal of Political Economy*, *71*(5), 481-486.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, *77*(6), 1121.

Krugman, P. (1996). What economists can learn from evolutionary theorists. *A talk given to the European Association for Evolutionary Political Economy*. http://www.mit.edu/~krugman/evolute.html

Krugman, P. (2009, April 9). Making banking boring. *New York Times*.

Krugman, P. (2009, September 2). How did economists get it so wrong? *New York Times*.

Krugman, P. (2013). Friedman and the Austrians. *New York Times*, August 11.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, *108*(3), 480.

Kurzban, R., & Houser, D. (2005). Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences*, *102*(5), 1803-1807.

Kuziemko, I., Norton, M. I., Saez, E., & Stantcheva, S. (2015). How elastic are preferences for redistribution? Evidence from randomized survey experiments. *American Economic Review*, *105*(4), 1478-1508.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1996). Trust in large organizations.

Laërtius, D. (1925). *Lives and Opinions of Eminent Philosophers*. Translated by Hicks, Robert Drew. Harvard University Press. (Original work published 3rd century A.D.)
https://penelope.uchicago.edu/Thayer/E/Roman/Texts/Diogenes_Laertius/Lives_of_the_Eminent_Philosophers/1/Anacharsis*.html

Laeven, L., & Levine, R. (2009). Bank governance, regulation and risk taking. *Journal of financial economics*, *93*(2), 259-275.

Laffont, J. J., & Tirole, J. (1993). A theory of incentives in procurement and regulation. MIT press.

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1992). The impact of institutional trading on stock prices. *Journal of financial economics*, *32*(1), 23-43.

Landier, A., Sraer, D. A., & Thesmar, D. (2015, June). The risk-shifting hypothesis: Evidence from subprime originations. In *AFA 2012 Chicago Meetings Paper*.

Lang, P. (2015). Science: Antiquity and its Legacy. Bloomsbury.

Lange, O. (1938). On the Economic Theory of Socialism. B. Lippincott (Ed.). University of Minnesota Press.

Laver, M. (1997). Private desires, political action: An invitation to the politics of rational choice. Sage.

Law, S. H., & Singh, N. (2014). Does too much finance harm economic growth?. *Journal of Banking & Finance*, *41*, 36-44.

Layte, R., & Whelan, C. T. (2014). Who feels inferior? A test of the status anxiety hypothesis of social inequalities in health. *European Sociological Review*, *30*(4), 525-535.

Le Clec'h, W., Chevalier, F.D., Genty, L., Bertaux, J., Bouchon, D. & Sicard, M., 2013. Cannibalism and predation as paths for horizontal passage of Wolbachia between terrestrial isopods. *PloS one*, *8*(4), p.e60232.

Lee, J., Murphy, K. J., Oh, P. S. H., & Vance, M. D. (2015). Inside debt and corporate investment. *Inside Debt and Corporate Investment (January 26, 2018). AAA*.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate behavioral research*, *39*(2), 329-358.

Lehman, J., Clune, J., Misevic, D., Adami, C., Altenberg, L., Beaulieu, J., Bentley, P.J., Bernard, S., Beslon, G., Bryson, D.M. & Chrabaszcz, P. (2018). The surprising creativity of digital evolution: A collection of anecdotes from the evolutionary computation and artificial life research communities. *arXiv preprint arXiv:1803.03453*.

Lehtonen, J., 2020. The Price equation and the unity of social evolution theory. *Philosophical Transactions of the Royal Society B*, *375*(1797), p.20190362.

Leimar, O., & Hammerstein, P. (2001). Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1468), 745-753.

Lerner, A (1938). Theory and Practice in Socialist Economics. *The Review of Economic Studies*. 6(1), 71–5.

Lessenski, M. (2022). How It Started, How It is Going: Media Literacy Index 2022. Media Literacy Index project of the European Policies Initiative (EuPI) of Open Society Institute–Sofia Foundation (OSI-Sofia). https://osis.bg/wp-content/uploads/2022/10/HowItStarted_MediaLiteracyIndex2022_ENG_.pdf

Levi, M. (1998). A state of trust. In V. Braithwaite & M. Levi (Eds.), *Trust and Governance*. New York: Russell Sage.

Levi, M., & Stoker, L. (2000). Political trust and trustworthiness. *Annual review of political science*, *3*(1), 475-507.

Levitsky, S., & Way, L. (2022). Revolution and Dictatorship. In *Revolution and Dictatorship*. Princeton University Press.

Levitsky, S., & Ziblatt, D. (2018). *How democracies die*. Crown.

Levitt, K. & McRobbie, K. (2006). *Karl Polanyi in Vienna: the contemporary significance of the Great Transformation* (Vol. 8). Black Rose Books Limited.

Levitt, S. D. (1994). Using repeat challengers to estimate the effect of campaign spending on election outcomes in the US House. *Journal of Political Economy*, *102*(4), 777-798.

Lewis, P. (2011, February 21). The great rock 'n' coal swindle. *Australian Broadcasting Corporation*.

Lewontin, R.C. (1970). The units of selection. *Annual review of ecology and systematics*, *1*(1), 1-18.

Li, P., Guo, Y., & Meng, H. (2022). The default contagion of contingent convertible bonds in financial network. *The North American Journal of Economics and Finance*, *60*, 101661.

Li, W., Mills, D., & Nower, L. (2019). The relationship of loot box purchases to problem video gaming and problem gambling. *Addictive behaviors*, *97*, 27-34.

Li, Z. F., Lin, S., Sun, S., & Tucker, A. (2018). Risk-adjusted inside debt. *Global Finance Journal*, *35*, 12-42.

Liebrand, W.B. (1984). The effect of social motives, communication and group size on behaviour in an N-person multi-stage mixed-motive game. *European journal of social psychology*, *14*(3), 239-264.

Lindholm, M. (2014). Morphologically conservative but physiologically diverse: the mode of stasis in Anostraca (Crustacea: Branchiopoda). *Evolutionary Biology*, *41*, 503-507.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*. 47(1), 13–37. doi:10.2307/1924119

Liu, Y., Mauer, D. C., & Zhang, Y. (2014). Firm cash holdings and CEO inside debt. *Journal of Banking & Finance*, *42*, 83-100.

Lively, C.M. (2010). A review of Red Queen models for the persistence of obligate sexual reproduction. *Journal of Heredity*, *101*(1), S13-S20.

Loewenstein, G., Friedman, J. Y., McGill, B., Ahmad, S., Linck, S., Sinkula, S., ... & Volpp, K. G. (2013). Consumers' misunderstanding of health insurance. *Journal of Health Economics*, *32*(5), 850-862.

Lovelock, J.E. and Margulis, L. (1974). Atmospheric homeostasis by and for the biosphere: the Gaia hypothesis. *Tellus*, 26(1-2), 2-10.

Lucas Jr, R. E. (2013). Glass-Steagall: A requiem. *American Economic Review*, 103(3), 43-47.

Lyons, B., & Mehta, J. (1997). Contracts, opportunism and trust: self-interest and social orientation. *Cambridge journal of economics*, *21*(2), 239-257.

Ma, H., & O'Farrell, P. H. (2016). Selfish drive can trump function when animal mitochondrial genomes compete. *Nature genetics*, *48*(7), 798-802.

Mace, R., & Holden, C. J. (2005). A phylogenetic approach to cultural evolution. *Trends in ecology & evolution*, *20*(3), 116-121.

Machiavelli, N. (1532). *The Prince*. Antonio Blado d'Asola, Italy.

Makinson, L. (1992). Political contributions from the health and insurance industries. *Health Affairs*, *11*(4), 119-134.

Maclean, J. C., Mallatt, J., Ruhm, C. J., & Simon, K. (2020). Economic studies on the opioid crisis: A review.

Macro International. (2008). Design and testing of effective Truth in Lending disclosures. U.S. Federal Reserve. https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20081218a8.pdf

Madison, J. (2016). The federalist papers. In *Democracy: A reader* (52-57). Columbia University Press.

Maguire, S., Philips, N. & Hardy, C. (2001). When 'silence=death', keep talking: Trust, control and the discursive construction of identity in the Canadian HIV/AIDS treatment domain. *Organization Studies*, 22(2), 285-310.

Majolo, B., Schino, G. & Aureli, F. (2012). The relative prevalence of direct, indirect and generalized reciprocity in macaque grooming exchanges. *Animal behaviour*, *83*(3), 763-771.

Makimura, Y., & Yamagishi, T. (2003). Ongoing group interaction, ingroup favoritism, and reward allocation. *Shinrigaku kenkyu: The Japanese journal of psychology*, *73*(6), 488-493.

Malthus, T. R. (1798). *An essay on the principle of population*. J. Johnson, London.

Mandeville, B. (1989). *Fable of the bees: Or, private vices, public benefits* (P. Harth, Ed.). Penguin Classics. (Original work published 1714)

Manzetti, L., & Wilson, C. J. (2007). Why do corrupt governments maintain public support?. *Comparative political studies*, *40*(8), 949-970.

Margolis, M. F. (2020). Who wants to make America great again? Understanding evangelical support for Donald Trump. *Politics and Religion*, *13*(1), 89-118.

Mariathasan, M., & Merrouche, O. (2014). The manipulation of basel risk-weights. *Journal of Financial Intermediation*, *23*(3), 300-321.

Markowitz, H. M. (1952). Portfolio Selection. *The Journal of Finance*. 7(1), 77–91.

Markus, G. B. (1988). The impact of personal and national economic conditions on the presidential vote: A pooled cross-sectional analysis. *American Journal of Political Science*, 137-154.

Marsh, A. A. (2013). What can we learn about emotion by studying psychopathy?. *Frontiers in human neuroscience*, *7*, 181.

Marsh, A. A., Stoycos, S. A., Brethel-Haurwitz, K. M., Robinson, P., VanMeter, J. W., & Cardinale, E. M. (2014). Neural and cognitive characteristics of extraordinary altruists. *Proceedings of the National Academy of Sciences*, *111*(42), 15036-15041.

Marshall, A. (1920). *Principles of Economics 8th edn*. London Macmillan.

Marshall, C. & Scott, P. (2012, November 14). Sacred women's business and harsh tribal punishments. *Australian Broadcasting Corporation*. https://www.abc.net.au/local/audio/2012/05/25/3629680.htm

Marshall, J.A. (2011). Group selection and kin selection: formally equivalent approaches. *Trends in Ecology & Evolution*, *26*(7), 325-332.

Martincorena, I., Raine, K.M., Gerstung, M., Dawson, K.J., Haase, K., Van Loo, P., Davies, H., Stratton, M.R. & Campbell, P.J. (2017). Universal patterns of selection in cancer and somatic tissues. *Cell*, *171*(5), 1029-1041.

Martynova, N., & Perotti, E. (2018). Convertible bonds and bank risk-taking. *Journal of Financial Intermediation*, *35*, 61-80.

Maynard Smith, J. (1976). Letters, Haldane. Letter to the editor from J. Maynard Smith, *New Scientist*, p. 247, July 29. New Science Publications, London.

Maynard-Smith, J. (1978). *The evolution of sex* (Vol. 4). Cambridge: Cambridge University Press.

Maynard Smith, J. (1988). *Evolutionary progress and levels of selection. In: Evolutionary Progress* (M.H. Nitecki, ed.), 219-230. University of Chicago Press, Chicago

Mazzucato, M. & Skidelsky, R. (2020). Toward a New Fiscal Constitution. *Project Syndicate*. July 10.

McCraw, T. K. (1986). *Prophets of regulation*. Harvard University Press.

McElreath, R., Boyd, R., & Richerson, P. (2003). Shared norms and the evolution of ethnic markers. *Current anthropology*, *44*(1), 122-130.

McGuire, M. C., & Olson, M. (1996). The economics of autocracy and majority rule: The invisible hand and the use of force. *Journal of economic literature*, *34*(1), 72-96.

McLannahan, B. & Dye, J. (2017, April 21). Deutsche Bank fined $156.6 m over currency and Volcker violations. *Financial Times*. https://www.ft.com/content/7dc4e7ec-2620-11e7-8691-d5f7e0cd0a16

McLean, I. (2002). William H. Riker and the Invention of Heresthetic (s). *British Journal of Political Science*, *32*(3), 535-558.

McLean, I., & Hewitt, F. (Eds.). (1994). *Condorcet: foundations of social choice and political theory*. Edward Elgar Publishing.

Meloy, J. R. (1988). *The psychopathic mind: Origins, dynamics, and treatment*. Rowman & Littlefield.

Mehran, H., & Tracy, J. S. (2016). Deferred cash compensation: Enhancing stability in the financial services industry. *Economic Policy Review, Issue Aug*, 61-75.

Mian, A., Sufi, A., & Trebbi, F. (2010). The political economy of the US mortgage default crisis. *American Economic Review*, *100*(5), 1967-1998.

Michaels, J. D. (2017). The American deep state. *Notre Dame L. Rev.*, *93*, 1653.

Mihet, R., & Philippon, T. (2019). The economics of big data and artificial intelligence. In *Disruptive Innovation in Business and Finance in the Digital World*. Emerald Publishing Limited.

Mill, J. S. (1861). *Considerations on representative government*. London: Parker, Son, and Bourn.

Miller, M. (2021). Choosing the narrative: The shadow banking crisis in light of Covid. *Open Economies Review*, *32*(2), 291-310.

Milyo, J. (1997). Electoral and financial effects of changes in committee power: the Gramm-Rudman-Hollings budget reform, the Tax Reform Act of 1986, and the money committees in the House. *The Journal of Law and Economics*, *40*(1), 93-112.

Milyo, J., Primo, D., & Groseclose, T. (2000). Corporate PAC campaign contributions in perspective. *Business and Politics*, *2*(1), 75-88.

Miron, J., Sollenberger, G. M., & Nicolae, L. (2019). Overdosing on regulation: How government caused the opioid epidemic. *Cato Institute Policy Analysis*, (864).

Mises, L. (1922). *Socialism: An economic and sociological analysis*. Gustav Fischer Verlag.

Mitchell, A. P., Trivedi, N. U., Gennarelli, R. L., Chimonas, S., Tabatabai, S. M., Goldberg, J., ... & Korenstein, D. (2021). Are financial payments from the pharmaceutical industry associated with physician prescribing? A systematic review. *Annals of internal medicine*, *174*(3), 353-361.

Mitnick, B. M. (1975). The theory of agency: The policing paradox and regulatory behavior. *Public Choice*, *24*(1), 27-42.

Modigliani, F. & Miller, M.H. (1958). The cost of capital, corporation finance and the theory of investment. *The American Economic Review*, *48*(3), 261-297.

Mohr, P. B., & Larsen, K. (1998). Ingroup favoritism in umpiring decisions in Australian football. *The Journal of Social Psychology*, *138*(4), 495-504.

Mokyr, J. (1998). *Neither chance nor necessity: Evolutionary models and economic history*. Princeton University Press.

Montalvo, J. G., & Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. *American economic review*, *95*(3), 796-816.

Montalvo, J. G., & Reynal-Querol, M. (2021). Ethnic diversity and growth: Revisiting the evidence. *The review of economics and statistics*, *103*(3), 521-532.

Moravec, P., Minas, R., & Dennis, A. R. (2018). Fake news on social media: People believe what they want to believe when it makes no sense at all. *Kelley School of Business research paper*, (18-87).

Moreno-Fenoll, C., Cavaliere, M., Martínez-García, E. & Poyatos, J.F. (2016). Exploitation by cheaters facilitates the preservation of essential public goods in microbial communities. *bioRxiv*, p.040964.

Morran, L.T., Schmidt, O.G., Gelarden, I.A., Parrish, R.C. & Lively, C.M. (2011). Running with the Red Queen: host-parasite coevolution selects for biparental sex. *Science*, *333*(6039), 216-218.

Moschella, M., & Tsingou, E. (2013). Regulating finance after the crisis: Unveiling the different dynamics of the regulatory process. *Regulation & Governance*, *7*(4), 407-416.

Mossner, E. C. (1941). An apology for David Hume, historian. *PMLA*, 56(3), 657-690.

Mounk, Y. (2018). *The people vs. democracy: Why our freedom is in danger & how to save it*. Harvard University Press.

Mudde, C. (2004). The populist zeitgeist. *Government and opposition*, *39*(4), 541-563.

Mudde, C. (2007). *Populist radical right parties in Europe*. Cambridge University Press.

Mukand, S., & Rodrik, D. (2018). *The political economy of ideas: On ideas versus interests in policymaking* (No. w24467). National Bureau of Economic Research.

Mukand, S. W., & Rodrik, D. (2020). The political economy of liberal democracy. *The Economic Journal*, *130*(627), 765-792.

Mullainathan, S., & Washington, E. (2009). Sticking with your vote: Cognitive dissonance and political attitudes. *American Economic Journal: Applied Economics*, *1*(1), 86-111.

Mummendey, A., Kessler, T., Klink, A., & Mielke, R. (1999). Strategies to cope with negative social identity: predictions by social identity theory and relative deprivation theory. *Journal of Personality and Social Psychology*, *76*(2), 229.

Murphy, K.M., Shleifer, A. & Vishny, R.W. (1991). The allocation of talent: Implications for growth. *The Quarterly Journal of Economics*, *106*(2), 503-530.

Murphy, T. (2013). The first level of Super Mario Bros. is easy with lexicographic orderings and time travel. *The Association for Computational Heresy (SIGBOVIK)*, *2013*, 112-133.

Myerson, R. B., & Satterthwaite, M. A. (1983). Efficient mechanisms for bilateral trading. *Journal of economic theory*, *29*(2), 265-281.

Nakagawara, A., Azar, C.G., Scavarda, N.J. & Brodeur, G.M. (1994). Expression and function of TRK-B and BDNF in human neuroblastomas. *Molecular and cellular biology*, *14*(1), 759-767.

Nasiripour, S. (2018, June 21). Wild trading day at Deutsche Bank raises questions on risk. *Bloomberg*. https://www.bloomberg.com/news/articles/2018-06-20/wild-trading-day-at-deutsche-bank-raises-questions-on-u-s-risk

Nelson, R.R. & Winter, S.G. (1973). Toward an evolutionary theory of economic capabilities. *The American Economic Review*, *63*(2), 440-449.

Nelson, R.R. (2007). Universal Darwinism and evolutionary social science. *Biology & Philosophy*, *22*, 73-94.

Nelson, R.R., Dosi, G., Helfat, C.E., Pyka, A., Saviotti, P.P., Lee, K., Winter, S.G., Dopfer, K. & Malerba, F. (2018). *Modern evolutionary economics: An overview*. Cambridge University Press.

Neundorf, A., Gerschewski, J., & Olar, R. G. (2020). How do inclusionary and exclusionary autocracies affect ordinary people?. *Comparative Political Studies*, *53*(12), 1890-1925.

New America Foundation. (2021). Statement of concern: The threats to American democracy and the need for national voting and election administration standards. https://www.newamerica.org/political-reform/statements/statement-of-concern/

Newell, A. & Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-hall.

Nilsson-Møller, S., Poulsen, M., & Innocent, T. M. (2018). A visual guide for studying behavioral defenses to pathogen attacks in leaf-cutting ants. *Journal of visualized experiments: JoVE*, (140).

Noë, R. & Hammerstein, P. (1994). Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behavioral ecology and sociobiology*, *35*, 1-11.

Nooteboom, B. (2002). *Trust: Forms, foundations, functions, failures and figures*. Edward Elgar Publishing.

Nooteboom, B. (2007). Social capital, institutions and trust. *Review of social economy*, *65*(1), 29-53.

Norenzayan, A., Shariff, A. F., Gervais, W. M., Willard, A. K., McNamara, R. A., Slingerland, E., & Henrich, J. (2016). The cultural evolution of prosocial religions. *Behavioral and brain sciences*, *39*, e1.

Norris, P. (2021). Voters against democracy: the roots of autocratic resurgence. *Foreign Aff.*, *100*, 174.

North, D. C. (1979). A framework for analyzing the state in economic history. *Explorations in Economic History*, 16(3), 249.

North, D. C. (1986). The new institutional economics. *Journal of Institutional and Theoretical Economics*, 142(1), 230-237.

North, D.C. (1988). Ideology and political/economic institutions. *Cato J.*, *8*, 15.

North, D.C. (1990). A transaction cost theory of politics. *Journal of theoretical politics*, *2*(4), 355-367.

North, D.C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press.

North, D. C. (1992). *Transaction costs, institutions, and economic performance*. ICS Press.

North, D.C. (2005). *Understanding the process of economic change*. Princeton University Press.

North, D. C., & Thomas, R. P. (1973). *The rise of the western world: A new economic history*. Cambridge University Press.

Nowak, M. A., & Sigmund, K. (1998). The dynamics of indirect reciprocity. *Journal of theoretical Biology*, *194*(4), 561-574.

Nozick, R. (1974). *Anarchy, state, and utopia*. John Wiley & Sons.

Nygaard, S., Hu, H., Li, C., Schiøtt, M., Chen, Z., Yang, Z. & Boomsma, J. J. (2016). Reciprocal genomic evolution in the ant–fungus agricultural symbiosis. *Nature Communications*, *7*(1), 12233.

Ockenfels, A., & Werner, P. (2014). Beliefs and ingroup favoritism. *Journal of Economic Behavior & Organization*, *108*, 453-462.

OECD (2023). Financial indicators, stocks: Private sector debt. OECD.Stat. https://stats.oecd.org/index.aspx?queryid=34814

Oesch, D. (2008). Explaining workers' support for right-wing populist parties in Western Europe: Evidence from Austria, Belgium, France, Norway, and Switzerland. *International Political Science Review*, *29*(3), 349-373.

Ohtsuki, H. & Iwasa, Y. (2006). The leading eight: social norms that can maintain cooperation by indirect reciprocity. *Journal of theoretical biology*, *239*(4), 435-444.

Oliver, J. E., & Wood, T. J. (2014). Conspiracy theories and the paranoid style of mass opinion. *American journal of political science*, *58*(4), 952-966.

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press.

Olson, M. (1993). Dictatorship, democracy, and development. *American Political Science Review*, 87(3), 567-576.

Olson, M. (2000). *Power and prosperity: Outgrowing communist and capitalist dictatorships*. New York: Basic Books.

Opie (1963). *The Theory Of Economic Development. An Inquiry Into Profits, Capital, Credit, Interest, And The Business Cycle*, Oxford: Oxford University Press.

Orbell, J. M., & Dawes, R. M. (1993). Social welfare, cooperators' advantage, and the option of not playing the game. *American sociological review*, 787-800.

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter. *American Political Science Review*, *115*(3), 999-1015.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Ostrom, E. (2000a). Crowding out citizenship. *Scandinavian Political Studies*, *23*(1), 3-16.

Ostrom, E. (2000b). Collective action and the evolution of social norms. *Journal of economic perspectives*, *14*(3), 137-158.

Ostrom, E. (2000c). A behavioral approach to the rational choice theory of collective action. In M. D. McGinnis, *Polycentric games and institutions: readings from the Workshop in Political Theory and Policy Analysis*. University of Michigan Press.

Ostrom, E. (2010). Beyond markets and states: polycentric governance of complex economic systems. *American economic review*, *100*(3), 641-672.

Overbye, E. (1995). Making a case for the rational, self-regarding, 'ethical' voter and solving the 'paradox of not voting' in the process. *European Journal of Political Research*, 27 (3).

Pace, N., & Becchetti, L. (2006). The Economics of the trust game corporation. *Quaderni CEIS; 233*.

Padoa-Schioppa, T. (2004). *Regulating finance: balancing freedom and risk*. Oxford University Press.

Palda, F., & Palda, K. (1998). The impact of campaign expenditures on political competition in the French legislative elections of 1993. *Public choice*, *94*, 157-174.

Palley, T. I. (2013). Financialization: what it is and why it matters. In Palley, T. I. (2013). *Financialization: The Economics of Finance Capital Domination*. Palgrave Macmillan.

Palmer, M., & Schneer, B. (2016). Capitol gains: the returns to elected office from corporate board directorships. *The Journal of Politics*, *78*(1), 181-196.

Pande, R. (2011). Can informed voters enforce better governance? Experiments in low-income democracies. *Annu. Rev. Econ.*, *3*(1), 215-237.

Papagapitos, A., & Riley, R. (2009). Social trust and human capital formation. *Economics Letters*, *102*(3), 158-160.

Papaioannou, K. J., & Van Zanden, J. L. (2015). The dictator effect: how long years in office affect economic development. *Journal of Institutional Economics*, *11*(1), 111-139.

Pareto, V. (2014). *Manual of political economy: A critical and variorum edition*. (A. Zanni, A. Montesano, J. S. Chipman, L. Bruni & M. McLure, Eds.). Oxford University Press. (Original work published 1906).

Patrick, A. (2018, October 26). In the Federal Court, speed of justice depends on the judge. *Australian Financial Review*. https://www.afr.com/companies/professional-

services/in-the-federal-court-speed-of-justice-depends-on-the-judge-20181014-h16mk9

Payne, J.L. & Wagner, A. (2019). The causes of evolvability and their evolution. *Nature Reviews Genetics*, *20*(1), 24-38.

Pearson, D. & Shaw, S. (1993). *Freedom of Informed Choice: FDA Versus Nutrient Supplements*. Neptune, N.J.: Common Sense Press.

Peeters, S. H., & de Jonge, M. I. (2018). For the greater good: Programmed cell death in bacterial communities. *Microbiological research*, *207*, 161-169.

Peltzman, S. (1976). Toward a more general theory of regulation. *The Journal of Law and Economics*, 19(2), 211-240.

Perc, M., Jordan, J. J., Rand, D. G., Wang, Z., Boccaletti, S., & Szolnoki, A. (2017). Statistical physics of human cooperation. *Physics Reports*, *687*, 1-51.

Persaud, A. (2014). Why bail-in securities are fool's gold. *Available at SSRN 2643702*.

Petersen, M. B. (2012). Social welfare as small-scale help: evolutionary psychology and the deservingness heuristic. *American Journal of Political Science*, *56*(1), 1-16.

Pew Research Center. (2014, June 12). Political polarization in the American public: Section 5: Political engagement and activism. https://www.pewresearch.org/politics/2014/06/12/section-5-political-engagement-and-activism/

Peysakhovich, A., Nowak, M. A., & Rand, D. G. (2014). Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nature communications*, *5*(1), 4939.

Peysakhovich, A., & Rand, D. G. (2016). Habits of virtue: Creating norms of cooperation and defection in the laboratory. *Management Science*, *62*(3), 631-647.

Pham-Kanter, G., Mello, M. M., Lehmann, L. S., Campbell, E. G., & Carpenter, D. (2017). Public awareness of and contact with physicians who receive industry payments: a national survey. *Journal of general internal medicine*, *32*, 767-774.

Phan, H. V. (2014). Inside debt and mergers and acquisitions. *Journal of Financial and Quantitative Analysis*, *49*(5-6), 1365-1401.

Philippon, T. (2008). The evolution of the US financial industry from 1860 to 2007: Theory and evidence. NYU Stern Working Paper.

Philippon, T., & Reshef, A. (2012). Wages and human capital in the US finance industry: 1909–2006. *The Quarterly Journal of Economics*, *127*(4), 1551-1609.

Philippon, T. (2015). Has the US finance industry become less efficient? On the theory and measurement of financial intermediation. *American Economic Review*, *105*(4), 1408-1438.

Pienta, K.J., Hammarlund, E.U., Axelrod, R., Brown, J.S. & Amend, S.R. (2020). Poly-aneuploid cancer cells promote evolvability, generating lethal cancer. *Evolutionary Applications*, *13*(7), 1626-1634.

Pierre, A. & Fernandez, A.S. (2018). Going deeper into SMEs' innovation capacity: An empirical exploration of innovation capacity factors. *Journal of Innovation Economics Management*, (1), 139-181.

Pigou, A. C. (1920). *The Economics of Welfare*. London: Macmillan.

Pilster, U., & Böhmelt, T. (2012). Do Democracies Engage Less in Coup-Proofing? On the Relationship between Regime Type and Civil—Military Relations. *Foreign Policy Analysis*, *8*(4), 355-371.

Pinker, S. (1995). *The language instinct: The new science of language and mind* (Vol. 7529). Penguin UK.

Plantin, G. (2015). Shadow banking and bank capital regulation. *The Review of Financial Studies*, *28*(1), 146-175.

Plosser, M. C., & Santos, J. A. (2014). Banks' incentives and the quality of internal risk models. *Available at SSRN 2535856*.

Plotkin, H. (1994). *Darwin machines and the nature of knowledge*. Harvard University Press.

Plumptre, A. F. W. (1947). Keynes in Cambridge. *Canadian Journal of Economics and Political Science/Revue canadienne de economiques et science politique*, *13*(3), 366-371.

Polanyi, K. (2001). *The great transformation: The political and economic origins of our time*. Beacon Press. (Original work published 1944)

Polanyi, M. (1958). Personal knowledge: Towards a post-critical philosophy. University of Chicago Press.

Popper, K. (1945). *The open society and its enemies*. Routledge.

Popper, K. (1972). *Objective Knowledge An Evolutionary Approach*. London: Oxford University Press

Popper, K. (1994). *In search of a better world: Lectures and essays from thirty years*. Psychology Press.

Posner, E. A., & Weyl, E. G. (2018). *Radical Markets*. Princeton University Press.

Poteete, A.R., Janssen, M.A. and Ostrom, E. (2010). *Working together*. Princeton University Press.

Poulsen, M., Erhardt, D. P., Molinaro, D. J., Lin, T. L., & Currie, C. R. (2007). Antagonistic bacterial interactions help shape host-symbiont dynamics within the fungus-growing ant-microbe mutualism. *PLoS One*, *2*(9), e960.

Powell Jr, G. B., & Whitten, G. D. (1993). A cross-national analysis of economic voting: Taking account of the political context. *American journal of political science*, 391-414.

Powell, E. N., & Grimmer, J. (2016). Money in exile: Campaign contributions and committee access. *The Journal of Politics*, *78*(4), 974-988.

Powers, S.T., Taylor, D.J. and Bryson, J.J. (2012). Punishment can promote defection in group-structured populations. *Journal of theoretical biology*, *311*, 107-116.

Pradel, J., Euler, H. A., & Fetchenhauer, D. (2009). Spotting altruistic dictator game players and mingling with them: The elective assortation of classmates. *Evolution and Human behavior*, *30*(2), 103-113.

Prat, A. (2002). Campaign spending with office-seeking politicians, rational voters, and multiple lobbies. *Journal of Economic Theory*, *103*(1), 162-189.

Premo, L. S., & Hublin, J. J. (2009). Culture, population structure, and low genetic diversity in Pleistocene hominins. *Proceedings of the National Academy of Sciences*, *106*(1), 33-37.

Purewal, K., & Haini, H. (2021). Re-examining the effect of financial markets and institutions on economic growth: evidence from the OECD countries. *Economic Change and Restructuring*, 1-23.

Putnam, R. (2000). *Bowling Alone: The collapse and revival of American community*. Simon & Schuster.

Queller, D. C. (1992). A general model for kin selection. *Evolution*, *46*(2), 376-380.

Rajan, R. G. (2006). Has finance made the world riskier?. *European financial management*, *12*(4), 499-533.

Ray, J. L. (1993). Wars between democracies: rare, or nonexistent?. *International Interactions*, *18*(3), 251-276.

Read, A. F. (1994). The evolution of virulence. *Trends in microbiology*, *2*(3), 73-76.

Reed, W. J., & Hughes, B. D. (2002). From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. *Physical Review E*, *66*(6), 067103.

ReMine, W. J. (1993). The biotic message. *St. Paul Science, Saint Paul, MN*.

Ricardo, D. (1817). *On the principles of political economy and taxation*. London: J. Murray.

Rice, T. M., & Sumberg, A. F. (1997). Civic culture and government performance in the American states. *Publius: The Journal of Federalism*, *27*(1), 99-114.

Richerson, P., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., ... & Zefferman, M. (2016). Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behavioral and Brain Sciences*, *39*, e30.

Richerson, P. J., & Boyd, R. (1998). The evolution of human ultra-sociality. *Indoctrinability, ideology, and warfare: Evolutionary perspectives*, 71-95.

Richerson, P. J., & Boyd, R. (2001). The evolution of subjective commitment to groups: A tribal instincts hypothesis. *Evolution and the Capacity for Commitment*, *3*, 186-220.

Richerson, P. J., & Boyd, R. (2008). *Not by genes alone: How culture transformed human evolution*. University of Chicago press.

Richerson, P., & Henrich, J. (2009). Tribal social instincts and the cultural evolution of institutions to solve collective action problems. *Context and the Evolution of Mechanisms for Solving Collective Action Problems Paper*.

Richter, B. K., Samphantharak, K., & Timmons, J. F. (2009). Lobbying and taxes. *American Journal of Political Science*, *53*(4), 893-909.

Ridley, M. (1994). *The red queen: Sex and the evolution of human nature*. Penguin UK.

Ridley, M. (1996). *The Origins of Virtue*. London: Penguin Viking

Riker, W. H., & Ordeshook, P. C. (1968). A Theory of the Calculus of Voting. *American political science review*, *62*(1), 25-42.

Rioja, F., & Valev, N. (2004). Does one size fit all?: a reexamination of the finance and growth relationship. *Journal of Development economics*, *74*(2), 429-447.

Rissing, S., Pollock, G., & Higgens, M. (1996). Fate of ant cofoundresses containing a 'cheater'. *Naturwissenschaften* 83, 182-185.

Robertson, C. B. (2018). Judicial impartiality in a partisan era. *Fla. L. Rev.*, *70*, 739.

Robertson, K. (2023, March 8). 5 times Tucker Carlson privately reviled Trump: 'I hate him'. The New York Times. https://www.nytimes.com/2023/03/08/business/media/tucker-carlson-trump.html

Robinson, J. (1969). *Economics of imperfect competition* (2nd ed.). St Martin's Press.

Robinson, J. A., Torvik, R., & Verdier, T. (2006). Political foundations of the resource curse. *Journal of development Economics*, *79*(2), 447-468.

Robison, G. B. (1960). Rockers and rollers. *Mathematics Magazine,* 33, 139-144.

Rodrik, D. (1999). Where did all the growth go? External shocks, social conflict, and growth collapses. *Journal of economic growth*, *4*, 385-412.

Rodrik, D. (2018a). Populism and the economics of globalization. *Journal of international business policy*, *1*, 12-33.

Rodrik, D. (2018b). Is populism necessarily bad economics?. *AEA Papers and proceedings*. 108, 196-199.

Roig-Franzia, M. (2009, May 31). Brooksley Born warned about the risks of derivatives. The Seattle Times. https://www.seattletimes.com/business/brooksley-born-warned-about-the-risks-of-derivatives/

Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, *98*(5, Part 2), S71-S102.

Romer, P. M. (1996). Preferences, promises, and the politics of entitlement. In *Individual and social responsibility: Child care, education, medical care, and long-term care in America* (195-228). University of Chicago Press.

Romer, P. M. (2010). Process, responsibility, and Myron's Law. IMF Macro Conference Presentation.

https://www.imf.org/external/np/seminars/eng/2011/res/pdf/PMRpresentation.pdf

Romer, P. M. (2016). The Deep Structure of Economic Growth. *Paul Romer blog*, April 13. https://paulromer.net/economic-growth/

Romer, T., & Snyder Jr, J. M. (1994). An empirical investigation of the dynamics of PAC contributions. *American journal of political science*, 745-769.

Roper, M., Ellison, C., Taylor, J.W. & Glass, N.L. (2011). Nuclear and genome dynamics in multinucleate ascomycete fungi. *Current biology*, *21*(18), R786-R793.

Rosenberg, N. (1960). Some institutional aspects of the Wealth of Nations. *Journal of Political Economy*, *68*(6), 557-570.

Rosenberg, S.M. (2001). Evolving responsively: adaptive mutation. *Nature Reviews Genetics*, *2*(7), 504-515.

Ross, J. R. (1967). Constraints on variables in syntax: MIT dissertation. *Institute of Technology, Massachusetts*. https://files.eric.ed.gov/fulltext/ED016965.pdf

Ross, S. A. (1973). The economic theory of agency: The principal's problem. *The American economic review*, *63*(2), 134-139.

Rothstein, B., & Uslaner, E. M. (2005). All for all: Equality, corruption, and social trust. *World politics*, *58*(1), 41-72.

Rothwell, J. T., & Diego-Rosell, P. (2016). Explaining nationalist political views: The case of Donald Trump. *Available at SSRN 2822059*.

Roughgarden, J. (1975). Evolution of Marine Symbiosis--A Simple Cost-Benefit Model. *Ecology*, 56(5), 1201-1208.

Rousseau, J. J. (1985). *The government of Poland*. Hackett Publishing. (Original work published 1782)

Rousseau, P. L., & Wachtel, P. (2011). What is happening to the impact of financial deepening on economic growth?. *Economic inquiry*, *49*(1), 276-288.

Rundquist, B. S., Strom, G. S., & Peters, J. G. (1977). Corrupt politicians and their electoral support: some experimental observations. *American Political Science Review*, *71*(3), 954-963.

Rushton, J. P. (2004). Genetic and environmental contributions to pro-social attitudes: a twin study of social responsibility. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(1557), 2583-2585.

Sabatier, P., Hunter, S., & McLaughlin, S. (1987). The devil shift: Perceptions and misperceptions of opponents. *Western Political Quarterly*, *40*(3), 449-476.

Sacarny, A., Barnett, M. L., Le, J., Tetkoski, F., Yokum, D., & Agrawal, S. (2018). Effect of peer comparison letters for high-volume primary care prescribers of quetiapine in older and disabled adults: a randomized clinical trial. *JAMA psychiatry*, *75*(10), 1003-1011.

Sacarny, A., Yokum, D., Finkelstein, A., & Agrawal, S. (2016). Medicare letters to curb overprescribing of controlled substances had no detectable effect on providers. *Health Affairs*, *35*(3), 471-479.

Salali, G. D., Juda, M., & Henrich, J. (2015). Transmission and development of costly punishment in children. *Evolution and Human Behavior*, *36*(2), 86-94.

Salge, C., Lipski, C., Mahlmann, T. & Mathiak, B. (2008). Using genetically optimized artificial intelligence to improve gameplaying fun for strategical games. In *Proceedings of the 2008 ACM SIGGRAPH symposium on Video games,* 7-14.

Samargandi, N., Fidrmuc, J., & Ghosh, S. (2015). Is the relationship between financial development and economic growth monotonic? Evidence from a sample of middle-income countries. *World development*, *68*, 66-81.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, *5*(17), 61-71.

Samuelson, P.A., & Nordhaus, W.D. (1989). Economics (13th ed.). McGraw-Hill.

Sánchez Rodríguez, Á., Jetten, J., Willis Sánchez, G. B., & Rodríguez Bailón, R. M. (2019). High economic inequality makes us feel less wealthy.

Sanders, L. (2020, October 8). Washington Examiner/YouGov poll: By 47% to 34% voters oppose court packing. YouGov. https://today.yougov.com/topics/politics/articles-reports/2020/10/07/supreme-cort-packing-poll

Saunders, A., Strock, E., & Travlos, N. G. (1990). Ownership structure, deregulation, and bank risk taking. *the Journal of Finance*, *45*(2), 643-654.

Savage, D. (2021, December 31). Chief Justice Roberts calls for better enforcement of conflict laws involving judges' stock ownership. *Los Angeles Times*. https://www.latimes.com/politics/story/2021-12-31/roberts-calls-for-better-enforcement-of-conflict-laws-involving-judges-stock-ownership

Schedler, A. (1999). Conceptualizing accountability. In *The self-restraining state: Power and accountability in new democracies*. Lynne Yienner Publishers.

Schedler, A. (2002). Elections without democracy: The menu of manipulation. *Journal of democracy*, *13*(2), 36-50.

Schularick, M., & Taylor, A. M. (2012). Credit booms gone bust: monetary policy, leverage cycles, and financial crises, 1870–2008. *American Economic Review*, *102*(2), 1029-1061.

Schultz, J. (2021). Disrupting Media and Politics. In J. M. Lewis, & A. Tiernan (Eds.), *The Oxford Handbook of Australian Politics*. Oxford University Press.

Schumpeter, J. A. (1911). *Theorie Der Wirtschaftlichen Entwicklung. Eine Untersuchung Über Unternehmergewinn, Kapital, Kredit, Zins Und Den Konjunkturzyklus*. Translated By R.

Schumpeter, J. A. (1934). The theory of economic development: An inquiry into profits, capital, credit, interest and the business cycle (R. Opie, Trans.). Harvard University Press.

Schumpeter, J. A. (1994). *Capitalism, socialism and democracy*. New York: Routledge. Original work published 1942.

Schwartz, A., & Finighan, R. (2020). Impact Investing Won't Save Capitalism. *Harvard Business Review*, *17*.

Schwartz, M.W. & Hoeksema, J.D. (1998). Specialization and resource trade: biological markets as a model of mutualisms. *Ecology*, *79*(3), 1029-1038.

Sen, A. (2009). Introduction. In A. Smith (2009). The theory of moral sentiments. Penguin. (Original work published 1790, first published 1759)

Sengupta, R., & Xue, F. (2020). The global pandemic and run on shadow banks. *Main Street Views, Federal Reserve Bank of Kansas City, May*, *11*.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*. 19(3), 425–442.

Shen, C. H. H., & Zhang, H. (2020). What's good for you is good for me: The effect of CEO inside debt on the cost of equity. *Journal of Corporate Finance*, *64*, 101699.

Shrestha, M., Dyer, A.G., Dorin, A., Ren, Z.X. & Burd, M. (2020). Rewardlessness in orchids: how frequent and how rewardless?. *Plant Biology*, 22(4), 555-561.

Sidel, J. T. (2007). Riots, pogroms, jihad: Religious violence in Indonesia. NUS Press.

Silk, J. B., & House, B. R. (2016). The evolution of altruistic social preferences in human groups. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1687), 20150097.

Silk, J. B., Brosnan, S. F., Vonk, J., Henrich, J., Povinelli, D. J., Richardson, A. S., ... & Schapiro, S. J. (2005). Chimpanzees are indifferent to the welfare of unrelated group members. *Nature*, *437*(7063), 1357-1359.

Simon, H. A. (1986). Rationality in psychology and economics. *Journal of Business*, S209-S224.

Simon, H. A. (1990). A mechanism for social selection and successful altruism. *Science*, *250*(4988), 1665-1668.

Simon, H. A. (1991). Organizations and markets. *Journal of economic perspectives*, *5*(2), 25-44.

Simon, H. A. (2019). The sciences of the artificial (Reissue of the third edition with a new introduction by J. Laird). MIT Press. (Original work published 1968)

Singer, T., Seymour, B., O'doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, *303*(5661), 1157-1162.

Skidelsky, R. (2006). Hayek versus Keynes: The road to reconciliation. In E. Feser (Ed.), *The Cambridge Companion to Hayek* (82-110). Cambridge University Press.

Sismondo, S. (2021). Epistemic corruption, the pharmaceutical industry, and the body of medical science. *Frontiers in Research Metrics and Analytics*, *6*, 614013.

Smith, A. (1976). An inquiry into the nature and causes of the wealth of nations (R.H. Campbell, A.S. Skinner, & W.B. Todd, Eds.). Oxford University Press. (Original work published 1776)

Smith, A. (2009). The theory of moral sentiments. Penguin. (Original work published 1790, first published 1759)

Smith, S. E., Jakobsen, I. Gronlund, M. & Smith, F. A. (2011). Roles of arbuscular mycorrhizas in plant phosphorus nutrition: interactions between pathways of phosphorus uptake in arbuscular mycorrhizal roots have important implications for understanding and manipulating plant phosphorous acquisition. *Plant Physiol*. 156, 1050–1057.

Snell-Rood, E. (2016). Interdisciplinarity: Bring biologists into biomimetics. *Nature*, 529(7586), 277-278.

Snyder Jr, J. M. (1990). Campaign contributions as investments: The US House of Representatives, 1980-1986. *Journal of Political Economy*, *98*(6), 1195-1227.

Snyder Jr, J. M. (1992). Long-term investing in politicians; or, give early, give often. *The Journal of Law and Economics*, 35(1), 15-43.

Solaz, H., De Vries, C. E., & De Geus, R. A. (2019). In-group loyalty and the punishment of corruption. *Comparative Political Studies*, *52*(6), 896-926.

Solée, R.V., Valverde, S., Casals, M.R., Kauffman, S.A., Farmer, D. & Eldredge, N. (2013). The evolutionary ecology of technological innovations. *Complexity*, 18(4), 15-27.

Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*. 70(1), 65–94.

Soltis, J., Boyd, R., & Richerson, P. J. (1995). Can group-functional behaviors evolve by cultural group selection?: An empirical test. *Current anthropology*, *36*(3), 473-494.

Sonnemans, J., Oosterbeek, H., & Sloof, R. (2001). On the relation between asset ownership and specific investments. *The Economic Journal*, *111*(474), 791-820.

Somer, M. (2001). Cascades of ethnic polarization: Lessons from Yugoslavia. *The Annals of the American Academy of Political and Social Science*, *573*(1), 127-151.

Spencer, H. (1851). *Social statics: Or, the conditions essential to human happiness specified, and the first of them developed*. J. Chapman.

Srivastav, A., Armitage, S., & Hagendorff, J. (2014). CEO inside debt holdings and risk-shifting: Evidence from bank payout policies. *Journal of Banking & Finance*, *47*, 41-53.

Srivastav, A., Armitage, S., Hagendorff, J., & King, T. (2018). Better safe than sorry? CEO inside debt and risk-taking in bank acquisitions. *Journal of Financial Stability*, *36*, 208-224.

Steele, D.R. (1987). Hayek's theory of cultural group selection. *Journal of Libertarian Studies*, *8*(2), 171-95.

Steffensmeier, D., Ulmer, J., & Kramer, J. (1998). The interaction of race, gender, and age in criminal sentencing: The punishment cost of being young, black, and male. *Criminology*, *36*(4), 763-798.

Stein, E. & Lipton, P. (1989). Where guesses come from: Evolutionary epistemology and the anomaly of guided variation. *Biology and Philosophy*, 4, 33-56.

Stelter, B. (2021, November 16). This infamous Steve Bannon quote is key to understanding America's crazy politics. *CNN*. https://edition.cnn.com/2021/11/16/media/steve-bannon-reliable-sources/index.html

Stewart, D. (1980). Account of the Life and Writings of Adam Smith. In A. Smith, *Essays on Philosophical Subjects*. William Wightman,J. C. Bryce, and Ian Ross, eds. Oxford: University Press, 263–351.

Stigler, G. J. (1971). The theory of economic regulation. *The Bell journal of economics and management science*, 3-21.

Stout, L. A. (2001). Judges as altruistic hierarchs. *Wm. & Mary L. Rev.*, *43*, 1605.

Strassmann, J. E., & Queller, D. C. (2011). Evolution of cooperation and control of cheating in a social microbe. *Proceedings of the National Academy of Sciences*, 108(2), 10855-10862.

Stratmann, T. (1998). The market for congressional votes: Is timing of contributions everything?. *The Journal of Law and Economics*, *41*(1), 85-114.

Stratmann, T. (2002). Can special interests buy congressional votes? Evidence from financial services legislation. *The Journal of Law and Economics*, *45*(2), 345-373.

Stratmann, T. (2004). The effectiveness of money in ballot measure campaigns. *S. Cal. L. Rev.*, *78*, 1041.

Stratmann, T. (2005). Some talk: Money in politics. A (partial) review of the literature. *Policy challenges and political responses: Public choice perspectives on the post-9/11 world*, 135-156.

Suarez, S., & Kolodny, R. (2011). Paving the road to 'too big to fail': Business interests and the politics of financial deregulation in the United States. *Politics & Society*, *39*(1), 74-102.

Summers, L. H. (2015). Demand side secular stagnation. *American economic review*, *105*(5), 60-65.

Sundaram, R. K., & Yermack, D. L. (2007). Pay me later: Inside debt and its role in managerial compensation. *The Journal of Finance*, *62*(4), 1551-1588.

Sunstein, C. R. (Ed.). (2018). Can it happen here?: Authoritarianism in America. Dey Street Books.

Svolik, M. (2008). Authoritarian reversals and democratic consolidation. *American Political Science Review*, *102*(2), 153-168.

Swan, T. W. (1956). Economic growth and capital accumulation. *Economic Record*. 32(2), 334–361.

Swift, J. (1710, November 9). The Art of Political Lying. *Examiner*, 14, para. 9. Reprinted in Nordquist, R. (2016, March 20). *The Art of Political Lying, by Jonathan Swift*. ThoughtCo. https://www.thoughtco.com/art-of-political-lying-by-swift-1690138

Szathmáry, E. & Maynard Smith, J. (1995). The major evolutionary transitions. *Nature*, *374*(6519), 227-232.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American journal of political science*, *50*(3), 755-769.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European journal of social psychology*, *1*(2), 149-178.

Tan, J., & Hare, B. (2013). Bonobos share with strangers. *PLoS One*, *8*(1), e51922.

Mokyr, J. (1992). Technological inertia in economic history. *The Journal of Economic History*, *52*(2), 325-338.

Thaler, R. (2015). *Misbehaving*. New York: W.W. Norton & Company, Inc.

Thielmann, I., & Hilbig, B. E. (2015). The traits one can trust: Dissecting reciprocity and kindness as determinants of trustworthy behavior. *Personality and Social Psychology Bulletin*, *41*(11), 1523-1536.

Thomas, M. (2007). A review of developments in the Job Network. Research paper no. 15, 2007–08. Australian Parliament.

Thöni, C., Tyran, J. R., & Wengström, E. (2012). Microfoundations of social capital. *Journal of Public Economics*, *96*(7-8), 635-643.

Tilly, C. (2017). War making and state making as organized crime. In *Collective violence, contentious politics, and social change* (121-139). Routledge.

TIME (1924, December 15). Lazy judges. *TIME Magazine*. https://content.time.com/time/subscriber/article/0,33009,728221,00.html

Tirole, J. (1986). Hierarchies and bureaucracies: On the role of collusion in organizations. *The Journal of Law, Economics, and Organization*, *2*(2), 181-214.

Titmuss, R. (2018). *The gift relationship: From human blood to social policy*. Policy Press. (Original work published 1970)

Travisano, M., & Velicer, G. J. (2004). Strategies of microbial cheater control. *Trends in microbiology*, *12*(2), 72-78.

Tripathi, M., Ansolabehere, S., & Snyder, J. M. (2002). Are PAC contributions and lobbying linked? New evidence from the 1995 Lobby Disclosure Act. *Business and politics*, *4*(2), 131-155.

Tullock, G. (1967). The welfare costs of tariffs, monopolies and theft. Western Economic Journal 5: 224-232.

Tung, F., & Wang, X. (2012). Bank CEOs, inside debt compensation, and the global financial crisis. *Boston Univ. School of Law Working Paper*, (11-49).

Turchin, P. (2009). A theory for formation of large empires. *Journal of Global History*, *4*(2), 191-217.

Turner, A. (2010). What do banks do? Why do credit booms and busts occur and what can public policy do about it. *The Future of finance*, *5*.

Turner, R. H. (1964). Collective behavior. In R. E. L. Farris (Ed.), Handbook of modern sociology (397–409). Rand McNally.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, *5*(2), 207-232.

Tversky, A., & Kahneman, D. (1989). Rational choice and the framing of decisions. In *Multiple criteria decision making and risk analysis using microcomputers* (81-126). Springer Berlin Heidelberg.

U.S. Senate. (1934, June). Stock Exchange Practices: Report of the Committee of Banking and Currency (Senate Report No. 1455). 73rd Congress, 2nd session. https://www.senate.gov/about/resources/pdf/pecora-final-report.pdf

U.S. Senate. (2007, June 27). Paid to prescribe? Exploring the relationship between doctors and the drug industry. Special Committee on Aging. https://www.aging.senate.gov/imo/media/doc/6272007.pdf

Ugur, M., & Dasgupta, N. (2011). Evidence on the economic growth impacts of corruption in low-income countries and beyond: a systematic review. *EPPI-Centre Social Science Research Unit, Institute of Education, University of London*, 1-134.

Uslaner, E. M. (2002). The moral foundations of trust. *Available at SSRN 824504*.

Uslaner, E. M. (2004). Trust and corruption. In *The new institutional economics of corruption* (90-106). Routledge.

Valckx, N., Chan-Lau, J. A., Feng, A., Huston, B., Impavido, G., Jobst, A. A., ... & Yan, K. (2016). The Insurance Sector-Trends and Systemic Risk Implications. *Global Financial Stability Report, IMF*.

Van Bekkum, S. (2016). Inside debt and bank risk. *Journal of Financial and Quantitative Analysis*, *51*(2), 359-385.

Van Valen, L. (1973). A new evolutionary law. Evol Theory, 1, 1-30.

VanSickle, A. (2023, April 7). Justice Thomas says he was advised lavish gifts did not need to be reported. The New York Times. https://www.nytimes.com/2023/04/07/us/politics/clarence-thomas-supreme-court-travel.html

Vautrin, E. & Vavre, F. (2009). Interactions between vertically transmitted symbionts: cooperation or conflict?. *Trends in microbiology*, *17*(3), 95-99.

Veblen, T. (1898). Why is economics not an evolutionary science?. The Quarterly Journal of Economics, 12(4), 373-397.

Veblen, T. (1909). The limitations of marginal utility. *Journal of political Economy*, *17*(9), 620-636.

Verba, S., Schlozman, K. L., Brady, H. E., & Nie, N. (1990). American citizen participation study. ICPSR version available at *www.icpsr.umich.edu/cgi/archive.prl*

Voegeli, V. (2016, March 31). Credit Suisse had no blind spots on writedowns, Rohner says. *Bloomberg*. https://www.bloomberg.com/news/articles/2016-03-31/credit-suisse-had-no-blind-spots-on-illiquid-assets-rohner-says

Volcker, P. (2009). in Murray, A., 2009. Paul Volcker: think more boldly. *Wall Street Journal*, *14*.

Von Neumann J., & Morgenstern, O (1953). *The Theory of Games and Economic Behavior*. Third edition. Princeton University Press.

Vonk, J., Brosnan, S. F., Silk, J. B., Henrich, J., Richardson, A. S., Lambeth, S. P., ... & Povinelli, D. J. (2008). Chimpanzees do not take advantage of very low cost opportunities to deliver food to unrelated group members. *Animal Behaviour*, *75*(5), 1757-1770.

Walker, L., & Mann, L. (1987). Unemployment, relative deprivation, and social protest. *Personality and Social Psychology Bulletin*, *13*(2), 275-283.

Waller, M. J., & Yasmann, V. J. (1995). Russia's great criminal revolution: the role of the security services. *Journal of Contemporary Criminal Justice*, *11*(4), 276-297.

Walter, B. F. (2022). How civil wars start: And how to stop them. Crown.

Wang, C., Xie, F., & Xin, X. (2018). CEO inside debt and accounting conservatism. *Contemporary Accounting Research*, *35*(4), 2131-2159.

Wang, M., Schaefer, A. L., Dandekar, A. A., & Greenberg, E. P. (2015). Quorum sensing and policing of Pseudomonas aeruginosa social cheaters. *Proceedings of the National Academy of Sciences*, *112*(7), 2187-2191.

Wang, M., Schaefer, A.L., Dandekar, A.A. & Greenberg, E.P. (2015). Quorum sensing and policing of Pseudomonas aeruginosa social cheaters. *Proceedings of the National Academy of Sciences*, *112*(7), 2187-2191.

Wang, Q.J., Feng, G.F., Wang, H.J. & Chang, C.P., 2021. The impacts of democracy on innovation: Revisited evidence. *Technovation*, *108*, p.102333.

Wang, Z., Jetten, J., & Steffens, N. K. (2019). Do people want more wealth and status in unequal societies?. In *The social psychology of inequality* (289-303). Cham: Springer International Publishing.

Wang, Z., Walker, G. W., Muir, D. C., & Nagatani-Yoshida, K. (2020). Toward a global understanding of chemical pollution: a first comprehensive analysis of national and regional chemical inventories. *Environmental science & technology*, *54*(5), 2575-2584.

Wantchekon, L. (2003). Clientelism and voting behavior: Evidence from a field experiment in Benin. *World politics*, *55*(3), 399-422.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *science*, *311*(5765), 1301-1303.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy*, *11*(3), 271-294.

Warneken, F., & Tomasello, M. (2008). Extrinsic rewards undermine altruistic tendencies in 20-month-olds. *Developmental psychology*, *44*(6), 1785.

Wawro, G. (2001). A panel probit analysis of campaign contributions and roll-call votes. *American Journal of Political Science*, 563-579.

Weber, M. (2009). The theory of social and economic organization. Simon and Schuster.

Weeks, A.R., Turelli, M., Harcombe, W.R., Reynolds, K.T. & Hoffmann, A.A., 2007. From parasite to mutualist: rapid evolution of Wolbachia in natural populations of Drosophila. *PLoS biology*, *5*(5), p.e114.

Wei, C., & Yermack, D. (2011). Investor reactions to CEOs' inside debt incentives. *The Review of Financial Studies*, *24*(11), 3813-3840.

Weimer, W. (2013). Advances in automated program repair and a call to arms. In *Search Based Software Engineering: 5th International Symposium*. Proceedings 5 (1-3). Springer Berlin Heidelberg.

Wells, H. G. (1945). *The mind at the end of its tether*. William Heinemann Ltd.

West, S.A., Kiers, E.T., Pen, I. & Denison, R.F. (2002). Sanctions and mutualism stability: when should less beneficial mutualists be tolerated?. *Journal of Evolutionary Biology*, *15*(5), 830-837.

White, L. (2013). Krugman on Friedman, Hayek, and Liquidationism, *Alt-M: Ideas for an Alternative*.

Whitehead, H., & Rendell, L. (2014). *The cultural lives of whales and dolphins*. University of Chicago Press.

Whitehead, H., Richerson, P. J., & Boyd, R. (2002). Cultural selection and genetic diversity in humans. *Selection*, *3*(1), 115-125.

Whiteley, P. F. (2000). Economic growth and social capital. *Political studies*, *48*(3), 443-466.

Wilcox, A.J., Weinberg, C.R., O'Connor, J.F., Baird, D.D., Schlatterer, J.P., Canfield, R.E., Armstrong, E.G. & Nisula, B.C. (1988). Incidence of early loss of pregnancy. *New England Journal of Medicine*, *319*(4), 189-194.

Wilke, C.O., Wang, J.L., Ofria, C., Lenski, R.E. & Adami, C. (2001). Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, *412*(6844), 331-333.

Wilson, D. S., Hayes, S. C., Biglan, A., & Embry, D.D. (2014). Evolving the future: Toward a science of intentional change. Behavioral and Brain Sciences, 37(4), 395-416.

Wilson, D.S. (2003). *Darwin's Cathedral: Evolution, Religion and the Nature of Society*. University of Chicago Press:

Wilson, D.S., Hayes, S.C., Biglan, A. & Embry, D.D. (2014). Evolving the future: Toward a science of intentional change. *Behavioral and Brain Sciences*, *37*(4), 395-416.

Wilson, E.O. (1975). *Sociobiology: The new synthesis*. Harvard University Press.

Wilson, J. (2021, October 1). Top Republicans rub shoulders with extremists in secretive rightwing group, leak reveals. *The Guardian*. https://www.theguardian.com/us-news/2021/sep/30/revealed-council-national-policy-republicans-extremists

Wimsatt, W.C. & Schank, J.C., 1988. Two constraints on the evolution of complex adaptations and the means for their avoidance. *Evolutionary progress, ed. M. Nitecki. University of Chicago Press.[WCW].*

Winkler, H. A. (2015). *The Age of Catastrophe: A History of the West 1914–1945*. Yale University Press.

Winter, M., Schweiger, O., Klotz, S., Nentwig, W., Andriopoulos, P., Arianoutsou, M., ... & Kühn, I. (2009). Plant extinctions and introductions lead to phylogenetic and taxonomic homogenization of the European flora. *Proceedings of the National Academy of Sciences*, *106*(51), 21721-21725.

Winters, M. S., & Weitz-Shapiro, R. (2013). Lacking information or condoning corruption: When do voters support corrupt politicians?. *Comparative Politics*, *45*(4), 418-436.

Wintrobe, R. (2001). How to understand, and deal with dictatorship: an economist's view. *Economics of Governance*, *2*, 35-58.

Witko, C. (2006). PACs, issue context, and congressional decisionmaking. *Political Research Quarterly*, *59*(2), 283-295.

Wolf, M. (2010, April 21). The challenge of halting the financial doomsday machine. Financial Times. https://www.ft.com/content/f2e4dbb0-4caa-11df-9977-00144feab49a

Wolf, M. (2010, September 14). Basel: the mouse that did not roar. Financial Times.

Wolf, M. (2014). *The Shifts and the Shocks: What we've Learned–and have Still to Learn–from the Financial Crisis*. Penguin UK.

Womald, B. (1993) *Francis Bacon: History, Politics and Science, 1561–1626*. Cambridge University Press.

Woollacott, E. (2017, September 9). Amazon's fake review problem is now worse than ever, study suggests. *Forbes*.

Wright, J. R. (1990). Contributions, lobbying, and committee voting in the US House of Representatives. *American Political Science Review*, *84*(2), 417-438.

Wright, S. A., Dinsmore, J. B., & Kellaris, J. J. (2013). How group loyalties shape ethical judgment and punishment preferences. *Psychology & Marketing*, *30*(3), 203-210.

Wyatt, G.A., Kiers, E.T., Gardner, A. & West, S.A. (2014). A biological market analysis of the plant-mycorrhizal symbiosis. *Evolution*, *68*(9), 2603-2618.

Yamagishi T. (1995). Social dilemmas. In *Sociological Perspectives on Social Psychology*, ed. K Cook, GA Fine, JS House, 311–35. Boston: Allyn & Bacon

Yamagishi, T., & Kiyonari, T. (2000). The group as the container of generalized reciprocity. *Social Psychology Quarterly*, 116-132.

Yamagishi, T., Jin, N., & Kiyonari, T. (1999). Bounded generalized reciprocity: Ingroup boasting and ingroup favoritism. *Advances in group processes*, *16*(1), 161-197.

Yamamura, N. (1993). Vertical transmission and evolution of mutualism from parasitism. *Theoretical Population Biology*, *44*(1), 95-109.

Yeh, J. S., Franklin, J. M., Avorn, J., Landon, J., & Kesselheim, A. S. (2016). Association of industry payments to physicians with the prescribing of brand-name statins in Massachusetts. *JAMA internal medicine*, *176*(6), 763-768.

Yek, S. H., Boomsma, J. J., & Poulsen, M. (2012). Towards a better understanding of the evolution of specialized parasites of fungus-growing ant crops. *Psyche*, *2012*.

You, J. S. (2018). Trust and corruption. *The Oxford handbook of social and political trust*, 473-496.

Yu, F., & Yu, X. (2011). Corporate lobbying and fraud detection. *Journal of Financial and Quantitative Analysis*, *46*(6), 1865-1891.

Yudkin, D. A., Rothmund, T., Twardawski, M., Thalla, N., & Van Bavel, J. J. (2016). Reflexive intergroup bias in third-party punishment. *Journal of experimental psychology: general*, *145*(11), 1448.

Zak, P. J., & Knack, S. (2001). Trust and growth. *The economic journal*, *111*(470), 295-321.

Zak, P.J. & Denzau, A.T. (2001). Economics is an evolutionary science. Evolutionary approaches in the behavioral sciences: Toward a better understanding of human nature, 31-65.

Zech, J.R., Badgeley, M.A., Liu, M., Costa, A.B., Titano, J.J. & Oermann, E.K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS medicine*, *15*(11), p.e1002683.

Zendle, D., Meyer, R., Cairns, P., Waters, S., & Ballou, N. (2020). The prevalence of loot boxes in mobile and desktop games. *Addiction*, *115*(9), 1768-1772.

Zettler, I., & Hilbig, B. E. (2010). Honesty–humility and a person–situation interaction at work. *European Journal of Personality*, *24*(7), 569-582.

Zettler, I., Hilbig, B. E., & Heydasch, T. (2013). Two sides of one coin: Honesty–Humility and situational factors mutually shape social dilemma decision making. *Journal of Research in Personality*, *47*(4), 286-295.

Zhukov, Y. M., & Talibova, R. (2018). Stalin's terror and the long-term political effects of mass repression. *Journal of Peace Research*, *55*(2), 267-283.

Zirin, J. (2022, April 1). Has the Supreme Court lost its compass? The Hill. https://thehill.com/opinion/judiciary/588032-has-the-supreme-court-lost-its-compass/