London School of Economics and Political Science

The Foundations of Theoretical Finance

Stuart Theobald

A thesis submitted to the Department of Philosophy, Logic and Scientific Method
of the London School of Economics and Political Science
for the degree of Doctor of Philosophy
September 2016

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work.

I declare that my thesis consists of 64 873 words.

Parts of this thesis were copy edited for conventions of language, spelling and grammar by Olivia Magennis and Colin Anthony.

Stuart Theobald

Abstract

This thesis provides an account of the ontological, methodological and epistemological foundations of theoretical finance. I argue that these are linked: financial theory is not just a positive enquiry into the nature of the world, but also a means to engineer it. A core feature of its methodology is a modelling approach that conceives of risk as being identical with the volatility of financial market returns. I argue that this can be justified from the perspective of finance as a positive science, because it successfully points to some real features by using idealisations and simplifications. However, I argue that this feature makes theoretical finance inadequate for the task of designing institutions in certain specific ways. I argue that the epistemic demands we make of models in the service of finance theory are different to the demands we should make of models that are used as blueprints for institutions. A failure to appreciate this difference contributed to the financial crisis and our failure to anticipate it.

Finance theory has various effects on the financial system so is causally caught up in social ontology, a notion known as "performativity". I argue that financial theory and its models affect the workings of the financial system in three ways: they provide positive accounts of the financial system that individuals can learn from, they provide normative calculative devices to determine optimal decisions, and finally they provides blueprints for the design of financial institutions. We have in the past assumed that models that succeed in supporting financial theories because of their coherence and validation by data will also be successful blueprints. However, good tests of blueprints should be inductive and assess for unexpected properties, quite unlike how we test theory models. It follows that we have insufficient reason to assume good theory models are good blueprints.

Contents

Acknowledgements

Thank you firstly to my supervisor Jason Alexander who has been a source of considerable guidance in the development of this thesis, exercising utmost patience as I have grappled with the demands of producing it. More recently, Luc Bovens stepped in as a second supervisor and provided an important additional perspective and many helpful comments. At the beginning of the process, Nancy Cartwright served as my second supervisor and disciplinarian on the importance of rigorous philosophical method. All of them have been crucial in my development as a philosopher, though I have so much more to learn.

Thanks also to Bryan Roberts, Katie Steel, Mike Otsuka, Armin Schulz, and other members of staff who have served as helpful sounding boards as my ideas have developed as well as guidance of a more pastoral nature. Outside of the department, Daniel Beunza in the Department of Management and David Webb in the Department of Finance provided helpful input on some of my ideas. Thanks also to Roy Havemann for inviting me to present some of the thoughts in this thesis to a research seminar at South Africa's National Treasury.

Many of my colleagues in the PhD cohort have provided helpful comments. Thanks particularly to Nicolas Weuthrich who provided extensive feedback on two chapters. In the course of various seminars with colleagues at the LSE, and in the corridors around them, I've received helpful input from James Nguyen, Orri Stefánsson, Johannes Himmelreich, Susanne Burri, Seamus Bradley, Chlump Chatkupt, Esha Senchaudhuri, Kamilla Buchter, Fernando Morrett, Catherine Greene, Chris Blunt, Roberto Fumagalli and Katherine Furman.

During the first two years of my PhD I benefitted from a Commonwealth Scholarship funded by the UK Department of International Development which I gratefully acknowledge. For the last four years this thesis was completed on a part time basis. This means the opportunity cost was shared by my colleagues at Intellidex who have had to carry the additional weight caused by my divided attention, for which I am grateful. One of those colleagues, Colin Anthony, helpfully proofread two chapters.

Thanks go also to my parents Lois Theobald and Graham Theobald for their encouragement throughout. Finally thank you to my fiancée Olivia Magennis for her encouragement and patience with my frequent distractions, and willingness to engage my philosophical musings. She also bravely volunteered to proof read this thesis and made many helpful corrections. Of course, all remaining errors are mine alone.

Chapter 1
Introduction

1. Setting the scene

The seed for this thesis was planted in 2008 during the global financial crisis. At the time I was preparing to take the exams for the chartered financial analyst (CFA) qualification, requiring the rote learning of a large body of financial models, formulae, and details about institutions of the financial industry. That task was made more difficult by the stream of news about the chaos enveloping the financial system, which seemed so starkly in contrast to the elegant, mathematically complete models I was studying in the textbooks. Recalling my undergraduate philosophy of science lessons from many years earlier, I felt sure we were witnessing falsification writ large, a methodological crisis that would induce a major paradigm shift in theorising about finance. At times I doubted the utility of studying the standard body of knowledge that I was absorbing, concerned that by the time it came to practice it would have changed beyond recognition.

I remember an exasperated hedge fund manager who had just lost the capital of his fund describe the calamity as a “one in 10,000 year event” on a financial news channel. At one level this was an obviously ridiculous claim – the financial system as we know it for the most part is not even 50 years old, let alone 10,000, and we do not have any reason to assume it will continue to exist in a remotely similar form in another ten millennia. But at another level I knew from my studies that his claim was backed by state-of-the-art risk management tools. One of these is Value-at-Risk modelling which extrapolates from the historic volatility and correlations of the assets in a portfolio to determine the probability of losses. I was sure that hedge fund manager had relied on such analysis to determine that a daily loss of sufficient magnitude to result in bankruptcy would happen only once in 10,000 years. I knew from my text books that this was standard practice across the industry, from fund managers through to banking regulators employed by central banks, and allowed them to sleep at night, comfortable that the system was not too risky.

Those early thoughts graduated to puzzlement as the main bodies presiding over the financial profession grappled with how to deal with the financial crisis in the years that followed. The CFA qualification claims to be the “most respected and recognized investment management designation in the world” and has 120,000 holders working across the investment industry (CFA Institute 2016). Thomas Kuhn may as well have been writing about it and the many similar investment qualifications produced by business schools when he described periods of normal science as “a strenuous and devoted attempt to force nature into the conceptual boxes supplied by professional education” (Kuhn 1962: 5). The response from such bodies following the crisis appeared to be robust but clung valiantly to the conceptual foundations that had preceded the crisis. In a monograph reviewing the crisis, the Research Foundation of the CFA Institute considered the role of Value-at-Risk modelling, declaring, “Rather than blame this measure for its failure to identify the possibility of a financial crisis, however, one must instead blame the way that risk measures were (or were not) used by investors, their advisers, and asset managers” (Fabozzi, Focardi, and Jonas 2010: 2). So the blame was directed at the users, not the models. This was of course personally convenient for me, as I imagine it was for the rest of the profession, who need not discard the tools of their trade. Having qualified, I could happily go on to apply the techniques I had been schooled in, having now apparently learnt to apply them “correctly”. Thankfully, I could still find gainful employment in the market place.

Perhaps more surprising was the role of official government-backed reports on the financial crisis. The US Congress commissioned the Financial Crisis Inquiry Commission which delivered a 668-page report on the crisis in 2011 (Financial Crisis Inquiry Commission 2011). Nowhere in that report did it raise the possibility that there were problems with the foundational models of finance theory. Instead it blamed individuals for the inappropriate use of theory, declaring that "The crisis was the result of human action and inaction, not of Mother Nature or computer models gone haywire" (Financial Crisis Inquiry Commission 2011: xvii). It also focused much attention on the engineering of the financial system, such as the role of ratings agencies and over-the-counter derivatives which were hidden from view.

In contrast, the UK's equivalent, The Turner Review (Turner 2009), did have some things to say about some foundational concepts in finance. It argued that the crisis "raises important questions about the intellectual assumptions on which previous regulatory approaches have largely been built", adding, "At the core of these assumptions has been the theory of efficient and rational markets" (Turner 2009: 39). In discussing the increased use of computers to capture relationships between market prices and the effects of diversification, Turner wrote "the underlying methodological assumption was straightforward: the idea that analysis of past price movement patterns could deliver statistically robust inferences relating to the probability of price movements in future". He then criticises this approach in strong terms (Turner 2009: 44):

> The financial crisis has revealed, however, severe problems with these techniques. They suggest at very least the need for significant changes in the way that [Value-at-Risk]-based methodologies have been applied: some, however, pose more fundamental questions about our ability in principle to infer future risk from past observed patterns.

Unlike the implicit defence of foundational theories seen in the profession's own research efforts, here was a direct challenge to some of the core theoretical precepts of the discipline, particularly the use of past data to determine future risk. But

Turner's recommendations, like his American counterparts', focused on the engineering of the system, calling for more investment in properly skilled regulatory personnel and more intensive surveillance of financial markets and institutions. This clearly reflected the politics of the time (Engelen 2011). Turner had been appointed head of the Financial Services Authority some months before producing the report. The institution was then battling for survival amid severe criticism for having failed to anticipate the crisis. An emphasis on the importance of oversight and strong institutional capabilities for the regulator served his purposes. While he was willing to challenge the conceptual heart of finance theory, this was in effect to shift the blame from the people working within the programme. His solution was more resources, particularly better paid and trained people.

My previous education in philosophy of science had prepared me to expect far more focus on the theories of finance and what the crisis meant for their reliability. Instead, vested interests and politics seemed to play an outsized role. Of course, sociological factors had been well discussed in the philosophy I knew, but the application of this and other methodological concerns in the literature was absent from the broader debates that followed the crisis. The logic embodied in financial theory was not appraised, particularly in respect of how financial theory should respond to empirical issues. Indeed, philosophical questions of relevance to the financial system, and practices linked to it, just didn't seem that important.

David Johnstone writes that "there has been a cultural shift away from critical philosophical analysis of financial logic and financial methods within finance" (2013: 2). There has been little consideration of just what core beliefs underpinned the main financial theories of the time, let alone any consideration of alternative approaches that might suggest a switch to new research programmes or a paradigm shift within the research community at large.

The seed germinated and the ambition grew to make a contribution to closing this gap, leading to the thesis you now hold. It represents an effort to come to terms with the way theoretical finance works and how it affects the institutions that form the financial system. The arguments I have developed show that my initial philosophical thoughts on the consequences of the crisis were naïve. The crisis has been damaging to theoretical finance, but this was mostly because financial theory played an inappropriate role in institutional design. The theory can still be good science without being good engineering. That is the key the argument of this thesis.

2. Seeking out foundations

One might think that the foundations of a discipline are among the most secure concepts that practitioners share. But in the everyday activity of those who operate within a discipline, the focus is usually on producing the next bit of work, meeting the next deadline, and seldom reflecting on what the foundational set of beliefs actually is. As a result, practice can divorce itself from its foundations, developing methods without much thought about the foundational assumptions they entail. Theoretical finance, as developed by banks, investors, regulators and the academy does not often involve much consideration of its foundations. But questions about the foundations of a scientific discipline, as Leonard Savage put it in his seminal *Foundations of Statistics*, have “a certain continuing importance as the science develops, influencing, and being influenced by, the more immediately practical parts of the science” (Savage 1954: 1). This thesis is a work of practical philosophy, in which the ambition is to show how the financial crisis might enable new insights into the foundations of theoretical finance, but also how questions about its foundations might influence practice. While the crisis plays only a background role in my arguments, it serves to open new avenues of exploration in an effort to understand what exactly the foundations of theoretical finance consist of.

Not everything needs to have a clear set of foundations. Our everyday activities happen at a level that ignores the foundational beliefs that underpin them. We may pause when confronted by contradictions in our thinking, but most people quite happily go about their lives without the effort of determining what the foundations of their worldviews are. Nevertheless, questions about the foundations of our knowledge are as old as philosophy. Aristotle demanded that we should identify and commit to a set of basic beliefs in order to escape an infinite regress of reasons for every reason. Descartes pushed the search for foundations to the extreme with his *cogito ergo sum*. Foundationalism of this sort has remained important in philosophy since, including during the development of the philosophy of science over the last century.

When it comes to a body of research like theoretical finance, questions about its foundations are in part empirical. Just what set of beliefs about the way the world is underpins the work of theorists? This calls for an investigation of the commitments of practitioners, an exploration of what their worldviews are. This thesis does not attempt such an empirical exercise, but instead brings to the fore the commitments that are implied by theoretical finance, the beliefs about the world one must have in order for the theories to be useful. This is different to the question of what the financial world in fact consists of. This second question is important because it allows us to appraise the ontological commitments of the programme of theoretical finance. In light of the dialectic between practice and questions about foundations, there is an evaluative task ahead; a question of whether the foundations are the right ones. The commitments of financial theorists have to be teased out by considering some of the seminal theories of finance, but in order to make sense and evaluate those foundations, I will provide some arguments about how social institutions, like the financial system, actually exist.

Theoretical finance occupies a somewhat unique status in that it is both a positive science and a constructive one. It tries to develop theories about how the world works, but it is also actively engaged in shaping the institutions that form part of the financial system. Questions about its foundations are therefore both methodological and ontological. Methodological questions are about how it reasons and develops truth claims about the world. Ontological questions are about what really is in the world. These two questions are intertwined with a further question: how does knowledge of the financial system affect what it is?

3. The arc of the argument

This last question underpins this thesis. I start by considering the methodology of finance but this leads to questions over its ontological commitments. I argue that theoretical finance rests on an ontological commitment to the world being one that generates stable probability distributions, so that risk can be measured as the variance of returns in financial markets. That commitment is justified from the standpoint of operationalism, which is a philosophical view that takes measurements to be the only proper ontological content for a science.

Finance is positive and constructive, but it is also normative. It is a variety of normative microeconomics, a toolbox to be used by individual investors to decide what investment decisions they should make. This is another way that finance changes the world, creating decision rules that tell people what behaviour is optimal. This threefold character of finance theories leads to a complex influence on financial institutions, providing arguments for how institutions may be expected to work and arguments for how they *should* work.

In appraising these foundations there are many questions that have to be answered. One is whether the world really works the way the positive methodology of theoretical finance depends on it working. To make sense of that question I develop an account of the ontology of the financial system. This allows me to assess the ontological commitments of theoretical finance, but it also supports an exploration of how finance affects the world, for the ontology of the financial system is certainly not independent of theoretical finance. My argument is that the financial system is dynamic and constantly in process. It exhibits feedback mechanisms between the institutional facts that emerge from the people and institutions that form components of it. Among those components are theories about what people do and should do in making financial decisions. Those theories inform deliberate efforts to shape institutions in society; efforts that meet with mixed success. The failure of institutions drives new research questions and informs future institutional designs.

The methods that financial theorists use in such theoretical and design work differ depending on whether their purpose is to describe mechanisms that are claimed to exist, or to develop blueprints for new institutions. Theories about the financial system use models to explain and understand properties of it, but these are best seen as "socioeconomic machines", shielded theoretical worlds that exhibit the theorised property. They can be appraised in terms of their internal coherence, success as rhetorical devices in explaining tendencies in society and empirical corroboration, although the last aspect is difficult given the complexities present in any real world example. On the other hand, blueprints for institutions need to focus on detail, spelling out regulatory measures, staffing, professional codes and much else that an institution will need in order to function. Blueprints are certainly informed by socioeconomic machines, but their appraisal is different – it is how they actually work as guides to developing successful new institutions. We can test how they will work by using scale models of the institutions, using game theoretic assessments, randomised controlled trials, simulations, and other experimental methods I will discuss. But the clear point is that theory work and design work in finance are different. I argue that this has been confused in the past, particularly in the example of risk and its identification with variance, which began its role as an idealised simplification, evolved into an operationalist ontological commitment, and culminated as a design feature of institutions. This was ultimately a result of confusing a positive theory with a blueprint for design. I conclude this thesis by attributing some of the blame for the financial crisis to this error.

4. “Theoretical finance”

There are many ways to define what theoretical finance is. My approach is mostly Lakatosian, in that I see theoretical finance as a particular research programme. This programme consists of a group of theories that is often described as “modern portfolio theory” (Fabozzi, Gupta, and Markowitz 2002). I will focus on two exemplars of this programme, the Mean Variance Optimisation (MVO) theory of portfolio selection developed by Harry Markowitz (1952, 1959) and the Capital Asset Pricing Model (CAPM) developed by William Sharpe (1964) and John Lintner (1965). These models provide the foundations for other important work in the programme including the efficient markets hypothesis and random walk theory, the capital irrelevance theorem, arbitrage pricing theory, the Black-Scholes model for options pricing, and Value-at-Risk modelling. Of course, many theorists working in finance would object to this characterisation, pointing to the great deal of other work outside this programme such as behavioural finance, game theory, and multifactor models. I will at times consider aspects of this work too, but I consider these outside the programme of theoretical finance as I conceive of it. As I will discuss, some of this work is directly threatening to the core commitments of the theoretical finance programme and may present alternative approaches that will grow in importance.

I will argue in Chapter 2 of this thesis that the theoretical finance programme is unified by a particular set of ontological entailments. As Uskali Mäki puts it, “The contents and confines of the economic realm are suggested by the ontological commitments that economists hold. A set of such commitments has the nature of a *Weltanschauung*, a fundamental world view with a focus on a selection of economically relevant aspects of the world” (Mäki 2001: 4-5). In the case of financial theories, an important ontological commitment is that the world is such that it generates stable probability distributions. This allows the theory to define risk as variance in asset prices. Finance theorists work on the basis that risk is identical with variance, so that the variance of returns to assets is taken as an entirely sufficient concept of risk. This is done via operationalism, a philosophical position stemming from Percy Bridgman (1927), by which concepts are defined by their measurement. This strategy has allowed theoretical finance to work within a frequentist philosophy of probability, which I describe as a "hard core" commitment in the programme.

The term "hard core" comes from Lakatosian analysis, though traditionally it refers to a theoretical claim that is considered central to a scientific research programme. I show that this approach doesn't work in theoretical finance, because its central theoretical claims fail to distinguish it from any number of other programmes within economics. However, its ontological commitments and notion of risk, do delineate it.

As the architecture of the modern financial system was assembled using this theoretical approach, this operational concept of risk has been built into its institutions. I consider and dismiss an argument that in this process a particular notion of risk, along with the rest of the commitments of theoretical finance, was "performed" into reality. Ultimately the concept of risk used in theoretical finance was at best an abstract idealisation for the purpose of explaining a theorised mechanism. As it came to be used as a fundamental precept for the architecture of the financial system, its validity as a foundational assumption became less and less justifiable. A key argument of this thesis is that the concept of risk underpinning financial theory is inadequate, even dangerous, to use in designing institutions in the financial system.

5. Social ontology

Theoretical finance is both about the social world and causally relevant to it. An important ambition of this thesis is to clarify these two roles. In one role, theoretical finance is a positive science, developing hypotheses about what is true in the world. It is positive in the sense that it aims to generate knowledge about the way the world is, looking at it as a physicist might see the universe. However, in any social science the thing being studied can respond to the work of those who study it. This is clear in finance as theories are partly normative, in that they argue that people should act in certain ways to achieve their objectives. But even in its positive guise, theories can change behaviour if people come to accept them as true or adopt the concepts and reasoning from the theories to apply in their engagement with the world. Theories therefore can change society simply because individuals engage with them. But theoretical finance is also tightly bound to the project of "making" the social world, of shaping the institutions of the financial system. In this role it is in the service of design and engineering. This implies two "stances" for finance theorists: as positivist scientists and as social engineers and each of these may require a different methodology. I make a normative claim of my own: that these stances *should* employ different "types" of models, arguing that the core models of theoretical finance are socioeconomic machines and not blueprints. While there are superficial similarities, I argue that these two types of model are fundamentally different. I argue that socioeconomic machines are epistemically useful in the theory stance, but dangerous in the design stance where it is important that blueprints generate robust institutions. This difference comes down to the fact that the epistemic virtues of models in the two stances are different. In the theory stance we want models to explain much with little. In the design stance we want blueprints to contain detail for institutions that can manage all the risks we might face in the real world.

This argument depends on a proper theory of social ontology, and a significant part of this thesis is concerned with developing such a theory. In Chapter 3 I argue against constructivist notions of social reality, particularly that of John Searle, supporting instead an equilibrium account by which institutions emerge from the right set of causal conditions. A market, for instance, exists when there are buyers and sellers, goods or services, and a medium of exchange, all arranged in the right way, and continues to exist while this arrangement holds. It emerges when such an organisational structure comes about and does not exist before then, and will cease to exist when that structure is changed in critical ways. Such arrangements are at least to some extent outside of the intentions of any human agent. Social facts and institutions can emerge without being intended by anyone. They are always in process, and one can never completely anticipate all the features of an institution from one moment to the next.

I present this argument by considering arguments for constructivist or performative views of social ontology, ones that describe social reality as a result of collective intentionality or the work of economists. Such arguments present problems for the possibility of social science by characterising it as constructivist rather than positive. According to these accounts, social scientists are in the business of shaping reality and not studying it, and even their attempts to study it inescapably affect the world. It is by interrogating these arguments that I clarify that social science has the two stances described above and separate out the methods involved in each.

6. A summary of the chapters

Chapter 2 is largely descriptive in characterising the theoretical finance research programme and the MVO and CAPM models as exemplars of it. I consider the development of theoretical finance from the standpoint of the 20th century arguments in philosophy of science, metaphysics, and their interpretation in debates over economic methodology. I include a discussion of the philosophy of risk and show theoretical finance evolved from a subjective notion of probability to an operationalist notion according to which risk was identified with the volatility of market returns. I argue that a standard Lakatosian analysis of theoretical finance as a research programme fails to identify a unique theoretical hard core. I argue that we can instead look at theoretical finance by examining its ontological commitments, the internal metaphysics that the world must conform to if such models are to be seen as true. Among those, I identify the assumption of a world that generates stable probability distributions, such that risk can be equated with returns volatility, as the core ontological commitment of the programme.

I conclude the chapter by considering the role of empiricism in the research programme. I argue that key predictions in theoretical finance have faced empirical disconfirmation, to which the response has been various strategies to protect the core. I characterise two strategies to defend the core: first, a Millian strategy that dismisses disconfirmation as unpersuasive about the reality of the theorised tendency or mechanism and, second, an argument that disconfirmation amounts to a measurement error. I argue that these strategies can be successful only in certain respects and point to multifactor models as a threat to the ontological commitment at the core of the programme.

In Chapter 3 I begin a consideration of social ontology with a detailed examination of John Searle's theory, focusing on the example of money. I argue that Searle's constructivist account is fundamentally mistaken for two main reasons. First, it relies on a premise that physical objects are essential to social reality. Using the example of money, I argue that money is a concept of value, and not any physical thing. Physical things like bank notes are only representations of money, just like the records on a computer at a bank. I argue that money is prior to these things, and serves as a means of quantifying value and expressing deontic obligations to each other. I argue that the financial system is an elaborate record-keeping device of these obligations. Second, Searle's account cannot provide an explanation for the collapse of institutions. This is a key weakness, particularly in light of the financial crisis. I provide an account of money that argues money is a concept of value that emerges in human societies as individuals synthesise their subjective feeling of how much the value something with their experience of others' willingness to exchange. Money provides a means of quantifying this sense of value. It also works as a store of this value and the financial system works as a memory device for these allocations of value. There may be various such concepts used in any society, and these may become more or less prominent as individuals update their beliefs about the value of money itself.

In Chapter 4 I connect the arguments on ontology to the arguments on methodology by analysing the concept of performativity. This is the idea that economists are causally authoritative over the economy and that economics is a programme for creating the economy rather than studying it. I analyse this claim into four possible interpretations, strong and weak, and design and theory. I argue strong performativity is false, weak performativity is trivial, design performativity is largely banal, but that theory performativity is important if properly construed.

In Chapter 5 I return to the methodology of theoretical finance consider how the exemplar theories work using models. I argue that MVO and CAPM are best seen as idealised models that aim to draw out real features of social reality. They also work as normative decision tools in that they are able to guide investment decision makers by providing a calculative device to determine optimal strategies. I argue that these models are also used as “blueprints” in the task of social engineering. Designs for social institutions guide the assembly of the right causal conditions for those institutions to emerge. However I argue that the means by which we assess the epistemic value of theory models is fundamentally different to the way we should assess the value of blueprints. I argue this is under-appreciated, so we mistakenly think theory models which are successful also make for good blueprints of financial institutions. I argue that the exemplar theories of theoretical finance meet certain standards for good theory models, such as being simple and therefore easy to use, but fail to meet the standards desirable for blueprints. For this reason, the use of theory models in the design of the financial system contributed to its weakness.

I conclude by arguing that a better appreciation of the theory and design stances in theoretical finance would be an advance of our understanding of society and interventions to make it serve the needs of everyone. I argue that the confusion of these contributed to the design flaws in the financial system. I argue that the financial crisis is not a disconfirmation of the exemplar models of theoretical finance, but it should be an indictment of their use as blueprints. Some institutions are already taking steps toward de-emphasising theoretical finance and its conception of risk in favour of, for example, heuristics-based approaches that appear to be better at managing uncertainty. I call for better testing of institutional designs, including scale models, randomised controlled trials and other experiments in order to make the institutions of the financial system more robust.

Chapter 2

Theoretical Finance as a Research Programme

1. Introduction

The research programme that has become theoretical finance had its genesis in 1952 with Harry Markowitz's "Portfolio Selection" paper. In the following decades, important theories and models were developed that form a canon with a particular set of foundations that make for an identifiable programme. The canon also includes the Capital Asset Pricing Model, the Black-Scholes model, the Efficient Markets Hypothesis, the capital irrelevance theorem and Value-at-Risk modelling, among others. The challenge I take up in this chapter is to provide an account of this programme that draws out its foundational features.

I consider these features in the categories of methodological, epistemological and ontological commitments of theoretical finance. I characterise theoretical finance as a research programme by examining two exemplar models, Mean Variance Optimisation and the Capital Asset Pricing Model. Through this examination I argue that the methodology and ontology of theoretical finance are all bound up with the way *risk* is conceived of in finance. Finance works by measuring variance in returns to financial assets and using this to form expectations about future variance. Variance is accepted as not just a measurement involving some degree of approximation to "real" risk, but as a complete definition of risk. This excludes the possibility of uncertainty, in the Knightian sense[1], from modern portfolio theory.

The account I provide straddles the methodological, epistemological and ontological and a first step is to be clear about how these different spheres of study intersect. Daniel Dennett connects method to epistemology as follows (Dennett 1999, his italics):

> The methods of science, like everything else under the sun, are themselves objects of scientific scrutiny, as *method* becomes *methodology*, the analysis of methods. Methodology in turn falls under the gaze of *epistemology*, the investigation of investigation itself – nothing is off limits to scientific questioning.

A methodology serves to govern the methods used by scientists, to dictate what methods can be used in the generation of knowledge by the science. The link between methods and methodology can sometimes be vague, and methodologies themselves often have to admit exceptions when applied to actual practice. Finance, however, has developed a very clear set of methods, which makes it amenable to methodological analysis. I will argue the set of models used in finance reveal a methodology that sees a particular kind of modelling and statistical analysis as a valid route to truths about the world.

A methodology is itself amenable to epistemological appraisal. Such appraisal operates at the meta-methodological level as methodologies are assessed for whether they provide reliable truths about the world. We can dismiss certain methodologies

[1] The economist Frank Knight distinguished between risk and uncertainty (Knight 1921). The former he defined as a measurable quantity that can be determined, for instance the failure rate of a factory production line. The latter could not be determined, consisting of completely unexpected events.

as "pseudo-scientific" when it is clear that the practices within the methodology are not capable of establishing the truth. Astrology, for example, has a methodology that we cannot rely on to produce true facts about the world. Such assessments occur at the epistemological level as we try to determine whether a methodology is truth-seeking.

There is an obvious sense in which science is about ontology – it aims to discover or explain things that it claims to exist, at least in common realist interpretations of scientific practice. But any methodology depends on a set of ontological commitments, preconditions that the world must meet in order for any methodology to engage with it. So a science is said to have "an ontology", which is a foundational system of categories, properties and entities that it holds as existing in the world, what might be called an "internal metaphysics" (Pratten 2007), that precedes any ontological finding the science may aim to elaborate. An internal metaphysics need not correspond with an external metaphysics – whether and in what way it does correspond is often a crucial research question for the research programme. Success in predicting certain phenomena concerning the external metaphysics may be taken as confirmation of the internal metaphysics. The ontological commitments are often implicit, not spelled out by practitioners, but implied by the laws and theories that are proposed to explain or predict phenomena. Drawing out the ontological commitments of a science is therefore a process of reverse engineering from the methodology to discover the ontology that theorists are committed to. Methods provide the basis to assess methodology, methodology provides the basis to assess epistemology, and both methodology and epistemology provide the basis to assess ontology. These domains of enquiry concern the foundations of a science that we might collectively call a "paradigm", using Thomas Kuhn's term (Kuhn 1962).

Theoretical finance is partly a practical programme that supports investor decision-making. It has been called a branch of "applied economics" because of this practical orientation (Eatwell, Milgate, and Newman 1992). Many of its theories and models work as decision aids, providing the tools to calculate and determine optimal investment strategies. This practical normative orientation gives it a strong operationalist character. Rather than becoming bogged down with vague notions of uncertainty, finance theorists embrace the practical use of statistical analysis to determine risk. However, finance still consists of theories that make positive claims about the world through models that are supposed to have some coherence with the world. The empirical investigation of finance theories, as I will discuss below, is just one way that this theoretical aspect of finance is clear within the programme. This has clear implications for the internal metaphysics of theoretical finance.

Many financial theorists will object to my characterisation of its concept of risk, saying they obviously recognise that there is more to risk than the volatility of returns in financial markets. Certainly, financial theorists have become much more aware of weaknesses in reducing risk to simple statistics derived from market data. A recent professional monograph on risk management (Coleman 2011) contains an extensive discussion on the distinction between statistics on market data and subjective probability as well as the presence of Knightian uncertainty in financial markets. However, these concerns are used to counsel practitioners against blind faith in modern portfolio theory, rather than to question the validity of modern portfolio theory itself. Coleman writes, for instance, that "Markowitz's framework provides immense insight into the investment process and portfolio allocation process, but it is an idealised model" (Coleman 2011: 15). I will argue that the theoretical finance research programme that has evolved since Markowitz's seminal 1952 paper remains a coherent one with a methodology and conceptual framework. Being aware of its limitations is much like a physicist being aware that the Newtonian model breaks down at very large scales and very small scales, but is useful at least in some applications. It remains valuable to look specifically at the Newtownian research programme and assess its foundations, just as it is to look at those of modern portfolio theory. In claiming that modern portfolio theory is committed to a particular concept of risk, it does not follow that all financial theorists are similarly committed.

Indeed, by making the commitments of the theory explicit, financial theorists may be much better able to understand what their views on risk really are. It also ensures that when using the body of theory, one can be aware of what commitments are implied. In Chapter 4 I will argue that modern portfolio theory has various important effects on financial institutions, some of which stem directly from its internal metaphysics.

The statistical analysis of financial market returns is used to determine properties of that data including its mean and standard deviation. Methodologically, these data form the objects of study for theoretical finance. The epistemology involves claims to knowledge through theories and models that aim to show the relationship between returns and risk using this data. A core notion in theoretical finance is that returns reflect the compensation for taking risk; the greater the standard deviation of returns, the higher must be their mean. This framework implies an ontology that has markets existing as stable systems generating randomness such that asset returns in each period conform to a knowable probability distribution.

I begin with expositions of Mean Variance Optimisation and the Capital Asset Pricing Model. These are referred to often in this thesis and I spell out the models in detail in this chapter. They are, in a Lakatosian sense, "exemplars" of the programme and are critical building blocks in the full body of modern portfolio theory that includes the efficient markets hypothesis, Black-Scholes model, and Value at Risk modelling.

I then examine these exemplars to draw out the philosophical implications of the way probability is treated in the models. I argue that while Markowitz initially positioned himself as a subjectivist about probability, he outlined a method that is fundamentally frequentist. I argue that Sharpe takes an operationalist approach to probability in which the concept is defined by its measurement procedure, namely to establish volatility from the data.

I then consider finance through the prism of Lakatosian analysis (Lakatos 1968, 1970, 1978; Lakatos, Worrall, and Currie 1980). This is because Lakatos understood science as a series of research programmes and my approach to theoretical finance is to see it as a programme. Lakatosian analysis was once popular as a way of understanding methodology in economics, including in finance. On examination, however, I show that Lakatosian analysis provides a poor account of methodology in finance. Instead, the defining features of the programme are its reliance on frequentist analysis and in the ontological commitments that implies. These commitments are to a world that has the propensity to generate stable probability distributions of financial market returns.

I then examine the operationalism inherent in finance's methodology. Operationalism reduces all concepts to their measurement and does not permit concepts that cannot be measured. In theoretical finance the casualty of this approach is the denial of any form of Knightian uncertainty, which is incapable of an operationalist definition. So the internal metaphysics of theoretical finance is a world in which risk is precisely that determined by the measurement of variance and there is nothing else to it. This implies a form of ontological commitment to the world being such that risk in financial markets stems from frequencies of returns in the long run, and financial markets have the propensity to generate such stable probability distributions.

The last section of this chapter considers the role of empiricism in the development of the programme, in the context of Popperian ideas of falsification (Popper 1959) and how they might apply to theoretical finance. I identify rival programmes that can be delineated according to their relationship to the ontological commitment at the core of theoretical finance. I consider what I call Millian defences to putative disconfirmation of finance theories. I argue that programmes which threaten theoretical finance are those with fundamentally different ontological commitments and concepts in respect of risk.

2. Markowitz's Mean Variance Optimisation (MVO)

Many writers identify Harry Markowtiz's 1952 paper, "Portfolio Selection", as an important historical milestone in the development of finance as a programme. Nobel laureate Merton Miller described it as finance's "big bang" (Miller 1999: 8). Financial historian Peter Bernstein calls it "a landmark in the history of ideas" (Bernstein 1993: 41). Berkeley financial economist Mark Rubinstein describes it as "the moment of the birth of modern financial economics" (Rubinstein 2002: 1041). Finance historian Haim Levy uses Kuhnian terms, describing it as a "breakthrough", adding that "a glance at finance textbooks that were used in teaching before 1952 and textbooks that are currently used suffices to reveal the revolution induced in the finance profession by the publication of this 1952 Mean Variance article" (Levy 2011: xi).

The challenge is to identify just what it is that made Markowitz's paper so different. I will argue in this section that at first the differences were merely superficial – Markowitz relied on statistical analysis, but the framework of mean variance optimisation was methodologically similar to other models in economics before it. Rather, its uniqueness only became clear as subsequent development of the programme led to a specific epistemological and ontological framework.

MVO is a simplified model of financial markets and is primarily concerned with the question of how to combine assets into a portfolio such that the portfolio is "optimal". The answer it gives is that optimal portfolios are those that maximise return for a given amount of risk, or minimise risk for a given amount of return. The model shows that this happens when the right assets are combined in the right proportions.

According to the model, each asset has three properties:

1. The expected return of that asset, denoted $E(R)$

2. The expected variance of returns for that asset, denoted σ^2
3. The expected correlation of those returns with returns of other assets in the portfolio, denoted ρ

All of these are "expected", ex ante rather than ex post, so Markowitz's approach was forward-looking, at least in his formal description. The 1952 paper itself did not address how these estimates should be developed, but Markowitz addressed himself to those operational questions in his 1959 book that developed the ideas, which I will come to.

These three measures are statistical properties of data. The returns are expressed as a percentage on the value of the asset in the previous period, for example, in the form of daily figures for the percentage change of a share price. Variance is a statistical property of a set of numbers that indicates how much the numbers vary from their mean, calculated as the square of their standard deviation. In this context, the relevant numbers are return figures for a specified time interval, such as daily, expressed as percentage losses or gains. So a particular share may appreciate by 1.0% on a day and then depreciate by 1.5% the next day, and so on, accumulating a data set from which mean and variance can be determined. If all the figures are identical, the variance will be zero, while a higher variance indicates that the returns are widely dispersed from their mean. This is often depicted using a probability distribution, often assuming returns are normally distributed, where the random variable is the percentage daily return. The standard deviation of asset price returns is also often called the volatility of returns, so volatility is the square root of variance. Volatility and variance are often used as synonyms to refer to the degree to which an asset's price varies from the mean, so in common financial parlance a "highly volatile stock" is one whose returns tend to range far wider than the market.

Correlation indicates how closely the returns of two assets match. If one asset's return is negative when the other is positive, the correlation will be negative, while it will be positive if the two assets have returns that move in the same direction, reaching 1 if the two match perfectly. It follows that correlation captures the impact of diversification. A two asset portfolio will have a lower overall variance if the correlation of assets within the portfolio is less than 1.

Intuitively, a portfolio containing only railway securities (to use Markowitz's example) would have a higher variance than one containing railway, utility, mining, and manufacturing securities, because the correlations between these assets would be less than 1. The correlation between securities in the railway portfolio is likely to be close to one as the securities are likely to all be affected in the same way by changing economic conditions. In the diverse portfolio, however, the securities are likely to be affected differently by changing conditions, so the correlation of the price movements will not be as high. It follows that by seeking to combine assets that minimise variance, one would usually combine assets that don't have a high positive correlation, obtaining diversified portfolios. This is the mean variance optimisation strategy.

The expected return of the portfolio is calculated as (using notation similar to Markowitz 1952 though with some adjustments to bring it in line with modern expressions of the approach):

$$E(R_P) = \sum_i w_i E(R_i)$$

Where the expected return of the portfolio, $E(R_p)$, is the sum of the weighted expected returns of each asset, i, in a portfolio. So, for a two asset portfolio consisting of assets A and B the expected return is:

$$E(R_P) = w_A E(R_A) + \ w_B E(R_B)$$

Next, the expected variance of the portfolio is calculated:

$$\sigma_P^2 = \ \sum_i w_i^2 \sigma_i^2 + \ \sum_i \sum_{j \neq i} w_i w_j \sigma_i \sigma_j \rho_{ij}$$

Where the expected variance of the portfolio (σ_P^2) is the sum of the squared weighted variance ($w_i^2 \sigma_i^2$) for each asset *i* plus the sum of weighted standard deviation of each asset multiplied by the correlation between them (ρ_{ij}). This second term is alternatively described as the "covariance matrix" of the assets in the portfolio as it sums up the covariance of each asset with each other asset. If the correlation between assets is zero, implying that the returns on assets are completely independent from each other, then the second term of the equation is unnecessary. However, it is usually assumed that the prices of financial assets have some level of correlation because they are all ultimately a function of economic activity (Brentani 2004).

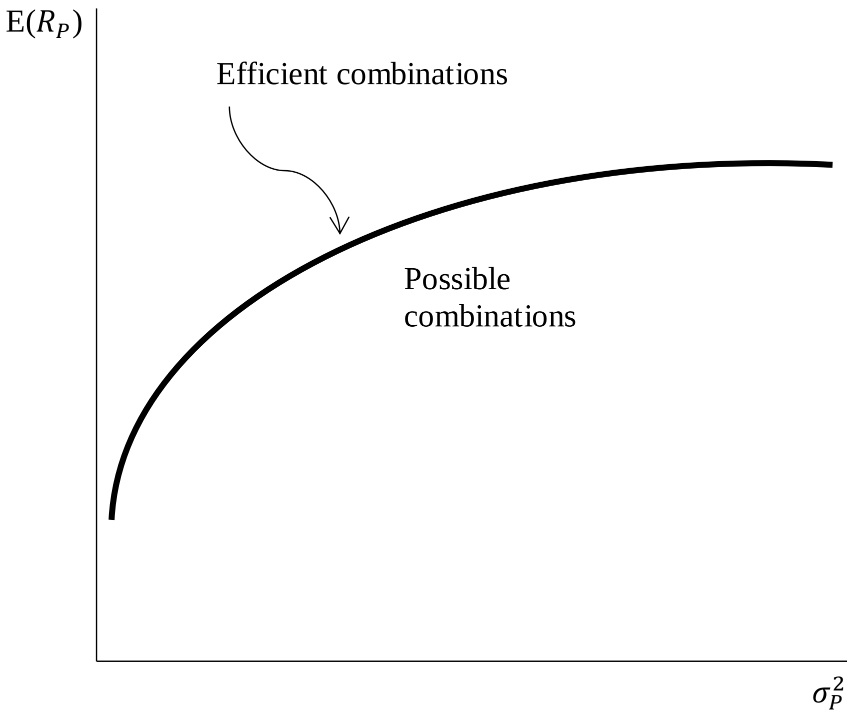
For example, for a portfolio of two assets A and B, the variance is determined as:

$$\sigma_P^2 = w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2 w_A w_B \sigma_A \sigma_B \rho_{AB}$$

Each additional asset substantially increases the computational load, requiring the calculation of the correlation between that asset and each other asset in the portfolio to complete the matrix. This made the practical application of MVO cumbersome, particularly before the advent of computers. This was an important practical limitation at the time which, as we will see, drove the further development of theoretical finance.

Once a set of return and variance expectations is determined for each asset, MVO provides a technique to determine an optimal portfolio, which is one where expected returns cannot be increased for a level of expected variance, or expected variance cannot be decreased for a given level of expected returns. In modern treatments (e.g. Ruppert 2004) this is expressed graphically as follows, although Markowitz's original 1952 expression had variance and return on opposite axes:

Figure 1: Efficient portfolios



Only portfolios that fell on the line could be rational, where returns and risk are at efficient levels. A simple decision rule was thus established. Investors should seek to combine risk and return so that the former is minimised and the latter maximised. To accept a lower than possible return for a given amount of risk, or to accept a higher than necessary level of risk for a given amount of return, was irrational.

3. Sharpe's Capital Asset Pricing Model

The main CAPM paper was published in 1964, 12 years after Markowitz's key paper. The CAPM is now pervasive in finance teaching in business schools and in the techniques of professional asset managers. As Fama & French put it, "It is the centrepiece of MBA investment courses. Indeed, it is often the only asset pricing model taught in these courses" (Fama and French 2004: 25). This recalls Kuhn's words quoted in my introduction, of normal science being "a strenuous and devoted attempt to force nature into the conceptual boxes supplied by professional education" (Kuhn 1962: 5). The CAPM also dominates in research papers published in finance, playing a role in up to 50% of the articles published in the leading journals of the field (Levy 2011).

Sharpe had been a protégé of Markowitz at the Rand Corporation and acknowledges an intellectual debt to Markowitz in his early papers. The two had focussed on reducing the computational complexity of identifying efficient portfolios, bearing in mind that the MVO exhibited a combinatorial explosion as additional assets were added to the portfolio. The main consequence of Sharpe's 1964 paper was to dramatically reduce the computational complexity of determining "efficient" portfolios. The publication was closely followed by Lintner (1965) who is often described as a co-developer with Sharpe of the CAPM although they had worked separately.

Sharpe (1964) introduces a number of important concepts that have become entrenched in theoretical finance, particularly the notions of systematic and unsystematic risk. Drawing on Markowitz's MVO and earlier work by James Tobin, he argued that the market as a whole represents an efficient portfolio. It is the maximally diversified portfolio. This will exhibit volatility that Sharpe terms the "systematic risk", which will reflect all stocks' "common dependence on the over-all level of economic activity" (Sharpe 1964: 441). This risk is the variance of the market. However, each individual asset in the market will have its own volatility that differs from the market as a whole. In order for that asset to find itself a place in the market portfolio, its return has to compensate for that individual risk. This is an equilibrium condition for the price of the asset – it would not be bought by investors if its price was too high relative to its risk, and would be bought if it was too low, so normal demand and supply forces would ensure the prices of securities in equilibrium reflect the relative contribution of risk and return to the market portfolio. This aspect of Sharpe's approach is clearly consistent with Marshallian equilibrium analysis.

It follows that if the overall market return and variance can be estimated, and the covariance of the individual asset's return with the market, then a "required return" can be backed out. This is a much simpler set of estimates than that required for the MVO, in effect treating the investment decision as one about two assets: the market and the individual asset. It also provides a way to determine the "expected return" to go alongside the expected risk required for MVO.

The calculation of required return is expressed as:

$$E(R_i) = R_f + \beta_i(E(R_m) - R_f)$$

Where $E(R_i)$ is the expected return of asset i, $E(R_m)$ is the expected return of the market as a whole, R_f is the risk free interest rate, and $(E(R_m) - R_f)$ is the "equity risk premium". The expected return of any one asset should be the risk free rate, such as that paid by a government bond, plus some premium representing an additional return for having taken the risk. The intuition is that you wouldn't choose a risky investment unless it paid you more than a government bond, which is usually seen as risk free. But the calculation of the riskiness of the asset is revealing.

The equity risk premium is the return one expects for investing in the market as whole. Sharpe provides little discussion of its calculation, saying only that that most discussions of market values "assert" that it exists. There is now a considerable literature on the equity risk premium but finance textbooks usually advise that this can be determined by examining historic data, for example by taking the average outperformance of an equity index over the risk-free rate (for instance, Koller, Goedhart, and Wessels 2005).

Any one individual asset contributes to the risk of the portfolio either positively, so increasing the average risk, or negatively, so decreasing the average risk. This makes sense as soon as you see risk as variance; the same approach as Markowtiz. The idea is that an asset should pay a higher return if its variance increases the average variance of the market, but a lower return if it decreases it.

To assess this contribution of variance to the market average, the CAPM relies on the all-important beta of finance, represented by β_i. This is calculated as:

$$\beta_i = \frac{Cov(R_{i,}R_m)}{Var(R_m)}$$

In other words, the risk premium an investor should earn for holding an asset *i* is a function of the covariance of that asset's returns with the returns of the market, per unit of variance in the overall market. An asset that is just as volatile as the market as a whole will have a beta of 1, while assets with a volatility uncorrelated with the market will have a beta less of than one, while an asset with a volatility that exaggerates market price movements will have a beta greater than 1. For example, a utility company with predictable income and little risk will have a beta lower than 1 as it is less volatile than the market, but still tends to move somewhat with the market as a whole. An asset like gold has a price which has little correlation with the stock market even though it is still highly volatile, so usually would have a beta less than 1 too. A cyclical stock like a luxury goods company geared toward the overall economic cycle will tend to show exaggerated price movements with the market as a whole in the same direction, giving it a beta greater than 1.

A beta of 1 implies that the return for holding *i* should be the same as the overall return for the market, a beta of less than 1 implies it should be less, and a beta of more than 1 implies the returns should be higher. The beta is used to weight the equity risk premium so the investor's expected return is a function of that asset's contribution to the overall variance of the market. Beta has become one of the most pervasive notions in all of theoretical finance. There are clearly intuitively appealing aspects of the CAPM in that it appears to provide a justification for the common sense view that riskier assets should provide the prospect of higher returns.

Alternatively, beta can be calculated by regression analysis of the returns of the individual asset compared to the returns of the market as a whole. Beta is the slope of the regression line of the asset's returns as the dependent variable and the market returns as the independent variable. This provides the same outcome as the formula for beta discussed above. I provide an example below.

Sharpe's paper did not provide any analysis of the concept of probability. Sharpe assumes that any individual "views the outcome of any investment in probabilistic terms; that is, he thinks of the possible results in terms of some probability distribution" (Sharpe 1964: 427). Sharpe, like Markowitz, assumes the individual acts on the basis of only two parameters – the expected return and the standard deviation, which form his utility function. As in the case of MVO, the CAPM entails an ontological commitment to the existence of expected returns and variance.

Sharpe's 1964 paper does not shy away from how risk should be assessed, as Markowitz's 1952 paper did. The simple answer is that risk is equivalent to variance. Sharpe does not dwell on a theoretical justification of this approach, choosing instead to define it in terms of how it can be measured. He sets it out in this passage (Sharpe 1964: 438-39 parentheses and italics as per original):

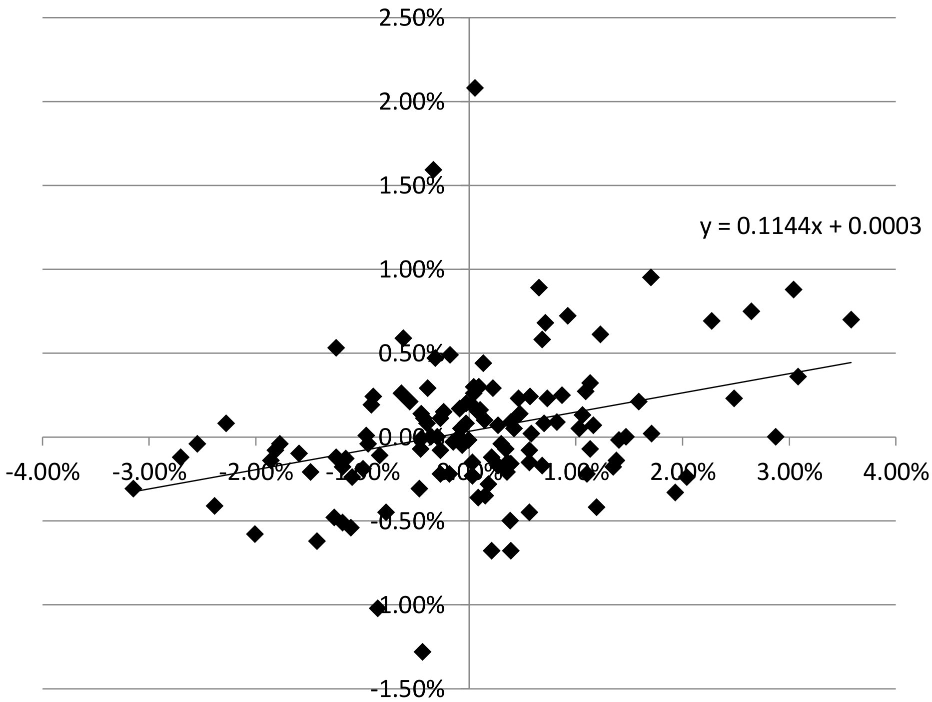
> Imagine [in the case of asset i and combination g] that we were given a number of (ex post) observations of the return of the two investments…The scatter of the R_i observations around their mean (which will approximate E_{Ri}) is, of course, evidence of the total risk of the asset $- \sigma_{Ri}$. But part of the scatter is due to an underlying relationship with the return on combination g, shown by B_{ig}, the slope of the regression line. The response of R_i, to changes in R_g, (and variations in R_g, itself) account for much of the variation in R_i. It is this component of the asset's total risk which we term the *systematic* risk. The remainder being uncorrelated with R_g, is the *unsystematic* component. This formulation of the relationship between R_g and R_i, can be employed ex ante as a predictive model. B_{ig} [slope of the regression] becomes the predicted response of R_i to changes in

> R_g. Then, given σ_{Rg} (the predicted risk of R_g), the systematic portion of the predicted risk of each asset can be determined.

Put simply, Sharpe's method is to take a dataset of historic returns for two assets (i is the individual asset and g is the market portfolio), and determine the volatility (standard deviation) of each set of returns data. This would represent the "total risk of the asset" which makes it clear that in Sharpe's eyes, there is nothing to risk apart from the volatility of returns, a hallmark of operationalism, as I will discuss in the next section. The data allows for further analysis to determine the relationship between the returns of the asset (i) and the market (g). By plotting these on a graph and using regression analysis one can plot a regression line. The slope of that line, what Sharpe describes as B_{ig}, is the beta. This measure, according to Sharpe, now becomes an "ex ante" predictive tool. So, put simply, Sharpe's point is that by determining the historic correlation of asset returns, one can obtain a means to determine expected volatility, which is the risk. This is a nakedly frequentist approach to the statistics, allowing one to make claims about the future based on the historic data. Markowitz's nod to subjective probability has been shed.

To help make Sharpe's approach clear, let me work through a small example. The graph below represents six months' of daily returns until the end of July 2016 for the FTSE 100 and for one share that is contained in the index, SABMiller, making a total of 127 observations (investing.com 2016). The FTSE 100 is the independent variable on the X axis and SABMiller is treated as the dependent variable on the Y axis. This is the resulting graph:

Graph 1: Regression of the returns to the SABMiller share against returns to the FTSE100 (investing.com 2016)



The positive slope of the trend line indicates that there is a positive correlation between the two variables. When the market has a positive day, the SABMiller share tends to have a positive day. The coefficient of the x value is beta, which in this case is 0.1144. This implies that the variance of SABMiller's share price is quite weakly correlated with the variance of the market as a whole (using the FTSE100 as a proxy for the market). SABMiller therefore would be considered a relatively low risk share and the expected return to be earned from investing in it would be lower than the average for the market as a whole. Perhaps this is because SABMiller's main product, beer, is consumed no matter what the economic conditions, so the firm's earnings are less volatile.

Sharpe separates risk into "systematic" and "unsystematic" components. The systematic risk is that accounted for by the market as a whole and represented by beta. In addition to that risk, there is residual risk that is contained within an asset's return volatility that is not driven by the market. For example, the takeover bid for

SABMiller by Anheuser Busch during the period considered for this example would have led to quite different pricing dynamics than the rest of the market and be contained in its unsystematic risk.

Sharpe implies that there is a causal mechanism at work. R_i is caused by R_g and the volatility of the asset's price, at least that part that is "systematic", is caused by the volatility of the market. In other words, a move in prices of the global portfolio will cause a predictable movement in the prices of an individual asset with the probability measured by beta. The claim Sharpe is making is a mechanistic one – that individual asset variance is determined, to some extent, by the market portfolio. The slope of the regression line is β (denoted B_{ig} in the extract above) and provides the connection between the two: a measure of the relationship between volatility of the individual asset and volatility of the market as a whole. This systematic component is now a predictive tool – we can use it to predict the future variance of the asset and therefore determine the expected return via the CAPM.

Beneath the hood, as Daniel Hausman (1994) might put it, the CAPM contains a theoretical mechanism which allows for ex ante predictions of variance, that in turn allows for forecasts of expected returns. This suggestion of a mechanism can be contrasted with a strict regularity approach to causation stemming from David Hume, in which nothing more can be said about two events than that one follows the other. An explanation of a mechanism might provide a way to explain *why* we obtain some regularity, rather than just a claim that a law-like regularity exists. A mechanistic explanation provides a story about the structures and processes that underlie the relationship between two variables (Reiss 2013). The mechanism within the CAPM can be described as an individual causal relation, which does not mean anything beyond causality of one variable to another, in this case of market variance to the variance of the individual asset. The model does not shed any light on the processes that might underpin this. Seeing as the market variance is nothing other than the combined variance of the prices of all the assets in the market, it might be more compelling to argue that the causal relationship goes the other way around: from individual price variance to the variance of the market. Sharpe's view is that market variance is caused by broad macroeconomic factors which affect individual securities to different degrees, so the mechanism is from the macroeconomy to individual company performance, and market variance serves as a measure of the macroeconomy. I consider this an open question, though there is some empirical support for Sharpe's implicit view, in that futures on broad market indices like the S&P 500 do usually lead the prices of underlying stocks, even if only by a few milliseconds (MacKenzie 2017). But what the extract quoted above makes clear is that the basis for determining expected variance is the ex post observations of historic variance. Historic data is like an expert we can consult to determine probability beliefs.

4. Markowitz's methodology

Having sketched out the two exemplars of the programme, I now want to examine the methodological positions they represent. There are interesting differences between Markowitz and Sharpe, representing an evolution of methodology between the two. The latter came after 12 years of development of the programme and by then certain theoretical commitments had solidified. However, the course was set toward the later methodology in Markowitz (1952) even though he at times appeared to explicitly contradict what later become accepted.

Superficial features of his 1952 paper have been noted by various authors. In its edition of the Journal of Finance, it was the only article that included equations and graphs. Bernstein (2005) examined the next seven years of the journal and could count only five that were theoretical rather than descriptive. By descriptive, Bernstein meant simple explanatory accounts of relevant issues, such as Federal Reserve policy or data on money supply, prices and business output. Markowitz's paper stood out for its focus on decision making and use of mathematical notation.

It provided an argument for an intuitively plausible idea: that a portfolio of assets should be structured so as to maximise returns *and* to minimise risk. In this sense it was making a positive claim about a mechanism that exists in the world, and therefore can be understood as positive. But it was also presenting a normative case for how investors should make decisions.

The popularity of Markowitz's approach was at least partly influenced by factors that had nothing to do with the positive or normative claims the theory was advancing, but rather with the more prosaic consequences for research productivity for those involved at the time. In the 1950s and 1960s early computers were being developed and large databases of financial market statistics were being assembled for the first time. This had a galvanising effect on the popularity of Markowitz's approach – it provided a theory to support practical research into the data that were being accumulated. The databases of stock prices and returns lent themselves to statistical analysis of the sort that could derive ex post variance and correlations between asset returns. There was a mutually reinforcing link between the development of the theory and the development of technology that allowed for the analysis of historic data.

The normative implications of Markowitz's theory were initially awkwardly received. Markowitz reported that Milton Friedman, when serving on his PhD committee, had said his thesis was interesting but did not qualify as a work of economics (Markowitz 1991), though the committee still granted the PhD in economics. Friedman later denied saying this but did contend that the thesis had been more of a mathematical exercise than a work of economics (MacKenzie 2008). The story is meant to suggest that there was something very different in Markowitz's work. It was not the mathematics or statistics itself that marked the difference. Mary Morgan (1992) notes that Friedman supported the use of statistical economic data to derive theories and test hypotheses which contributed to the development of econometrics at the time. Rather, Friedman seemed concerned that the work was normative in character, a work of decision theory rather than a positive claim about a tendency or mechanism in the economy yielding predictions that could be tested. As such, MVO had more in common with statistics than with economics.

Markowitz clearly saw his 1952 paper as presenting both positive and normative claims. “Various reasons recommend the use of the expected return-variance of return rule, both as a hypothesis to explain well-established investment behaviour and as a maxim to guide one’s own action” (Markowitz 1952: 87). He therefore believed he was providing a theory to explain the existing behaviour of investors as well as a normative guide to how investors should make decisions. This positive/normative dual function of models in theoretical finance is common to several others in the programme, including CAPM.

4.1 MVO as positive theory

The equations of MVO serve as an abstract representation of purportedly real properties of markets, in a highly idealised form. MVO makes a positive claim about the way the world really is: that securities prices reflect investors rationally pursuing an efficient investing strategy, by balancing their expectations about return with their expectations about risk. The MVO framework serves as an idealised model of the behaviour of rational investors. The simplified notion of “expected return” is all that investors are attempting to maximise while simultaneously minimising a simplified notion of risk, which is represented as “expected variance”. This is an idealised world similar to the expected utility maximisation assumption that is common in much of consumer choice economics, and Markowitz saw his approach as similar to expected utility maximisation.

In their discussion of models in science, Frigg and Hartmann divide the idealisation found in models in science into two main kinds, Aristotelian and Galilean (Frigg and Hartmann 2012). Aristotelian idealisation strips out all factors that are not relevant to the particular subject of interest. For example, most models of the solar system ignore the gravitational influence of any celestial objects outside of it. Galilean idealisations involve distortions so as to simplify reality. Galileo used point masses and frictionless planes and economists use fictions such as infinitely divisible goods in their models.

There are both Aristotelian and Galilean idealisations in the MVO framework. It is Aristotelian in that all other factors are stripped away to focus on the ones that are proposed to be important. So whether you happen to like or dislike an investment for other reasons than its expected return and variance, perhaps because you work at the company concerned, is outside of the model. The idealisation is also Galilean in that "risk" is reduced to variance. This reduction of risk to a simple statistical measure, strips the concept of its usual connotations, such as that it refers to the probability of danger or harm, and unexpected events generally. Now risk becomes something that can be expected, is symmetrical in harms and benefits (at least when the normal probability distribution is used, as is the most common practice), and is blind to any uncertainties. Returns are reduced to a financial percentage, which ignores other forms of utility that may come from ownership. The reduction of risk to variance helps to make the model practical and operational, which was particularly important for the CAPM, as I will discuss in section 6 below. As a positive model of investment behaviour, the Aristotelian and Galilean idealisations in MVO help to provide a tractable model in which the relationship between risk and variance can be made explicit using relatively few, simple, terms.

Assumptions in economic models serve to shield models and simplify particular aspects of models. The *ceteris paribus* clauses that shield models from outside influences serve to establish the Aristotelian idealisation. The "unrealism" of assumptions in economics, such as perfect information or zero transaction costs, help to idealise the world of the model in a Galilean sense. All of these are, in a sense, "unreal" assumptions. This lack of realism was famously defended by Friedman who argued it was the success of predictions that mattered, not the realism of assumptions (Friedman 1953). Indeed, he argued that models might be better if, though their assumptions, they are made easier to use and more accurate in predictions. So he would not have resisted Makowitz's methods on the grounds of the idealisations in his models.

4.2 MVO as normative

The positive interpretation of the model was seldom made in subsequent literature. Sharpe, who was Markowitz's protégé, later described Markowitz's work as entirely normative (Sharpe 1964). On this reading, MVO is really a tool to use in making decisions, based on analysing a portfolio to determine if it is "efficient" and then to make buying and selling decisions in order to move closer to the efficient frontier. The gap between the normative and positive interpretation of MVO depends on whether we can make the claim that investors really are aiming to maximise the efficiency of their portfolios in a rational way. There have been various attempts to deny that this is what investors are really doing, including much of behavioural economics which argues investors are not always rational (Kahneman and Tversky 1979; Simon 1959). But the usual arguments in response, that rationality is the only way to prevent exploitation by other investors, are difficult to respond to.

In Chapter 4 (section 6.1) I will discuss the nature of normative reasoning in economics at large in some detail. Many economic models have a dual positive and normative role, when normative is understood as how one can achieve some particular goal. This is a natural consequence of discovering something about the world in that it then becomes amenable to being changed. The MVO provides a decision rule about whether to buy or sell investments, based on the goal of maximising returns and minimising risk. The decision is guided by the objective of creating an "efficient" portfolio. Doing so takes some statistical analysis to understand the properties of assets and then selecting them accordingly. This normative character of the model is an important feature, but it is also true that the MVO model is a positive claim about the world and that has been clear since Markowitz's earliest work.

5. Markowitz's philosophy of probability

There is in fact a clear difference between the way Markowitz understood probability and the way Sharpe did. In this section I consider Markowitz and in the next section, Sharpe. As will become clear, Markowitz was initially hesitant and unsure about the nature of probability, which gave way in his later work and even more so in Sharpe's work, to an operationalist approach that treated risk as nothing other than long run limiting frequency of mean and volatility in data sets.

Markowitz's 1952 paper is decidedly vague about how expectations about return and variance should be formed by investors. He notes in a footnote that his paper "does not consider the difficult question of how investors do (or should) form their probability beliefs" (Markowitz 1952: 81). In another footnote he notes that if investors are consistent they will possess a system of probability beliefs, which would be "in part subjective". He speculates that "perhaps there are ways, by combining statistical techniques and the judgment of experts, to form reasonable probability beliefs" (Markowitz 1952: 82). Markowitz was conscious of the computational load of performing the necessary calculations for a portfolio containing a handful of assets, given the combinatorial explosion that ensues as you add assets and have to calculate the correlation between every possible combination of them. The computer was still some time away; three years after his paper he found himself at Yale with an early computer that could only handle the calculation for 25 securities. But he concludes his paper with a cautious comment about statistical analysis (Markowitz 1952: 91):

> My feeling is that the statistical computations should be used to arrive at a tentative set of [expected returns] and [variance]. Judgment should then be used in increasing or decreasing some of these [expected returns] and [variances] on the basis of factors or nuances not taken into account by the formal computations…One suggestion as to tentative [expected returns]

> and [variances] is to use the observed [expected returns and variances] for some period of the past.

While tentative, this represents a mix of subjective and objective interpretations of probability. Individual beliefs were “in part subjective” yet statistical computations, which could be carried out on the databases then being assembled, could be used to determine the relative frequency of returns in the long run. Once this is done “judgement should…be used” to fine tune the resulting expectations. One could interpret this approach as Bayesian: Markowitz can be taken to say that one should use historic data to form “priors”, and then use subjective judgment to condition them. This is reminiscent of some modern Bayesian approaches that integrate data analysis with background knowledge such as Bayesian network analysis (Heckerman 1998). But Markowtiz is modest on this point saying that the determination of probabilities is “another story” of which “I have read only the first page of the first chapter” (Markowitz 1952: 91).

This caution diminishes by the time of his 1959 book, where the practical use of statistics for data analysis is the focus. Much of the book provides financial analysts with guidance on how to measure returns and variance from market data. Markowitz writes that (Markowitz 1959: 14):

> Portfolio selection should be based on reasonable beliefs about the future rather than past performances *per se*. Choice based on past performances alone assumes, in effect, that average returns of the past are good estimates of the “likely” return in the future; and variability of return in the past is a good measure of the uncertainty of return in the future. Later we shall see how considerations other than past performances can be introduced into a portfolio analysis. For the present it is convenient to discuss an analysis based on past performances alone.

So here Markowitz is setting the scene for the use of historic data as the basis to determine expected risk and returns. The “later” that he refers to comes in the last quarter of the book where there are two chapters on foundational principles, including expected utility (explaining why investors trade off returns and risk) and probability beliefs. The probability chapter starts by acknowledging a debt to Frank Ramsey and Leonard Savage. Ramsey's 1926 paper outlined a method of measuring personal probabilities using gambles to determine at what probability individuals are indifferent to the gamble. This reveals the beliefs individuals have, providing a subjective probability conceived of as "degrees of belief", or credences, in a proposition. I might, for instance, believe that there is a 50% chance that my team will win a football game, and reveal this by accepting anything better than 2:1 odds that it will win.

The method has pragmatic appeal in finance because securities prices can serve a similar function to Ramsey's gambles in revealing the beliefs of investors. If the investor buys security A at price P, it follows she must believe that the returns for doing so compensate for the risk. If she was indifferent at P, we can assume her expectations are roughly the same as those implied by P, and if she sold at P it follows that her risk and return expectations are such that P is too high for them to be met. This approach is fundamentally behavioural, in the Skinnerian sense (Skinner 1953), requiring no need to speculate about what investors might actually be thinking. Key economic figures like Paul Samuelson and Friedman strongly endorsed such approaches at the time, and behaviourism of this sort continues to be an influential, if not dominant, approach in economics to this day[2] (Dietrich and List 2016).

---

[2] The sense of "behavioural" used here is quite different from what is referred to by writers about "behavioural" economics or finance (Kahneman and Tversky 1979). In behavioural economics it is precisely the speculations about psychological motivations of actors that drive the theories. This is in sharp contrast to the behavioural psychology propounded by BF Skinner (1953) which eschews speculation about mental states in favour of restricting the scientific domain to what people actually do, which was more amenable to empirical observation. This is consistent with the revealed preference approach to analysing economic choices preferred by Samuelson.

However, while Ramsey gambles allow us to determine what a specific individual's credences are, it is not clear how we should extrapolate from *market prices* to individual beliefs. Prices are not set by individuals alone, the way Ramsey gambles are. If we accept that market prices represent the revealed probability beliefs of investors through some form of aggregating process, we have a further requirement to determine how daily changes in prices, or variance, should be connected to subjective probability beliefs.

Markowitz's answer was to frame individual probability beliefs as *beliefs about probability distributions.* Using an approach based on Savage's decision theory (Savage 1954), Markowitz argued subjective probability beliefs are about states of the world in which each state is a different probability distribution. But while our credences in each world are subjective, within each world the probability distribution is *objective*. Markowitz (1959: 258) explains it as follows:

> We shall…assume that there is a finite (though perhaps extremely large) number of hypotheses about the world, one and only one of which is true. One hypothesis about the world might include the statement that stock prices behave as if generated by a joint probability distribution which remains the same through time. Another hypothesis might assert that this probability distribution changes through time in a particular way. Each such hypothesis will be referred to as a possible *Nature of the World.* As these examples illustrate, we admit the possibility that objective random variables and probability distributions may be used in the definition of a Nature of the World.

This passage indicates that Markowitz believed that hypotheses about risk are hypotheses about particular objective probability distributions governing the returns to securities, leaving open the possibility that the objective probability distributions may change over time. His term, "Nature of the World", is similar to David Lewis's "possible worlds" which represent logically possible propositions about which we may have some degree of belief (Lewis 1980). I interpret Markowitz as suggesting that there is a large number of possible worlds, each with an objective probability distribution, about which we hold subjective degrees of belief. He also leaves open that individuals can condition their beliefs over time as a result of experience. He illustrates this with the example of a possibly biased coin that is going to be tossed 100 times. An agent sees three possible states of the world: heads with chance 0.5; heads with chance 0.4 and heads with chance 0.6. He has a credence in these three states of the world of 0.6, 0.2 and 0.2 respectively. This is summarised:

Credence of…	That chance of heads is…
0.6	0.5
0.2	0.4
0.2	0.6

Given these beliefs, when asked what he believes the chance of heads on the first throw is, he answers 0.5. This is a simple compounding of these probabilities: (0.6 x 0.5) + (0.2 x 0.4) + (0.2 x 0.6) = 0.5. If he is asked at *t* what the probability of heads on the *i*th toss is, he would also say 0.5. This is consistent with the starting distribution of probabilities that he sees but does not imply that, by the law of large numbers, he should conclude the probability of heads is 0.5 *a priori*. In fact, he is only 60% “sure” the coin is a fair one. So, although the subjective beliefs he holds are a form of second order probability, Markowitz argues they cannot be merely compounded. There are two reasons in the offing for this: that the probabilities are of two different “types”, i.e. a subjective credence and an objective chance, and that subjective credences may change subject to evidence. Markowitz (1959: 272) goes on:

> …the probability belief he attaches to the statement, “Heads will appear on the second flip,” is not independent of what happens on the first flip. In terms of probability beliefs, the outcomes of the flips are correlated to such an extent that the average of a large number of flips has a variance substantially greater than zero.

This suggests Markowitz has held on to the tentative notion expressed in his 1952 paper that beliefs can be conditioned over time by the evidence. This seems to be the reason that the probabilities in his coin example cannot be compounded. If the "real" chance of heads is 0.6 with enough repeated iterations, we will come to believe that the probability is 0.6 and give up our starting view that it is 0.5. This is a static version of the use of beta density functions in decision theory (Rapoport 2013), whereby a decision maker's beliefs about possible objective probabilities are conditioned dynamically over time as the decision maker becomes more or less certain about any particular objective probability following each event such as a coin toss. However, Markowitz stops far short of developing his ideas to this extent. Rather he seems to be arguing simply that beliefs should be conditioned by the statistical frequency actually seen in the market.

The general idea of conditioning through observation goes back to Ramsey (1990). This idea can be expressed as:

$$P(A|relativefrequency=x)=x$$

The subjective probability belief *P* that *A* is the case, given that the relative frequency of *A* being the case is *x*, equals *x*. Here the assumption is that relative frequency provides the right calibration evidence to condition our probability beliefs. This position is shared by Van Fraassen who explains the process of calibration with the example of a weather forecaster:

> …consider a weather forecaster who says in the morning that the probability of rain equals 0.8. That day it either rains or does not. How good a forecaster is he? Clearly to evaluate him we must look at his performance over a longer period of time. Calibration is a measure of agreement between judgments and actual frequencies. Without going into detail, it is still easy to explain perfect calibration. This forecaster was perfectly calibrated over the past year, for example, if, for every number *r*, the proportion of rainy days among those days on which he announced probability *r* for rain, equalled *r*.

Markowitz appears to have this sort of process in mind in forming beliefs about the variance of returns to financial assets. Investors can be calibrated much like a weather forecaster. Relative frequency is like an expert we can consult to determine what our subjective probability belief should be (Hájek 2012) and if we do so our beliefs will coincide with the relative frequency. This can be dynamic. As we have seen, Markowitz was open to the possibility that the objective probability of returns can change over time, presumably in response to evolving economic conditions.

Markowitz's approach is much like David Lewis's Principal Principle (Lewis 1980). Lewis argued that subjective probability beliefs should conform to objective chances, when they are known. The knowledge of objective chances is not necessarily provided by relative frequencies. It could instead by determined by logical analysis of the setup, noting for example equiprobable outcomes, such as the throw of dice. If the information at hand is that a dice is fair and six sided, we can determine that the chance of any one number is 1/6.

Lewis's approach is important because it distinguishes beliefs about *A* from beliefs about the *chance* of *A*. If *A* is a coin flip at $t - 1$ I will obviously know if it came up heads, but that does not imply that the *chance* of it coming up heads at $t -1$ was 1. I may know that the chance was 0.5, even though I know with certainty that it came up heads. My knowledge that it came up heads, Lewis argues, is *inadmissible* evidence in my determination of what the chance actually was. The Principal Principle allows one to infer credences about frequencies from credences about chance (Loewer 2004). So if you know the chance of heads is 0.5, you can make the claim that the frequency of heads will be 50% in a set of observations. If we take the Principal Principle and assume that historic frequency plays the role of an expert in telling us the objective chance, we can then infer a probability distribution, or variance of market returns, for the future.

This seems to me a reasonable interpretation of what Markowitz set out in his 1959 book. This interpretation is certainly consistent with later practice, particularly as set out by William Sharpe, as I will discuss in the next section.

Markowitz was never explicit in this interpretation of risk. He wrote simply that if one replaced the term "risk" with the "variance of return", "little change of apparent meaning would result" (Markowitz 1952: 89). Donald MacKenzie asked why he had identified risk with variance some decades later and Markowitz had replied simply that he had come across the use of standard deviation this way in the statistics courses he had taken (MacKenzie 2008). But Markowitz, influenced by Savage, is eager to position his approach within the subjective probabilist framework. In his last book-length discussion of MVO (Markowitz, Todd, and Sharpe 2000) he accuses detractors of not having read and understood his chapters on expected utility theory and subjective probability in his 1959 book and reiterates his initial arguments. He makes a somewhat revealing comment about the 1959 book, declaring that he left the theoretical discussion of expected utility and subjective probability to the end because the book was aimed at working professionals who needed to understand how to apply MVO and not why they should apply it. "I imagine that many of these would have been 'turned off' if chapters 10-13 had come first" (Markowitz, Todd, and Sharpe 2000: 52).

Markowitz's equating of risk and volatility in this way implies an operationalist definition of risk. MVO casts the world as consisting of two properties: expected returns and variance. These can be determined by decision makers by consulting historic return frequencies. This operationalism matured as the theoretical finance programme developed, finding its full expression in Sharpe, as I will show next.

6. Sharpe as an operationalist

Sharpe's CAPM was in part developed to overcome the practical difficulty of undertaking the statistical analysis needed for the MVO approach to portfolio construction. CAPM made it far simpler by reducing the problem to one of assessing whether the particular asset in question added to the overall return profile of a portfolio relative to the risk it contributed. The programme developed this intensively practical orientation as financial theorists positioned themselves as providing tools to facilitate investment decision making. This practical orientation supported a commitment to operationalism.

Operationalism is the notion that concepts should be reduced to operational definitions. For instance, there should be no such thing as "length" independent of the act of measuring it. So an operational definition of length would be the result of a measurement using a ruler. This is a distinctly positivist outlook that rejects terms that do not have any clear empirical foundations. Operationalist approaches were influenced by the logical positivism that was popular at the time operationalism was first developed, which demanded that terms in a language must have unambiguous real referents. The relationship between statements in language and the world is a substantial area of philosophical investigation, and operationalism provides one argument. This is that a concept or term is "operationally meaningful" only if it can be characterised by a set of operations, and that set of operations defines it (Hands 2001).

The general idea of operationalism was initially argued by the physicist Percy Bridgman in 1927. He explained his approach as follows (Bridgman 1927: 5, quoted in Hands 2001: 62):

> What do we mean by the length of an object? We evidently know what we mean by length if we can tell what the length of any and every object is, and for the physicist nothing more is required. To find the length of an object, we have to perform certain physical operations. The concept of length is therefore fixed when the operations by which length is measured are fixed: that is, the concept of length involves as much as and nothing more than the

> set of operations by which length is determined. In general we mean by any concept nothing more than a set of operations; *the concept is synonymous with the corresponding set of operations.*

So “length” has meaning precisely in terms of its operational content, and nothing else. Classically, economists measure some feature of the world as

$$m = q + e$$

where the measurement *m* is the underlying “true” property of the world *q* plus some error term *e*. The error term exists because we cannot measure *q* precisely. Operationalism is the claim that *q* and *e* do not exist (Holton 2004). All that exists is *m* – the measurement itself.

The emphasis on a practical, operational approach to social science had adherents among both economists and probability theorists. In economics, Paul Samuelson is often described as operationalist in orientation, having described his objective in his seminal *Foundations of Economic Analysis* to find “operationally meaningful” theorems (Samuelson 2013 (1947)). His view was that scientific theories should describe the empirical evidence, rather than attempting to formulate claims about hidden causes of phenomena (Hands 2001). As noted above, this outlook also supported a behavioural approach to understanding consumer choice. Samuelson’s revealed preference theory relied only on the data of actual consumer choices, rather than any speculation about their psychological states or intentions. Samuelson’s approach was grounded in developing theories that applied to identifiable and measurable properties in the world.

In probability theory, operationalist approaches reduce probability to its measurement. One influential such approach is frequentism; the notion that probability is nothing more than the limiting frequency of phenomena in the long run, developed by Richard von Mises (1957). Von Mises proposed a strict delineation of probability as a concept that can only be applied in cases of large numbers of repetitions of a phenomenon, such as coin flipping. He limited the concept of probability as follows (Von Mises 1957: 11):

> The rational concept of probability, which is the only basis of probability calculus, applies only to problems in which either the same event repeats itself again and again, or a great number of uniform elements are involved at the same time

By flipping a coin or throwing a dice a large number of times we can determine probabilities for "heads" or of the number six coming up on the dice. The limiting frequency of the event ("heads" or number six) is the probability, and there is nothing else to it. This position makes Von Mises an operationalist because the concept of a probability is nothing other than the measurement of its limiting frequency (Gillies 2000).

I described in the previous section how Markowitz proclaimed to support a subjectivist philosophy of probability but in fact embraced statistical analysis of datasets as a means to determine ex ante probabilities. Sharpe's approach makes no such effort at a subjective interpretation and clearly advocates for the use of historic price volatility data as a means to determine ex ante probabilities. Operationalism provides the philosophical support for this strategy, reducing the concepts of probability and risk to this measurement. Markowitz and Sharpe stand as exemplars of this approach and set up this concept of risk as core to the methodology of finance.

There are many arguments in the literature against operationalism and its use in probability theory. Von Mises, for instance, was heavily criticised by Richard Jeffrey who described Von Mises's frequentism as a "metaphysical conceit" because it fudged the difference between sets of data and infinite repetitions (Jeffrey 1992). Clearly there is no such thing as an infinitely flipped coin and no dataset can be infinite either. The probability measured from any particular set of finite data will shift by some small margin for each additional observation. Also, Von Mises depends on the notion of a "collective" as a set of data that is generated under stable "generating conditions". It is impossible to be an operationalist about the generating conditions because they seem to be like dispositions, inherent features of a particular set up (Childers 2013). Talk of inherent features is precisely what operationalism is meant to avoid. So while Von Mises is an operationalist in respect of the probabilities that can be determined from a set of data, he has something like a propensity view of the generating conditions for that data that is closer to the Keynesian propensity theory of probability (Keynes 1921). This philosophical dissonance may explain why finance has embraced an operationalist concept of risk with little consideration of the implications for the nature of the financial system, for the generating conditions of the data. Classical frequentism ultimately depends on a particular set up to generate a random series. The set up must be imbued with the propensity to create the probability distribution that is then be measured. While frequentism is meant to be operationalist at heart, the demands of the background setup imply a strict metaphysics, one that demands the environment consist of stable generating conditions. This has important implications in finance, which I will come back to in section 9 below.

There is a further metaphysical slight of hand at work in theoretical finance to note here. We have seen that Markowitz's philosophy of probability conceived of expected variance as a subjective degree of belief in an objective probability distribution. The operationalism that is clear in Sharpe collapses the difference between the subjective and objective. Expected variance is defined by its measurement. While it is of course the historic variance that is actually measured in any dataset, the metaphysical slight of hand is to hold this as equivalent to expected variance. The ex post has become the ex ante.

This makes clear the ontological commitment at the heart of finance. The world must be such that it has the propensity to generate a stable frequency. In finance this frequency is analysed in terms of a probability distribution that remains stable over time. By studying the distribution of returns to financial assets, the distribution is measured and assumed to apply in future. The internal metaphysics of finance demands a world that is risky only in this sense, admitting of no other concept of risk. There is no room for Knightian uncertainty in modern portfolio theory. Its methodology has no way of accommodating it.

7. Finance as a Lakatosian research programme

So far I have argued that one hallmark of the methodology of theoretical finance is an operationalist approach to concepts. Risk and return are defined as measurable statistical properties of data sets. I suggested above that this entails a commitment to a particular metaphysics, a world that has the propensity to produce objective probability density functions. I will return to this ontological claim in section 9, below, but first I want to contrast the discussion so far to an approach to methodology that was highly influential in the second half of the last century, Lakatosian analysis. I will show that the classical Lakatosian approach is not very successful in describing finance methodology, but I reconstruct it to accommodate the methodological approach discussed above.

Lakatos aimed to reconcile the logical approach of Popper and his focus on falsification (Popper 1959) with the sociological approach of Kuhn (Kuhn 1962). To Lakatos, science progressed via the methodology of scientific research programmes (MSRPs). Research programmes are in constant tension with each other, ultimately failing or succeeding based on their empirical success relative to rival programmes. Programmes were defined by their "pornographic metaphors" (as Hacking 1979 described them) of hard cores and protective belts. The hard core is the theoretical heart of a programme that is defended at all costs, while the protective belt consists of derivative propositions and novel facts that can be put to the test. A "negative heuristic" says that researchers may not question the hard core, while a "positive heuristic" describes a process for updating the protective belt. A programme is considered "progressive" if the novel facts predicted by propositions in the protective belt are met with empirical support, or "degenerating" if they encounter disconfirmation. A programme exhibits a progressive "problem shift" as it moves from one theory to a more successful theory, replacing what came before it.

Debates over Lakatos and the wider methodology of economics had an extensive airing in the 1970s through to the mid-1980s and had a substantial impact on economists over this period. De Marchi, one of the key proponents of Lakatosian analysis of economic methodology, argued that economists adopted MSRPs, Lakatos's main methodological account, as their own "self-image" (De Marchi 1991: 2) such that MSRP had a performative[3] effect on economics, driving the way economics worked.

Lakatos's work was quickly popularised among economists thanks to the efforts of one of his students, Spiro Latsis (1976) who helped Lakatos organise a conference in 1974 that brought economists and physical scientists together to discuss his work and economics. Lakatos's novel facts fitted well with Friedman's influential methodological approach that focused on the prediction of phenomena not yet observed. Blaug (1992) followed the 1974 conference with a systematic study of important theories in economics from a Lakatosian standpoint. His perspective, however, lent strongly toward Popperian falsificationism, which remained in his view the way to "discriminate between wheat and chaff" in scientific research programmes (Blaug 1992: 264). A cottage industry soon emerged in applying Lakatosian analysis to various programmes in economics, ranging from the theory of the firm (Latsis 1972; Ahonen 1990) to game theory (Bianchi and Moulin 1991) and a great many others (Hands 2001). Among them is a rare one dealing specifically

[3] Performativity will be the subject of extensive analysis in Chapter 4. I use the term loosely here.

with theoretical finance by Reinhard Schmidt (1982) which I will consider in more detail in the next section.

Much of this literature trumpets the success of Lakatosian analysis in *validating* the methodology of economics, rather than a serious effort to appraise that methodology from a Lakatosian standpoint. Most of the case studies determine that Lakatos's MSRP "fits" research programmes in economics well. Little of it attempts to appraise any particular programme in economics in order to determine whether it meets Lakatos's standard in successfully predicting novel facts. It is taken as given that economics and the programmes are successful, and the question is whether an MSRP approach validates this success (the Schmidt analysis of finance is a good example of this in action, as we will see in the next section). This embrace of Lakatos was enthusiastic because the main alternative on offer, Popperian falsificationism, suggested that empirical failure of economic models should be their death knell.

This flourishing of Lakatosian work in economics went into decline from the mid-1980s, marked by a symposium in 1985 which revealed a level of hostility to Lakatosian ideas in economics. This was largely due to the growing recognition that Lakatos's "novel facts" are seldom provided by important economic theories (Backhouse 2012) and so programmes cannot be progressive. Economic theories and models often did not conform to the notion of progressive problem shifts, particularly in that economic models did not appear to replace ones that came before them, but rather complement, fill in gaps, or find specific applicability without contradicting any other more general models. When models or theories met empirical disconfirmation, the response was usually to explain away such disconfirmation by arguing that the *ceteris paribus* clauses that surrounded the model had been violated. Such responses seemed to fit Lakatos's hallmark of a degenerating science, one that does its work by explaining away failures of prediction, rather than building off of successful predictions. The vision of economics as a science that accumulates explanations has been recently argued by Dani Rodrik, who characterises economics as a “horizontal science” that progresses by assembling a wider and wider library of models, rather than a vertical science that attempts to make progressively more accurate models (Rodrik, 2015). I will return to this idea in Chapter 4 and argue that models might be inconsistent with each other, yet individually retain explanatory power and be useful in different situations. It is clear that this library conception of economic theories is fundamentally incompatible with MSRP and the idea of the success of a programme being determined by its progressive problem shift.

Another frustration was that “hard cores” appear to overlap, which was a problem for delineating different research programmes. Delineation is a requirement of Lakatosian analysis because programmes are in tension with each other, either progressing or degenerating depending on their empirical success. Kuhn had argued that rival paradigms were incommensurate, by which he meant fundamentally incompatible in that the assumptions and methodology of one paradigm contradicted those of rival paradigms. Lakatos did not accept this, arguing that scientists could work on multiple programmes at the same time without the cognitive dissonance that incommensurability implies. However, Lakatos thought that ultimately a programme was always degenerating or advancing depending on whether the problem shift was progressive or not. If multiple programmes share hard cores, it becomes difficult to know what theoretical claims are progressing or degenerating. Economics seems to follow a different methodology that embraces multiple different programmes using similar techniques and foundational assumptions, some overlapping and each individually shedding some light when explaining particular economic phenomena, but none capable of being declared a general advance over another one.

8. Schmidt’s Lakatosian analysis of finance

Among the applications of Lakatosian analysis to economics in the decade to the mid-1980s was one on theoretical finance by German finance academic Reinhard Schmidt. It is clear from the outset that Schmidt considers finance to be a successful research programme, and his project is to appraise how well Lakatoisan analysis does in explaining and justifying this success. Schmidt says that the flattering light that theoretical finance is cast in by the Lakatosian account renders it preferable to any Popperian alternative. “A very brief look at the recent history shows that finance can be considered as a *successful* research programme” (Schmidt 1982: 399).

Schmidt describes the finance research programme as “capital market theory” which he says has two central claims (Schmidt 1982: 395):

> 1. Excess returns on assets are unpredictable on the basis of publicly available information, and
>
> 2. The expected return on an asset is a linear function of its systematic risk only.

He writes that these two claims stem from finance’s core commitment to equilibrium in financial markets and are the main predictions of what he calls “capital market theory”. The two claims are key corollaries to the CAPM and the MVO. Systematic returns that are captured by the “expected return” of the model are the only predictable returns. The second claim is particularly clear from the CAPM which provides an explicit way to determine expected return from the systematic risk of an asset. Both of these claims also relate to other key theories in the programme including the efficient markets hypothesis, and the related random walk hypothesis which argues that residual price movements are random (Fama 1995).

Schmidt argues that the two basic claims he describes for the finance research programme meet the criteria for being good science set out by Karl Popper because they are capable of empirical assessment. They are in the protective belt of the programme and verifying them is part of the positive heuristic. In this respect, he claims, finance is unusual compared to the rest of economics which relies on *ceteris paribus* shielding clauses, which render economic models immune to empirical disconfirmation and falsification. This is clearly inconsistent with my view that the MVO involves Aristotelian idealisations, particularly when it is seen as a positive model of the sort necessary to be Popperian. But in any event, Schmidt does not think Popperian approaches are adequate with respect to finance because they are unable to determine whether apparent disconfirmation represents a market that is out of equilibrium, or whether the claim of equilibrium is false. He argues that when empirical assessment finds fault in finance theory, one should not reject the theory on Popperian grounds but rather hold on to the theory while working to make it more precise: "Finding a flaw in an existing theory is no reason to eliminate the theory – as no scientist would have done – but to investigate whether tolerating the deficiency prevents or facilitates further progress" (Schmidt 1982: 398). A similar argument has been made against Popperian methodology as a basis for assessment in economics as a whole, given that empirical disconfirmation can be explained away as a violation of the *ceteris paribus* clauses of economic models (Caldwell and Hutchison 1984) and that economists respond to empirical disconfirmation by tweaking rather than rejecting their models (Hausman 1994).

Having argued that Popper is inadequate, Schmidt embraces Lakatos's MSRPs as a preferable approach to assessing theoretical finance. He applies the Lakatosian scheme to the programme as follows: the hard core of the programme is its commitment to the concept of equilibrium in force in financial markets and at the level of individual decision making. Its protective belt consists of assumptions about the shape of investor utility functions (e.g. risk aversion), transaction costs, and return-generating mechanisms, with the central claims being those discussed above. He constructs the development of finance as a "progressive problem shift", beginning with Markowitz's MVO, to the CAPM and on to multi-period accounts of market price movements based on the efficient markets hypothesis. The Black-Scholes model is a further "outstanding example for the dynamics of scientific progress in finance" (Schmidt 1982: 400). Schmidt characterises these as progressive in the sense of gradually improving theories, each providing a better description of the world. He claims that theoretical finance works in this progressive way, steadily accumulating new insights and approaches to financial markets, all supported by the hard core.

This Lakatosian outline of the programme can be challenged by Lakatosian standards. I will outline three problems with the analysis Schmidt provides.

The first problem is Schmidt's claim that the hard core of finance is equilibrium, and that this is a theoretical proposition. "Equilibrium is not only the *central theoretical concept* which makes financial economics an empirical science, but it is also a *fact of life* – at least on the major world stock market[s]" (Schmidt 1982: 396). This appears to be begging the theoretical question – equilibrium is a fact of life and the central theoretical concept. If it is a fact of life then what theoretical work is being done in theoretical finance? Is it merely descriptive of the world? Clearly an answer in the affirmative would be unsatisfactory – if finance is a progressive research programme it needs to propose theories that contain novel facts. Indeed, one mark of difference between modern portfolio theory and the work produced by academic finance before it is that it took finance beyond descriptive work such as accounts of the quantity of money supply or details of financial instruments. So Schmidt's proclaimed core of finance is not really a theory at all, and not one that can be seen as a rival to other programmes that are advancing their own theories. When one remembers that Lakatos's analysis was written for a time when quantum and relativistic theories in physics were being appraised against each other, it becomes clearer how the approach is unproductive in finance. Relativity and quantum physics have core theoretical claims that are in tension, and both are aiming to progress by drawing out novel facts about the world and then corroborating them. Finance has no such core that is in tension with any other programme. It is, as Schmidt says, merely a fact of life rather than a hypothesis or theory.

This is related to the second problem, which is that even if we take equilibrium to be a theoretical claim, it is certainly not unique to finance. It is pervasive in a great deal of economic theory and different schools within it, mostly commonly related to the basic idea that prices represent a balance of supply and demand. Backhouse (2006) argues that Lakatosian analysis requires that the hard core of a programme be a defining and unique feature of the programme. It is the hard core that is protected from challenge within a programme while it poses research questions (the "positive heuristic") that make for the novel facts and protective belt.

Again, it is hard to see how we can therefore ever see finance as a programme distinct from any other in economics, if we must take equilibrium as the core. The conclusion is that either equilibrium is not the core of finance, or Lakatosian analysis is not a good account of its methodology, or both.

The third problem with the Lakatosian analysis provided by Schmidt is his notion of the progressive problem shift. He sees the accumulation of theories in finance as progressive, with the Black-Scholes model a pinnacle achievement. Lakatos provided two senses in which a problem shift may be progress – theoretical and empirical (Lakatos 1968). Theoretical progress happens when theories build on one another such that each new theory has greater content than the one that preceded it. Empirical progress happens when successive theories predict novel facts that are then corroborated. Schmidt's account implies that he sees the progress in finance of being primarily of the theoretical kind. Certainly it is the case that the progress of models and theories in finance exhibits increasing complexity in a sense. The Black-Scholes model appears in a form that is mathematically more complex than, say, the CAPM from which it was derived. However, one could not say the same of the MVO, which, as we've seen, exhibits a combinatorial explosion as the portfolio of assets to be analysed grows. The CAPM was therefore not considered a *theoretical* advance over MVO, but was rather celebrated as a *practical* advance because it made it much easier to make decisions over which assets to include in portfolios.

Schmidt concludes that his account is positive for financial economists (Schmidt 1982: 403):

> This position can be expected to be welcomed by financial economists for two reasons. It gives methodological support for their work, or in other words, the intuitive feeling that financial economists do good work can now be ‘rationally reconstructed’ in methodological terms. And besides, it protects their work from short-sighted grumbling with a pretended backing from philosophy of science.

This is typical of the Lakatos-inspired analysis of economics at the time in that it casts existing practice in a flattering light. His objective is to lobby for Lakatosian philosophy of science as a preferred basis of methodological assessment for financial theorists, rather than to assess the adequacy of Lakatosian analysis in assessing the finance research programme.

This is a meta-methodological point, an appraisal of a methodological account rather than the use of the methodological account to appraise theories. In comparing Lakatosian and Popperian analysis the meta-methodological point was usually argued by reference to the success of the methodological account in reconstructing the history of science, an approach known as the methodology of historiographical research programmes (Lakatos, Worrall, and Currie 1980). Such an approach requires the long gaze of history in order to see which research programmes have survived, for hard cores to have coalesced and for progress from one programme to the next to follow degeneration and advancement, in order to determine if MSRP provides an adequate account of this history. In such an analysis, the intuitive feelings of scientists about their work don’t seem to be important.

Schmidt describes behavioural theory and finance as two competing research programmes. As discussed, in Lakatosian analysis, research programmes are always either progressing or degenerating, striving to become dominant. It is interesting that over 30 years since Schmidt’s paper, it would still be difficult to determine which of those programmes has gained or lost. Behavioural research remains vibrant in finance. The financial crisis may have sunk the star of theoretical finance somewhat but it remains an important source of insight on financial markets. The two programmes might better be seen as asking different questions rather than competing explanations for the same phenomena, which would fit Rodrick's depiction of economics as a horizontal science. Ultimately it is this aspect of financial research that renders Lakatosian analysis inert.

The methodology of finance has at its core not a theoretical claim, but rather a way of analysing the world. The hallmark of its methodology is its operationalist approach to data and the concept of risk. Modern portfolio theory is a series of complementary rather than progressive models and theories about financial markets with a strong normative leaning. Threats to the programme come not from challenges to its novel facts, but from attacks at an epistemological level; challenges to its assumption that analysis of data provides a sufficient insight into risk. In section 10 I will consider how the programme deals with empirical challenge, and will argue that responses that challenge this operationalist core of modern portfolio theory are the most threatening.

9. Theories and ontological commitments

Ontology has played an important role in several discussions of methodology in social science including Cartwright (1995b), Lawson (2015); (Lawson 1996, 2012) and Mäki (2001). Mäki explains how a theory's ontological assumptions are different to those of the theorist (Mäki 2001: 6):

> Questions about the economic world view can often be transformed into questions about economic theories, taking on the general form, 'What does theory T presuppose concerning P?' For example, 'What exactly does theory T presuppose about the capacities and dispositions of economic agents, or of the market mechanism?' A refined account of the presuppositions of theory T does not necessarily serve as an adequate account of the ontological convictions of an economist using T: deep in her heart, the economist may not fully believe in the world view presupposed by T.

Mäki supports analysis focused on the "presuppositions" of a theory, the assumed nature of institutions and causes that underpin the theory. This represents an entailment approach to ontological commitments, where the focus of analysis is what entities, categories and properties must actually exist in the world for the proposed theory to be true (Pratten 2007). In economics, theories or models may rest on idealisations or patently false entities, assumptions that have an instrumental purpose in allowing theories or models to be parsimonious or easy to apply (Friedman 1953). This might be taken to mean that economics need not be committed to any particular things or categories actually existing. But despite acknowledged unrealism in (some) assumptions, there are nevertheless ontological requirements of the world that are entailed if the theory or model is to fit the world. The model of a perfectly competitive market, for instance, has a false assumption like infinitely divisible goods, for example, but requires the world to consist of agents competing with each other to supply and acquire goods in a market, for it to have any explanatory or predictive power over actual behaviour. The false assumptions stem from its Galilean idealisations, those parts of the model that are idealised in order to make it more tractable. Infinite divisibility of goods is an assumption in many economic models for convenience, because the actual divisibility of goods is too varied to be accommodated in a model designed for general application. However, the main theoretical import of the model, for instance that prices will be set where demand and supply reach equilibrium, require a specific ontology.

The claim I will make of theoretical finance is that its ontology is bound up in its assumptions. The equating of risk with volatility initially had the purpose of explanatory parsimony and operational practicality. But it has become an implicit claim about the way the world actually is. There is a critical realist flavour to my analysis, a philosophical outlook that emphasises the ontological implications of theories in social sciences (Bhaskar 2014; Mäki 2001). This body of work is rather diverse and often conflictual (Latsis 2015; Pratten 2007) and I will draw on it only to bring out the role of ontological commitments in theoretical finance. The commitments I will describe are not in the technical sense philosophers usually speak of, the existential quantifiers of theories much discussed by Quine (1948), but rather an understanding of the "internal metaphysics" of the programme (Pratten 2007), the set of beliefs about what the world is like that are necessary for the programme to function.

Nancy Cartwright writes (Cartwright 1995a: 73):

> A probability distribution as it functions in scientific theory is a nomological item. It is not part of the data, but a putative law of nature that (depending on your view) summarises, governs, or generates the data. You shouldn't think that the probability approach avoids ontology. It just chooses one ontology over another.

My claim for theoretical finance is that its choice of ontology comes through in its approach to probability. Cartwright argues that a probability distribution depends on a particular causal structure. Probabilities are generated by a "chance setup" that can have various degrees of stability and reliability. A car engine, for example, is a stable causal structure by which expectations of the probability of acceleration can be reliably formed given the level of the throttle. Social and economic contexts may not have stable generating conditions because their causal structures are much more complex and dynamic. Yet using probability distributions requires some form of order on this complexity, or at least the appearance of order. It requires some law of nature that summarises, governs or generates the data in a way that fits a pattern.

A probability distribution in finance, as in most statistical analysis, is assembled from a sample. The distribution "fills in the gaps" between different observations, creating the smooth bell curves around the mean that are often used to illustrate the probability distribution. While the observations are discrete, the distribution provides a probability for any observation, including those that have not actually been made. Often the distribution is also assumed to be Gaussian, without much further comment. It implies an ontological commitment to the world being a stable generating setup that provides the distribution.

We have seen that theoretical finance rests heavily on its notions of expected return and expected variance, which provides the mean and standard deviation for the distribution. These have operational definitions, determined from the analysis of market return data. But they imply a commitment to a stable set of generating conditions, such that the probabilities of the past will be the probabilities of the future. There is a "purchase on invariance" (Cartwright 1995a) in the ontology of the world that theoretical finance rests on. It is this approach to analysis of financial data that is the core feature of the research programme. Its operationalist approach to the concepts of risk and return are critical to its methodology and epistemology, but the reliance on probability also determines its ontology.

10. Empiricism, disconfirmation and theoretical finance

The CAPM in particular has been subjected to a great deal of empirical testing. Fama & French have become the most cited (2004, 1992, 1993) but others like Banz (1981), Douglas (1969), Black, Jensen, and Scholes (1972), Miller and Scholes (1972), Blume and Friend (1973), and Stambaugh (1982) have published important empirical assessments of the model. MVO has also been tested, particularly of its normative implications. For instance portfolios that are optimised are compared to portfolios assembled using a naïve 1/n approach, in which each asset is held in equal proportions regardless of correlations or variance, generally finding that such alternative strategies perform no worse than MVO (DeMiguel, Garlappi, and Uppal 2009).

The first observation to make about this body of literature is that empirical testing in theoretical finance is considered a worthwhile endeavour. This is consistent with a Popperian-cum-Friedmanite methodology that was readily applied by the financial research community and Sharpe's own view that his theory should be assessed according to the "acceptability of its implications" (Sharpe 1964: 434).

While none of these studies explicitly considers the methodological foundations of the CAPM, it is clear that they see the CAPM as a positive theory of market prices, and not just normative. This is natural enough. Even as a normative tool in the hands of financial practitioners, it would not be very useful if its advice for portfolio construction was not subsequently borne out by market returns.

These empirical investigations typically involve examining market data to test the CAPM's prediction that returns to any asset (or portfolio of assets) is a function of the relative riskiness (i.e. beta) of that asset to the market as a whole. Fama & French (1992) provided severe criticism of CAPM, finding it has no explanatory power at all, based on comparing different time series of market returns. They found instead that other explanatory factors led to better success in predicting returns, including price to earnings ratios, ratios of book value to market value, and the size of the firm. In their (1993) article they use time series regression and this time find the "market factor", their term for beta, has some explanatory power, but that the size of the firm and book value to market value ratio are also important factors in predicting returns. This has since come to be called the "three factor model" which is often considered to be a development of the CAPM, rather than a substitute for it, though there are interpretations that suggest it challenges the core of CAPM, and theoretical finance, which I will come to.

The three factors are reflected in the equation (Fama and French 2004):

$$E(R_i) - R_f \ = \beta_{iM}\big(E(R_m) - R_f\big) + \beta_{is}E(SMB) + \ \beta_{ih}E(HML)$$

This should be familiar as the standard CAPM model described in section 3 with two additional terms added. SMB is "small minus big" which is the small firm premium, determined as the difference between returns to a portfolio of small stocks and a portfolio of big stocks. HML is "high minus low", which is the difference between a portfolio of stocks with a high ratio of book value to market value, and a portfolio of stocks with a low ratio of book value to market value. There are also two additional betas for each term, determined from the slopes of the multiple regression equations.

There is no attempt at a theoretical foundation to the three factor model in Fama and French (1993); it is empirical rather than theoretical. In the style of Francis Bacon, the three factor model is derived from empirical observation. It is a work of statistical analysis that finds regularities, rather than a model of a causal mechanism. The results can broadly be considered Humean: the three factor model provides a regularity rule that connects beta, firm size, and a particular feature of firm value to the future returns. There is no theoretical account of why firm size and value affect returns in this way.

Such empirical investigations of the CAPM assume that historic data makes for a good testing ground. Analysis usually involves determining expected return by examining one time series and then studying another series, usually the next period of returns, to see if the expected returns and volatility materialises. So for example, a researcher can determine beta from the share price history for a period of five years, and then assess whether the CAPM-derived required return is in fact realised over the next five years.

Other studies have identified other factors that arguably provide better predictive power over stock returns, such as that stocks with low price:earnings ratios tend to have higher returns than those predicted by CAPM (Basu 1977); that small capitalisation stocks perform better than predicted relative to large capitalisation stocks (Banz 1981); that companies with high financial gearing have returns higher than those predicted by CAPM (Bhandari 1988). Models based on additional factors like these, as well as the three factor model of Fama & French, are collectively called "multi-factor models" to distinguish them from the single factor of beta that applies in the CAPM.

Are multi-factor models a challenge to the CAPM or a development of it in a way that might be seen as progressive in the Lakatosian sense? The answer depends on the way the empirical evidence is treated and the implications for the probability ontology of the programme. There is a methodological tension between the CAPM and multi-factor models. Fama & French's empirical methodology leads to a different "type" of model. The CAPM is a theoretical model: it proposes a mechanism by which market equilibrium obtains when prices adjust for the risk of securities. As discussed above, the CAPM holds that the volatility of the market *causes* price movements of the individual security, the so-called "systematic" component of risk. It also assumes that investors are rational decision makers seeking to trade off risk and return. In contrast, the three factor model is an empirical model that identifies regularities but provides no causal mechanism. In the language of Cartwright (2002), the CAPM is a "nomological" model while the three factor model is an empirical model, a "free-standing association", which, to Cartwright, is no model at all. Such models lack any story of a causal process and posit merely a statistical regularity. Such statistical regularities can be found in datasets, but that can be fallout from data mining and without an argument for a causal mechanism, there is no way to determine whether the regularity is real or an artefact of the particular dataset.

The three factor model is an example of "measurement without theory", an empiricist approach to economics, famously criticised by the economist Tjalling Koopmans (1947) in respect of studies of the business cycle. In criticising a book by Arthur F Burns and Wesley C Mitchell, *Measuring Business Cycles*, Koopmans wrote (1947: 161):

> The approach of the authors is here described as empirical in the following sense: The various choices as to what to "look for," what economic phenomena to observe, and what measures to define and compute, are made with a minimum of assistance from theoretical conceptions or hypotheses regarding the nature of the economic processes by which the variables studied are generated.

In other words, there is no theory involved in the analysis of data. Koopmans drew an analogy with the "Kepler stage" and the "Newton stage" of theories of planetary motion. The former identified empirical regularities that allowed for predictions of planetary motion. The latter identified fundamental laws of attraction of matter that provided theoretical insight as to why the regularities occur. Koopmans criticised Burns and Mitchell for delivering a work of Kepler-style empiricism and that "fuller utilisation of the concepts and hypothesis of economics theory… *as a part of the process of observation and measurement* promises to be the shorter road, perhaps even the only possible road, to the understanding of cyclical fluctuations" (1947: 162, italics as per original).

The absence of a theoretical explanation for the additional terms in the three factor model was an important focus of further research, with various authors vying to become the Newton to the Kepler position staked out by Fama and French. These were done both within the theoretical finance research programme and outside of it. Methodologically, the impact on the CAPM depends on the detail of the theory proposed, how the story is told to explain the regularity posited by the three factor model. Some of these explanations immunise the core, seen as the operationalist methodology and associated ontological commitments, while others directly threaten it.

10.1 Immunising the core

Efforts to protect the exemplar theories of finance from empirical refutation are largely Millian. Mill described economics as a science of tendencies. These are causes that play out in real economies but can be countered by other tendencies so theorised tendencies aren't always seen when observing real economies. "All laws of causation, in consequence of their liability to be counteracted, require to be stated in words affirmative of tendencies only, and not of actual results" (Mill 1884, p 319, quoted in Hands 2001: 20). Mill's methodology implies that any particular contrary empirical observation need not amount to evidence against a theorised tendency. For finance theory, a Millian defence is to hold the CAPM as correct, and explain the empirical failure as the influence of tendencies outside of the model. These can be behavioural, such as that investors are overly pessimistic about firms that have done badly recently, creating the value stock phenomenon, or that growth stocks are much more glamourous so attract more investor support than their riskiness demands (Lakonishok, Shleifer, and Vishny 1994). Other influential behavioural explanations along these lines include Bondt and Thaler (1987) and Haugen (1995). Alternatively, the outside factors can be market frictions like trading costs and liquidity shortages (MacKinlay 1995). While these responses are often invoked to criticise the CAPM as “unrealistic” they are vulnerable to the Millian response that the tendency proposed by the model has not been falsified. A further defence comes from Milton Friedman’s (1953) argument that unrealistic assumptions in a model are not only no barrier to empirical success, they may make the model more tractable and therefore more useful. The model itself therefore retains its status in the programme.

Another immunisation strategy is to blame the empirical failure on measurement problems. At first this seems effective. The strategy is to blame the data as inappropriate. For instance much CAPM analysis uses capital market indices as a proxy for the “market portfolio” at the heart of CAPM analysis, as I did with the example above that used the FTSE100 index. But any particular index is only a subset of the whole market, broadly construed. According to such a defence, the CAPM is right, it is only the dataset that is wrong. The most cited of such responses is Roll (1977). It is difficult, though, to push this response very far. It is hard to see how any possible measure of the market portfolio might escape this objection. It also flies in the face of the operationalism at the heart of the methodology. There has to be *some* way of measuring the market portfolio, and whatever that is should be the definition of the market portfolio. Given that the CAPM is often used as a normative decision rule for individual investors, it might even be argued that the “right” market portfolio is actually the complete choice set faced by the individual at the relevant moment in time. Such an abstract unmeasurable notion of a market portfolio would encounter a negative heuristic in the form of “don’t challenge the operational definitions”.

10.2 Threatening the core

These same strategies can lead to trouble for the core, as I’ve defined it. The three factor model, for example, argues that there is more to risk than variance, taking into account the size of companies and the accounting value of the companies’ assets net of liabilities (known as the book value). While the model is simply an equation without any claimed mechanisms, it is incommensurate with the CAPM’s concept of risk, measured as variance, as the sole driver of the systematic returns to an asset. This is a direct challenge to the operationalist methodology. If risk is affected by multiple factors then it is a different concept to that at the core of the theoretical finance research programme. This remains empirical in the sense that the other factors assessed as relevant are still measurable, but the operationalist treatment of risk can no longer be maintained. The three factors are all different contributors to risk, but risk itself has become a theoretical property of assets, it is now an abstract concept. Measurement of it through multifactor models is only approximate; effectively an error term is reintroduced and measurement is no longer the same as the concept.

Further arguments threaten the rationality assumption in theoretical finance, according to which rational investors minimise risk and maximise return. For example, some have argued that investors must have a preference for risk at least in some respects (Kahneman and Tversky 1979) or that utility is affected differently by downside risk and upside risk so the symmetrical nature of variance makes it invalid (Roy 1952). However, the CAPM can again be defended from such behavioural claims by emphasising its normative aspect; that while individuals may in fact act for irrational reasons, that is not a reason why they should do so. Such a response works only if one holds that eventually the market corrects such that prices do in fact reflect rational expectations about risk and return. So the positive claim at least has to be that investors are rational in the long run. If the market does not eventually price in rational expectations about risk and volatility, there is no useful normative sense in which investors should invest as if it does. As Keynes is supposed to have remarked, “There is nothing so disastrous as a rational investment policy in an irrational world.” It can be argued that both Roy and Kahneman and Tversky are making long-run claims, that investors have systemic and permanent biases.

Whenever a theory claims that there are other risks that are not captured by MVO and CAPM, the key conceptual commitment of theoretical finance is threatened, along with the ontological commitment to a world that generates a stable probability distribution. That happens the moment we concede that other factors constitute risk.

11. Conclusion

Theoretical finance is normal science in a Kuhnian sense. It is widely understood, has a clear set of beliefs, a language and a programme of work. It is a research programme, the core feature of which is an operationalist approach to the concept of risk and an associated ontological commitment to an objective probability distribution generated by stable conditions. This core sits at the heart of modern portfolio theory and the practices of the financial industry. In the rest of this thesis I will explore how this core affects the financial system and theoretical practice.

In this chapter I began by outlining Markowitz’s MVO model and Sharpe’s CAPM. I argued that Markowitz, while proclaiming to be a subjective probabilist, in fact laid the foundations for what, by the time of Sharpe, had solidified into an operationalist approach to risk that is fundamentally frequentist in philosophy. This means that risk can be measured as variance and remains stable over time. The frequentist philosophy implies that the environment has the propensity to generate the probability distribution that is revealed through the variance and mean returns of asset prices in the market.

I considered Lakatosian approaches to economics and Schmidt's Lakatosian analysis of finance, and argued that it does not provide an adequate methodological account because finance has no identifiable progressive problem shift or theoretical hard core. I argued that it is preferable to characterise the methodology of finance through its operationalism and ontological commitments.

I examined empirical testing of financial theories and argued that there are responses that protect the core, but also ones that threaten it. Protecting the core can be achieved by explaining away the disconfirmation as a violation of the *ceteris paribus* conditions for the model, that outside tendencies have interfered to disrupt what would otherwise have occurred. However threats to the core consist in undermining the operationalist approach to risk, providing explanations for market movements that depend on risk being an abstract concept rather than a specific measurable property.

Having set out these methodological and ontological foundations of theoretical finance as a research programme, in the next two chapters I examine in detail alternative conceptions of social ontology. This work allows me to return to methodological discussions in Chapter 5 where I consider how the models of theoretical finance are used in the design of social institutions and whether the ontological commitment to variance provides a sufficient basis on which to build financial institutions.

Chapter 3

The Ontology of Money

1. Introduction

In the previous chapter I argued that theoretical finance has a core ontological commitment to the world being one that produces stable probability functions. In this chapter I focus on general theories of social ontology. This will enrich the arguments in the subsequent two chapters in which I argue for a specific ontological role of theoretical finance.

Philosophical argument on the nature of the social world can be broadly divided into two types (Guala 2016). The first type focuses on the rules that are connected with institutions. In these arguments, the rules *constitute* social reality. Without rules, our institutions would not exist. The second type can broadly be described as equilibrium accounts. These see institutions and other aspects of the social world as the result of a balance of forces, including the competing interests of individuals, historic conventions, new technologies, human biology and other institutions. In the latter accounts, the rules are endogenous to social reality, they are the result of institutions, not the cause of them.

In this chapter I set out a negative argument against the most prominent rules-based account of social reality, that developed by John Searle (Searle 1995, 2010) and use that negative account to motivate for a form of equilibrium account, specifically in regard to money and the financial system. Searle argues that institutional facts come into existence when duly empowered individuals declare them to exist and assign certain functions to brute physical facts such that they play a social role that is collectively recognised and intended. Using this position as my foil, I argue that the deficiencies in Searle's account can be addressed by an alternative account that depends on individual beliefs, including beliefs about others' beliefs, and requires no declarative speech acts, physical objects or collective intentionality. I argue that money is a social convention used to quantify and express value, and many different physical and non-physical things and concepts can play the part. Individuals' qualitative subjective sense of value is synthesised with the experiences of exchange with others to arrive at an intersubjective measure of value that appears to us to be an objective fact in the world. That measure is money, and it comes prior to any physical representation of it.

I present two main arguments against Searle's account. First, drawing on an argument first developed by Barry Smith (2003), I argue that Searle's ontology is vulnerable to what has been called the "objection from freestanding Y terms". A Y term, in Searle's account, is a social fact that supervenes on some physical fact which he calls an X term. His claim is that every Y term *must* supervene on some X term. My claim is that we do not necessarily need *any* particular physical objects to exist in order for money to exist. I argue that money is an example of a freestanding Y term – it can exist without any physical thing standing for it.

Second, I argue that Searle's account has no way of accommodating institutional collapse. Searle focuses on the creation of institutional facts but does not provide any way that institutions can also cease to exist, particularly when the collapse is surprising or contrary to the will of society broadly. I reject one putative Searlean response to this objection which is that such collapse is an example of "fallout" from social ontology.

Any monetary system is subject to extensive regulation, but these regulations do not constitute money. The declarative speech acts that Searle focuses on are, in my account, part of the regulatory rules that a society may develop to try and govern conventions for quantifying and representing value. They come from institutional structures such as governments, and represent efforts to regulate other institutions.

Despite the existence of such rules, conventions may break down, such as in cases where there is a massive debasement of the link between money and individuals' sense of value, such as in cases of hyperinflation. My account provides a mechanism for such institutional collapse, even in the face of speech acts intended to shore up institutions. On my account the creation of money is best understood as the development of a commonly understood concept to quantify value. That concept is then used to allocate value between people. Representations of the concept, on ledgers or by designated physical tokens, serve as a means of recording these allocations. No physical thing is necessary for money to exist for this purpose, though at some point the practical constraints to remembering such allocations make it necessary to turn to external record keeping devices. These devices form the monetary system.

I argue that beliefs about the connection between money and value are subject to constant updating and therefore the relationship between value and money is dynamic. In section 6, I argue that money is intersubjective and emerges spontaneously in societies, drawing on the arguments of social philosopher Georg Simmel. I consider one potential Searlean objection to my account, which is the objection from an infinite regress of beliefs about others' beliefs. I reject this, arguing that the infinite regress is not required by a series of mutual beliefs. We can still form concepts that can be widely understood in society without a complete cascade of beliefs.

I start with an explanation of Searle's social ontology. I then present the two significant objections to his account. I then set out my account of the ontology of money and show how it overcomes these objections.

2. Searle's theory

Searle (1995, 2010) sets out to solve the puzzle of how social facts that are constructed in our minds fit into a physicalist world view. He provides an answer to the question "How are institutional facts possible? And what exactly is the structure of such facts?" (Searle 1995: 2).

> [There is a] continuous line that goes from molecules and mountains to screwdrivers, levers and beautiful sunsets, and then legislatures, money, and nation-states. The central span on the bridge from physics to society is collective intentionality, and the decisive movement on that bridge in the creation of social reality is the collective intentional imposition of function on entities that cannot perform these functions without imposition.

This extract positions Searle as a naturalist, eager to ground his theory in the physical world without recourse to any nonphysical metaphysics. Collective intentionality plays a critical role in this account, serving to impose social roles onto physical objects. He provides this formula as a general model for how this works (1995, p28, 43):

X counts as Y in context C

This is the general form of a *constitutive rule*. The X term is a "brute fact", borrowing Anscombe's (1957) term, a physical thing out there in the world. It counts as a Y term, a social fact, in the right context, C. So, for instance, a dollar bill (X) counts as money (Y) in the United States (C).

Constitutive rules enable us to assign *status functions* to particular brute facts. The assignment is, for instance, that these particular bits of paper with a certain set of physical features have the status of being money and therefore function as money. But for the assignment to work there needs to be general recognition of it in society which is achieved, according to Searle, via *collective intentionality*. One can imagine a central bank governor declaring that hence forth these bits of paper are money, thereby creating a collective intentionality that sees those bits of paper as money.

Constitutive rules, status functions and collective intentionality are critical notions in Searle's account and I discuss each of these concepts in the next three subsections. His argument amounts to the strong claim that these three notions can explain all of institutional reality (Viskovatoff 2003). They are key to creating *institutional facts* among the larger set of social facts.

2.1 Constitutive rules

Building on the work of Austin (1962), Searle's book *Speech Acts* (1969) set out a detailed theory of how speech acts serve to bring social facts such as promises and agreements into existence. Searle doesn't provide a precise definition of a "social fact", choosing instead to illustrate the term by examples, including "being an American citizen", money, a couple taking a walk in a park, traffic lights and others (1995). Some of these can be understood as "institutional facts" when they concern social institutions like governments, universities and basketball teams. Searle is not clear on the distinction between an institutional fact and a non-institutional social fact, but he appears to mean that institutional facts are those that involve brute facts, like a written contract, whereas other social facts, like "going for a walk together" (1995, p 26) are not. When a social fact depends on physical objects such as documents, then it is an institutional fact. These facts all constitute what he calls "social reality" which is the complex web of social facts that we are immersed in in our daily lives.

Constitutive rules are created through *performative utterances*, a class of speech acts that Searle calls "declarations" which do the job of making a particular fact true such as "I now pronounce you man and wife" or "war is declared" (1975). So, for example, a government would declare a certain set of pieces of paper to count as money and thereby make it true that they are money. While in 1995 he says declarations are key in most but "by no means all" social facts (1995, p 34), in 2010 he excludes the possibility of exceptions in the case of institutional facts: "All institutional facts, and therefore all status functions, are created by speech acts of a type that in 1975 I baptized as ‘declarations’” (Searle 2010: 11).

Searle uses the term “constitutive” to indicate that the rules are the defining features of institutional facts. Chess, for instance, is made through its set of rules, rather than through any particular pieces or board. Without the rules there is no such thing as chess. Similarly, an institution owes its existence to constitutive rules, which serve to assign status functions to physical objects.

X terms are usually physical objects like bank notes and traffic lights. They are imbued with an “agentive function” in that they serve to extend human agency. But other sorts of things can also count as X terms. People are often X terms. For example, a constitutive rule may specify that a person (X term) acts as a marriage officer (Y term) in the right context (a church service where he is duly authorised). X terms can also be written documents or spoken words or sounds. “Institutional facts exist, so to speak, on top of brute physical facts. Often, the brute facts will not be manifested as physical objects but as sounds coming from peoples’ mouths or as marks on paper – or even thoughts in their heads” (Searle 1995: 35).

The reference to “thoughts in their heads” seems surprising, given that Searle’s account depends on physical objects, his brute facts. Of course, brains are physical objects, so he might be interpreted as saying that thoughts can be reduced to neural activity in our brains. Searle does, in his philosophy of mind, subscribe to a form of mental materialism of this sort which he calls biological naturalism (Searle 2007). But he is content to argue at the “higher level” of mental concepts and leave the biological naturalism implicit, and it is at this level that his point on “thoughts in their heads” should be interpreted. He is allowing room for problematic cases like blindfold chess which has no board or pieces. There are mental representations of the X terms in the players’ heads that substitute for the physical items that are chess boards and pieces. While there are physical actions involved, being the two players uttering moves to each other, there are no objects that are assigned status functions, other than the objects that the players are imagining. As I will discuss below, this example remains problematic for Searle, even if we grant him that certain thoughts, being those that are imaginings of physical objects, are satisfactory substitutes for physical objects for the sake of his ontology. I will argue in section 3 that if we develop a game that depends on thought but has no physical pieces or board at all, we will have a good counterexample to Searle's ontology, a freestanding Y term, an institutional fact that does not depend on any X term.

The last thing to note about constitutive rules is that Searle distinguishes them from regulative rules, a distinction that will be important in my account of money. As he puts it, "Some rules regulate antecedently-existing activities. For example the rule 'drive on right-hand side of the road' regulates driving; but driving can exist prior to the existence of that rule" (Searle 1995: 27). This is an odd concession by Searle. While he says that driving exists prior to the existence of the regulatory rule, the example also shows how some rules can exist without having been declared by an authority. He contrasts this example of a regulatory rule with constitutive rules such as those for the game of chess. Searle does not give any account of why or how "drive on right-hand side" came to exist in the first place. One answer to this puzzle might be David Lewis's notion of social conventions (Lewis 1969). Lewis uses game theory to argue that societies settle on certain ways of doing things as a result of strategic interactions of individuals. Otherwise arbitrary choices, such as the side of the road to drive on, are made because everyone thinks everyone believes it is the right side to drive on, and that it would be harmful to drive on the other side. The convention can be sparked because it is the more salient to those choosing from otherwise identical choices, such as if a particularly prominent individual makes the first move. But if Lewis's conventions is the right answer to the origin of Searle's "antecdentally existing" activities, why can't something like Lewis's conventions be the answer for all of social reality? My point is not to suggest that Lewis's conventions is the right solution, rather it is that once Searle acknowledges that certain conventions can arise without any speech acts or constitutive rules, he has opened the door for equilibrium accounts of social reality that don't depend on his scheme.

2.2 Status functions

A constitutive rule assigns a certain status function to one or more X terms. For example, certain bits of paper have the status of being money and a priest has the status of being a marriage officer.

Constitutive rules can be iterative, generating successive status functions. For example an adult (X term) can become a marriage officer (Y term). At a ceremony the marriage officer (X) can declare a couple married (Y). A married couple (X) can obtain spousal rights (Y). And so on. At each stage a status function is assigned that thereby creates a Y term which then enables it to act as the X term in the next iteration.

Status functions are assigned by constitutive rules, but these have to be recognised by the rest of society for them to work. This final step in Searle's ontology is achieved by collective intentionality. While the rules assign status functions, Searle sometimes talks of collective intentions assigning status functions directly.

2.3 Collective intentionality

Intentionality is a philosophical concept that refers to important features of human identity and agency (Pierre 2014). It is what allows us to be directed toward specific goals and to have specific purposes for our actions. Hoping, believing, desiring, perceiving, knowing, feeling, and so on, all exhibit our intentionality. Our intentionality explains our actions: opening a door, "because I want to go through it", or turning on my computer "because I want to work on my paper".

The notion of collective intentionality extends this personal concept to groups. Theatre troupes, sports teams, and even a group push-starting a car have collective intentionality in the sense that Searle uses the term. Such groups have a kind of agency in that they have goals and collectively work toward them. In a similar way as we explain the actions of individuals, we can explain the actions of groups by referring to their intentionality. Searle introduces the idea of "we-intentions" as distinct from "I-intentions". "I-intentions" go no further than my own actions. This kind of intentionality seems easy to defend from a standpoint of personal intuition – we know our own intentionality. "We-intentions" are intentions individuals hold about what "we" are doing.

This idea of collective intentionality plays a key role in the creation of social facts. As Searle puts it (1995 p 46),

> …collective intentionality assigns a new status to some phenomenon, where that status has an accompanying function that cannot be performed solely in virtue of the intrinsic physical features of the phenomenon in question. This assignment creates a new fact, an institutional fact, a new fact created by human agreement.

So a group, via its "we-intentions" can collectively assign a status to some object (Searle uses the term "phenomenon" but seems to mean physical objects given the reference to their physical features) which then becomes an institutional fact. So, for example, bank notes are money because the status of being money is assigned to them, through some process of agreement over constitutive rules.

Note that collective intentions remain individual intentions. There is no Platonic collective intention that is divorced from individuals. The "we-intentions" in Searle's account remain personal intentions, they are merely about a group. It is difficult to understand just how Searle imagines this works. If we-intentions are personal, then the fact that other individuals have intentionality seems irrelevant. I simply need to know that other individuals are committed to the objectives of the group. Even if the other individuals are Zombies, acting on some primitive volition but not capable of thought, that seems sufficient for me to coordinate with them in a group such that my own thought could be described as a "we-intention". The role of we-intentions is that they allow us to predict the actions of others in order to be able to act in concert with them.

Collective intentionality can be contrasted to what Searle calls "mutual beliefs" about others' "I-intentions" (1995: 24). I may believe that you intend to play your violin and you believe that I intend to play my clarinet in order for us to perform a duet. While Searle uses the term "mutual beliefs" to describe such a cascade of beliefs, they are not the same beliefs. My belief "you will do X, because you believe that I will do Y" is not the same as your belief "you will do Y". Rather the notion of "mutual" here is meant to indicate that the beliefs mutually reinforce one another through this cascade, but each belief is an "I-intention" plus a belief about your "I-intention". Searle's notion of we-intentions replaces this cascade. Searle argues that an account based on mutual beliefs implies an infinite regress. If I believe you will play your instrument when I play my instrument and you believe that I will play mine for us to play this duet, we appear to enter an infinite regress of beliefs about each other's intentions. Searle describes we-intentions as biologically primitive, so the regress no longer occurs. We simply have joint we-intentions in the same way that wolves hunting in a pack have we-intentions.

I will argue in section 7 below that this infinite regress fear is unwarranted and that mutual beliefs are a sufficient basis for joint action. It follows that Searle's notion of we-intentions is unnecessary and best avoided on the grounds of parsimony. But there are several other difficulties with the concept of we-intentions. Searle wants to avoid any metaphysically difficult notions such as some kind of shared social consciousness or awareness hence his we-intentions are beliefs of individuals. He goes so far as to claim that a brain in a vat is capable of having we-intentions, writing (Searle 1990: 11):

> …I could have all the intentionality I do have even if I am radically mistaken, even if the apparent presence and cooperation of other people is an illusion, even if I am suffering a total hallucination, even if I am a brain in a vat. Collective intentionality in my head can make a purported reference to other members of a collective independently of the question whether or not there actually are such members.

On this telling, a we-intention differs from an I-intention only because it uses the first person plural rather than singular. Searle seems to think this allows him to avoid the problem of mutual beliefs. But it seems absurd to think such we-intentions can be formed without any reference whatsoever to what other members of the group are thinking. On Searle's account they don't need to have *the same* we-intention as I do and yet somehow institutional facts exist because collective intentionality recognises the agentive function of objects. This does not seem to be a consistent position. In contrast, my account allows individuals to speculate about the intentions of others, and it seems absurd to me to claim that people don't make assumptions about others' intentions and that collective action can exist without it.

There is another interpretation of "mutual beliefs" that I flag now, for further discussion later. Rather than beliefs about each other's beliefs, mutual beliefs may simply ones that are "mutualistic", to borrow a term from biology. An analogy from biology will help explain what I mean. Species are said to be involved in mutualistic cooperation when, in the pursuit of their own ends, they end up being helpful to other species, which may be helpful in return. One example is the oxpecker bird which sits on the back of zebras and other savannah grazers picking off ticks and other parasites. The zebra benefits from the control of parasites and the bird gets a healthy meal at a safe vantage point where predators can be detected. But neither animal is acting in the interest of the other, or with any assumption about the intentions of the other. The analogy is this – certain institutions may come into existence from the mutualistic beliefs of individuals. Their beliefs may be entirely independent, in the sense that they need not even be aware of the other's beliefs, and yet the two together serve to make a social institution real.

This idea plays a prominent role in economics, going all the way back to Adam Smith's notion of the invisible hand. It is worth quoting the original context for the term (Smith 1776: 593-94),

> By preferring the support of domestic to that of foreign industry, he intends only his own security; and by directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain, and he is in this, as in many other cases, *led by an invisible hand to promote an end which was no part of his intention*. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it. I have never known much good done by those who affected to trade for the public good.

I have italicised the key clause in this extract. Note Smith's comment that the institutional outcomes promoted by the actions of the individual were "no part of his intention". Smith is usually taken to mean that the institution that is the economy broadly is benefitted by the actions of individuals individually pursuing their own interest. This is a social example of mutualistic beliefs in action; institutions emerge from individuals pursuing their I-intentions.

This completes the basic overview of Searle's account of social facts. Next, I discuss some important objections.

3. Freestanding Y terms

One convincing objection to Searle's account is known as the objection from "freestanding Y terms", which are social facts that don't have any brute fact (X term) associated with them. The objection has been developed by Barry Smith (2003) who argues that freestanding Y terms are particularly common in "the higher reaches of social reality" such as sophisticated economic concepts (Smith 2003: 21):

> …We pool and securitise loans, we depreciate and collateralise and amortise assets, we consolidate and apportion debts, we

> annuitise savings – and these examples, along with...the example of money existing (somehow) in our banks' computers, make it clear that the realm of free-standing Y terms must be of great consequence for any theory of institutional reality.

On this account, a corporation, for example, does not have to have a building or boardroom or officers (people) or even a certificate of incorporation, or any of the other physical things we associate with a corporation for it still to be recognised as a corporation. Ultimately, of course, human beings must exist for the corporation to exist – that is true of any social fact. Corporations still require human beings to recognise that they do exist, but physical objects are not, on Smith's account, *necessary* for the corporation to exist. Such freestanding Y terms are, Smith argues, ubiquitous in social reality.

Searle has at times accepted and then backtracked on whether there really are such things as freestanding Y terms. Some of his earlier examples seem to suggest that his scheme would have to acknowledge their possibility. Consider again his example of blindfold chess. Searle used this example to hedge himself from the objection that certain institutions can happen purely at the level of mental representations. Searle argues that in such cases the X terms are the representations of physical objects in peoples' heads. The difficulty with this example is that it depends on the contingent fact that chess board and pieces exist. That chess is commonly played with a physical board and set doesn't seem to be a *necessary* feature of blindfold chess. We could come up with mental games and agree on rules with no physical instantiation of those games. There may still be utterances of the sort "pawn moves to E4", but it would be a stretch to transfer the agential function ascribed to a physical pawn in a chess set to the physical sounding of the terms. All the work of constituting the game is being done in people's imaginations, not in the sounding of the terms. The point is that such institutional facts appear to be possible without any need for brute facts or X terms.

In response to Smith, Searle appears to acknowledge corporations are freestanding Y terms (Searle 2003: 57, his emphasis):

> What we think of as social objects, such as governments, money and universities, are in fact just *placeholders for patterns of activities*. I hope it is clear that the whole operation of agentive functions and collective intentionality is a matter of ongoing activities and the creation of the possibility of more ongoing activities.

The notion of "placeholders for patterns of activities" appears to mean that he accepts that money, universities and governments are not necessarily physical things. A "placeholder" might be like a mind that is engaged in blindfold chess, although it is hard to see what such a placeholder would be for a university. If Y terms can be ongoing activities, we can dispense with X terms – they are not the necessary properties of the institution. The institutional fact is constituted by the Y terms.

Later, Searle appears to back away from this apparent acknowledgement of freestanding Y terms and recommits to the essential role played by X terms. He does this by arguing that, ultimately, all Y terms depend on particular people who have particular status functions (2010, p. 21-22):

> The freestanding Y terms do not bottom out in concrete objects, but they do bottom out in actual people who have the deontic powers in question. So there is no object or person who is the corporation, but there are the president, the board of directors, the stockholders, and others, and the deontic powers accrue to them. A corporation is just a placeholder for a set of actual power relationships among actual people.

Searle is here denying that Y terms are freestanding – they fundamentally depend on people, which serve as X terms. What Searle seems to be suggesting is that a corporation, Berkshire Hathaway, say, consists of a large number of office holders, with each office imbued with a set of deontic powers. Among those officeholders is the chairmanship of Berkshire, a position that carries a set of deontic powers. There is the person that is Warren Buffett and this person has, by virtue of occupying the office of chairman, accrued those deontic powers. He has, in effect, been assigned a status function, and therefore Searle's X counts as Y in C formulation holds.

His account seems to suggest that the corporation consists entirely of the people associated with it and the Y terms that are assigned to them. So the question, "what is Berkshire Hathaway?", is answered by specifying a set of office holders and the institutional roles they play. This seems to suggest that the corporation really is nothing more than this set and the corporation is not itself a causally distinct real thing. But a satisfactory notion of a corporation must surely allow it to exist as a thing in its own right. For one thing, Warren Buffett may resign from his position and, until the office is filled, there would be no person who is assigned the status function. Yet there is still an office that has certain responsibilities, such as convening an annual general meeting. These responsibilities do not go away while a search for his replacement is ongoing. The office itself therefore could be a freestanding Y term. Moreover, many corporations need not have people at all. The so-called "special purpose vehicle" that is used in financial transactions to hold assets is a corporation that exists only in a legal sense. While nominally there are individuals assigned to act as directors to such companies, they would serve such a function for thousands of such companies. The institutional fact of such companies cannot be bottomed out to these nominal officers whose entire responsibility is to sign off accounts once a year. Such companies exist because of the causal powers the company has, rather than that of the individual. This is Smith's main point – that eventually you have to acknowledge that the defining character of a corporation is that people recognise it as a corporation, not that it has any brute fact associated with it.

The point is that if we accept that institutions have causal powers as institutions, which I think Searle does, then the reduction strategy is closed to him. And once we see the corporation as existing, we can then see shares in its equity as existing, derivatives written on those shares as existing, and so on, getting further and further from any physical object relating to the corporation. Ultimately, it seems clear to me that Searle cannot escape the objection from freestanding Y terms, both at the lower level but, as Barry Smith argues, even less so as one moves further into the realm of complexity found in the financial system. There are many institutional facts that do not, of necessity, have physical objects or people assigned status functions in order for them to exist. This is not to say they aren't causally affected by people, buildings, documents, and so on, but only that the corporation is not identical with any of these physical things.

4. Money as a freestanding Y term

One of the issues that Searle grapples with is that not all money is in the form of physical notes and coins. As a matter of fact, in most economies, only about 3% of the money supply, as recorded in national accounts, is in physical cash with the rest recorded in the records of banks and other institutions. These records themselves might be kept on physical things, like computer disk drives, but it does not follow that these physical things are themselves money. All the physical things are doing is keeping records of financial obligations. If you and I agree that I will borrow $100 from you this fact need not be recorded anywhere to be true. If I then create a spreadsheet to record debts that are owed to me, that doesn't make them true; the spreadsheet is just a record-keeping device, to help me remember the facts that are true. I will argue that this is analogous to money.

I'm not aware of any location where Searle does this, but he may respond that a loan that exists only by agreement between people, and recorded only in their heads rather than through any external physical thing, is not an institutional fact. Such loans may be like his example of a couple taking a walk together, in which there is collective intentionality but no external representation of it and therefore no institution. But this response seems *ad hoc*. One can come up with many examples where individuals accept a set of deontic obligations to one another that constitute an institution. No physical document is *necessary* for such an institution to be said to exist. Indeed, we readily accept that a loan document can be destroyed, but the obligation can still be recognised by both parties and therefore still exist. As long as there are mental representations, the brute fact of a written loan agreement is not a necessary criterion for there to be a debt.

It is interesting to consider whether the converse is true. What if both of us, being the only two people in the world aware of a loan between us, forget about it *and* there was no external representation of it? My view is that the loan would cease to exist. This is why it is important that we do record such institutional facts so as to not leave them contingent on our memories. Should we forget, such representations work to recreate the deontic obligations and therefore the loan, perhaps mediated by other institutions such as the legal system. It may seem strange that debts can come in and out of existence in this way, but it is less strange if we see that an additional background belief is at work concerning the way we represent deontic obligations. We both believe that loan agreements tell us something about what our obligations "should" be. When I am reminded of the debt by the record of it, the fact of it comes back into existence because I accept that I am indebted. This suggests a line of argument that might help us to understand how institutions can collapse, which is that our recognition of the link between such representations and our subjective deontic obligations can collapse. Ultimately we cannot have external objects to mediate in every social fact – eventually we have to have a foundational belief in the representative role that the object itself will play. We have to accept that the loan agreement implies an obligation, even if we've forgotten the obligation itself. The agreement records the loan, but the social fact "loan agreements record obligations that must be respected" must itself be a free standing Y term.

The loan example shows, by analogy, roughly why money is a freestanding Y term too. I will present three arguments to make this more precise. First I will show that Searle's own logical construction of the coming into existence of money in society is weaker than an alternative account which requires no physical thing. Second, I will argue that cash money is already broadly recognised by several institutions as representations of money and not actual money. Third, I will consider examples from history of the destruction of external representations and mechanisms that have been used to reconstitute deontic obligations, which show that the representations are not essential to the facts. This argument is broadly theoretical but I draw on some real world examples to enrich the theoretical accounts.

4.1 Searle's account of the existence of money

In Searle's 1995 discussion of money he sets out a story of how money comes into existence, based on three sequential phases:

1: *Commodity money*, like gold and silver coins;

2: *Contract money*, like bills and other promissory notes that provide the holder the right to demand commodity money;

3: *Fiat money*, which are bills that have no commodity money backing them

Searle describes this evolution thus (Searle 1995: 42):

> The use of commodity money, such as gold and silver, is, in effect, a form of barter, because the form that the money takes is regarded as itself valuable. Thus the substance in question performs the function of money solely because of its physical nature, which will typically already have some function imposed on it. Thus, gold coins are valuable not because they are coins but because they are made of gold, and the value attached to the coin is exactly equal to the value attached to the gold in it. We impose the function of "value" on the substance gold because we desire to possess that kind of substance.

So gold and silver are valuable and therefore become useful as a medium of exchange. The next stage in the evolution of money was when certificates were issued as "a kind of substitute for gold. It had complete credibility as an object of value, because at any point, it was exchangeable for gold" (p 42). Next comes fiat money, which is money that has no commodity backing it.

This account of the development of money is not, I think, meant to be a claim about actual history. Rather it is a logical reconstruction to show how the institutional fact of money is logically connected to an underlying physical thing. This account, however, depends on a premise that, at the initial stage, gold coins were valuable in and of themselves and that this value was seamlessly borrowed in the status function that was given to money. In Searle's account the X term is *independent* of the Y term (clear in the extract quoted above: "gold coins are valuable not because they are coins but because they are made of gold"). The X term has intrinsic value that is not about any individual's beliefs about it. Its value precedes its status as money. I will argue in section 6.1, below, that the value of money is better seen as extrinsic.

Searle's first step is clearly inspired by certain historical accounts. Von Mises argued in his 1912 Theory of Money and Credit that gold and silver had value by virtue of the "natural qualities" of the two metals that allowed them to be used in jewellery (Mises 1934 [1912]: 33):

> For hundreds, even thousands, of years the choice of mankind [for a medium of exchange] has wavered undecided between gold and silver. The chief cause of this remarkable phenomenon is to be found in the natural qualities of the two metals. Being physically and chemically very similar, they are almost equally serviceable for the satisfaction of human wants. For the manufacture of ornament and jewellery of all kinds, the one has proved as good as the other.

Note that Von Mises provides an independent reason for the value of gold and silver, namely their use value, such as in the manufacture of jewellery. So, the commodity that is used as money has an intrinsic value. This is consistent with Searle's idea that this is borrowed by fiat money such that it works essentially as a form of barter. Von Mises and Searle's views fit with the barter theory of the origin of money, classically argued by Carl Menger in an 1892 article *On the Origin of Money*. Menger argued that money arose to solve the problem of a double coincidence of wants that normal barter exchanges seem to require, an account that remains standard in many economic text books today (Dodd 2016).

There are two counterarguments to this conception of money. Searle and Von Mises imply that the value of money depends on this first step of having independently valuable things that become money. But this account ignores the fact that the ability to use something as a medium of exchange provides an independent value to it. We could argue that a quarter dollar coin has a "use value", perhaps because it is useful as a decision device when I want a random outcome by flipping it. Of course, quarters aren't valuable because they are useful randomisation devices. Similarly, the fact that gold can be used for jewellery, and in many industrial applications, is not all there is to the value of gold and silver. The value of quarters, and gold and silver (though less so today) is because they provide a representation of money. That it may *also* have use value of this sort may have consequences for its use as a representation of money. For one thing, if it stops being a good representation of money, perhaps because no one else will accept it for exchange, its use value remains. The point is only that its function as a representation of money leads to it having a value that is *different* to the use value. This implies that there is an *extrinsic* value to money, a value that is not about the commodity used as money but rather arising from the fact that it is seen and used as money. I shall return to this point in section 6 below when I consider how money becomes valuable.

Some other speculations on how money came into existence are consistent with my account. The anthropologist Keith Hart argues that the traditional story that money arose as a solution to the double coincidence of wants, the idea that every buyer has to want what the seller is offering and vice-versa in order to barter, is "an origin myth" repeated by authors from Adam Smith to Karl Marx (2001: 269). He argues that the classical story requires institutions like private property to exist and markets in which strangers would encounter each other to barter, all of which are likely to only have emerged far later in economic history. Until then, exchanges of goods and services would be within kin groups in which we can imagine deontic obligations of the sort we have discussed above emerging as individuals gave each other goods and services. As he puts it (Hart 2001: 5),

> I argue that money's chief function is as a means of remembering. Indeed the origins of the institution in Europe drew a firm association between money and collective memory. And today money remains one of the main infrastructures of communication and shared memory, with its power now amplified by the widespread use of machines.

On Hart's account, deontic obligations arise between members of a kin group. As the complexity increases it become necessary to record these obligations. Money serves as a convention to express these deontic obligations. When we take the next step of recording these in some sort of proto-ledger, perhaps by designating a cowry shell to serve as a representation of some fixed value, and attributing that value to whoever has possession of that cowry shell, we are recording money via a representation rather than creating it.

I make no claims about the historic accuracy of the classical origin myth or Hart's alternative. But as a rational reconstruction of how money comes to play its role in society, Hart's account is at least as plausible as Searle's, and, I think, more so. If we see money as a social convention through which we represent value, that representation quite obviously need not be a physical object. If two people simply remember their debt in terms of some commonly understood representation, such as hours of labour or some other claim on the other person, no physical object need exist, nor do we even have to imagine any particular physical objects. We may decide to designate certain cowry shells as representing some quantum of obligation, but then they are representations that serve as mnemonic aids rather than intrinsically valuable.

Searle's later account of his theory acknowledges that representations are important, when money is not in cash form. "There need be no physical realisation of the money in the form of currency or specie; all that exists *physically* is the magnetic traces on the computer disk. But these traces are *representations* of money, not money" (2010: 20, original emphasis). This is a more substantial acknowledgement than Searle seems to realise.

Why are computer records representations while bank notes are the real thing? Searle doesn't provide any answer. In fact, the distinction is arbitrary. Bank notes represent my ability to obtain other goods and services in the same way that the balance on my online banking does. Both record my control over a certain representation of value.

Smit, Buekens and Du Plessis (2011) argue that Searle's mistake is that he depends on a "folk theory" of fiat money as being money "that is backed by nothing". This folk theory has it that fiat money works because governments have declared it as being a medium of exchange. If such government declarations were all there was to it, "fiat money" would indeed be a strange phenomenon, though completely consistent with Searle's speech act ontology. But in fact governments and their central banks actively work to maintain the value of their currencies, using the tools of monetary policy such as interest rates and banking regulation to do so. They support this with regulatory rules that require merchants to accept money as a medium of exchange and the physical integrity of cash to be respected. In these actions regulators and governments help to protect the value of money and maintain public confidence in it as a common reference point. But while these regulatory rules can reinforce a fiat money regime, they are *not* what constitute the regime. Indeed, the tools of monetary policy wouldn't be necessary if the rules constituted the regime. But central banks, consistent with my account, know that it is crucial that the public are able to use money to represent their subjective sense of value. Inflation targeting, interest rates, open market operations and bank cash reserve ratios are all mechanisms that affect this subjective sense.

Neither bank notes nor computer records are essential for money to exist. Many central banks will replace bank notes if they are destroyed. The Bank of England, for instance, says it will consider various claims for bank notes that have been destroyed, provided it can satisfy itself that it will not be paying out twice on the same banknote[4]. The Bank's behaviour is consistent with my claim that the note is a representation, not the thing itself, and when it is destroyed, it is only the representation that is destroyed. This attitude of the Bank makes sense when we see its role as managing the value of money, rather than managing particular representations of it.

A thought experiment provides further support for my argument. Imagine that by some remarkable accident, perhaps a solar flare with damaging electromagnetic consequences, the computer records of an entire banking system of an economy were deleted. What would we do? Would the obligations that the bank had to its customers, and its customers to it, suddenly cease to exist? The answer would be no. We'd simply have lost the record, or the external representations, of those obligations. What would remain would be the beliefs and memories we have about the obligations that we have to one another and to institutions like banks. These would be the representations of deontic powers that we think we hold.

We would certainly have a difficult task of reconstructing the external representations that make up the bank, but that is what the right response would seem to be. Some form of judicial process would have to consider everyone's claims to

[4] It helpfully adds, "The list of ways in which banknotes become damaged is almost endless – from those accidentally put through a washing machine to those chewed by the family pet. Banknotes hidden for safe keeping can often be overlooked. Those concealed in places such as ovens or microwaves run the risk of burning whilst banknotes hidden under floorboards or in gardens become damp and eventually decay" (Bank of England 2014). Of course the challenge is to prove that it won't be replacing a bank note twice, so bringing along the machine-washed, chewed, or microwaved remnants of the note may well be necessary to get a new one.

determine the real state of affairs. In this process the external representations of the deontic obligations between individuals and banks would be reconstructed. Money is the convention we use to quantify these deontic obligations.

While this is a thought experiment, it is not too far from reality. Something like it has occurred with respect to property records destroyed in revolutions. A property deed is a representation of a property right, it is not the right itself. If a property registry is destroyed a process of restoring it can follow in which representations are reassembled. Destruction and restoration has taken place in Iraq (Van der Auweraert 2007) and Rwanda, Cambodia, Kosovo, East Timor and Palestine (Manirakiza 2014). In such cases the records of ownership held in registries was destroyed and then reassembled through a process of restoration.

So, because bank accounts and bank notes are both representations, and because if such representations are destroyed we will replace them or recreate them, it follows that money *must* be a freestanding Y term. It exists as something quite apart from any of these representations. It is a concept we use to express our subjective senses of value and exists at an intangible level. In section 6 I will explicate the relationship between value and money more fully. First I will present a second objection to Searle's account, which is that it cannot explain institutional collapse.

5. Social facts and the collapse of institutions

Searle's social ontology does not seem to provide any way that institutional facts can come into existence unintentionally, but, also, no way that facts can cease to exist unintentionally. The "X counts as Y in C" formulation requires collective intentionality and the assignment of agentive functions. For such institutional facts to cease to exist, it must be that a constitutive rule needs to be undone with another rule like, "X no longer counts as Y in C". In the latter form, a collective intention has to reassign agentive functions, or remove them, from the X term for it to no longer count as a Y term. Searle's declarative speech acts would be essential for this to happen. But I will argue that not only do institutions collapse in the absence of such speech acts, they moreover collapse in the face of speech acts intended to affirm the ontological status of the institution.

Several examples show that even when the full might of social institutions are yoked to the job of performing desired institutional outcomes, they may fail. For example, during the 2007/2008 financial crisis, markets for certain types of derivatives such as collateralised debt obligations stopped working not because the government or any other agent declared they should, but due to a general loss of confidence in the markets. In some cases, governments were doing everything they could to keep the institutions working. It is hard to see any speech act declaring the failure. It is just as hard to see anything we might call collective intentionality regarding some form of intention to collapse such institutions. A far more plausible explanation is to use my concept of mutualistic beliefs, a series of individual intentions by relevant agents involved in the markets that, when bundled together, caused the collapse of such institutions, as if guided by a malevolent invisible hand. For example, individuals may be intending to protect the value of assets and believe that the only way to do so is to sell immediately at any price. A critical mass of such intentions may come together in a mutualistic way to cause the collapse of the market.

Another example further illustrates the point. At around the same time as the financial crisis, Zimbabwe had to abandon its currency after it was ravaged by hyperinflation, reaching an inflation rate of 79.6 billion percent per month. After the 2009 elections, newly installed finance minister Tendai Biti acknowledged that "since February this year, the Zimbabwe dollar is no longer a currency that the public and any trader will accept. Our national currency has, thus, become moribund" (quoted by Noko 2011: 348-49). The Zimbabwean dollar simply stopped being recognised as money. This was despite the currency remaining legal tender and the refusal to accept it being punishable as a crime. Despite declarations by the authorities, the Zimbabwean dollar became worthless because people lost confidence that it represented their subjective assessment of value.

How might Searle respond to such examples? Perhaps one strategy is the response he gave to a putative counterexample from Amie Thomasson, namely that of an economic recession. We are seldom aware of a recession existing at the time it comes into existence, so no declaration can have taken place, yet it is a social fact (Thomasson 2003). Searle's response is that recessions are not "institutional" facts because there is no collective intentionality at work, even though they are features of economies consisting of many institutional facts. Searle describes these as "intentionality-independent facts" about "intentionality-relative phenomena" and as "fallouts" from institutional facts (Searle 2010: 117). Something is a fallout, on his account, if and only if the truth of the fact that is the fallout is unrelated to whether people believe it is true or not. It is in this sense intentionality-independent. However, because a recession depends on individuals in a society all acting in a certain way, it is in this sense "intentionality-relative". But crucially, "They carry no additional deontology, so no new power relations are created by fallouts" (ibid.).

Might Searle's response to Thomasson also amount to a response to my objection from the collapse of institutions? A recession could be achieved by my concept of mutualistic beliefs whereas Searle requires collective intentionality for something to be a proper institutional fact. Could we describe the collapse of the Zimbabwean dollar as a "fallout"? I contend that Searle's distinction of intentionality-relative and intentionality-independent, cannot survive such examples. The collapse of the Zimbabwean dollar was clearly intentionality-*dependent* because it involved a change in beliefs about the value of the Zimbabwean dollar. It also involved mutual beliefs about the Zimbabwean dollar. My refusal to accept it in exchange for something I value is driven by my belief that others won't accept it in turn. It no longer serves its role as a common way of understanding and quantifying value. While no one has declared the Zimbabwean dollar to be worthless, the collapse of mutually reinforcing beliefs about its value has led to that result. It is therefore intentionality-dependent. The explanation for the collapse of markets in the financial crisis is similar – mutual beliefs about others' faith in the market institution to correctly facilitate and record trades collapsed, rather than any change to collective intentionality.

Neither a recession nor a financial crisis can be described as an "intentionality-independent" fact, although my view is that it is a *collective intentionality* independent fact. If we give up on the notion of collective intentionality in favour of a model of ontology by which institutions emerge from the confluence of individual intentionality, then depressions, institutional failure, and much else become explainable. For Searle a recession is not an institutional fact because it involves no new deontology and no collective intentionality. But the collapse of a currency does imply new deontology because individuals will no longer accept the currency, even though the formal apparatus of Searle's ontology, particularly the status functions assigned by government to the banknotes, still exists. On my account, both phenomena are intentionality-dependent, but not dependent on a notion of collective intentionality.

6. The role of value

I have so far loosely claimed that money represents value. In this section I develop my concept of value and how money is related to it. First, I argue that the value of money is extrinsic to money itself. Second, I argue that money emerges spontaneously in societies as a means of quantifying value, while bank notes and computer records serve to represent this quantified value.

6.1 The value of money is extrinsic

According to Searle, individuals value gold independently from its use as a medium of exchange. In the case of gold, Searle's (and Von Mises's, quoted above) position is that the value of gold has no collective intentionality about it. He seems committed to the view that the value is intrinsic to the gold itself, or at least its value comes from its instrumental use for purposes other than a means of exchange. He also implies that this value has nothing to do with the fact that gold is simultaneously a medium of exchange. On his origin account, fiat money inherits this intrinsic value, that the essence of its original worth in gold is somehow transferred to fiat money. However that might happen, this implies that Searle's view is that money has *intrinsic* value.

This notion of intrinsic value can be contrasted with *extrinsic* value. For objects that have extrinsic value, the source of value is not in the thing itself, but outside factors including the attitudes of individuals to that thing. I use the term "thing" loosely - I have already argued that money may be a freestanding Y term in that it can involve no physical objects. It consists of a concept, a common reference point, that members of a society recognise, which need not have a physical form. A modern form of money is Bitcoin, which form a thing of value without being a physical thing. Bitcoins have no intrinsic value. Their value consists entirely in the attitudes and beliefs of individuals about it and associated institutional infrastructure, so the value is extrinsic to Bitcoins.

My distinction between extrinsic and intrinsic value can be challenged. It may be argued that the instrumental value of gold is in fact extrinsic – it is the choice of humans to use gold in various ways, such as for jewellery, that makes it valuable. Such an argument depends on the premise that the choice is arbitrary, that humans might have settled on, for example, tin, just as easily as gold for adornment, such that the value in that thing consists in the human decision to value it, rather than because of any intrinsic property it has. I don't think many would accept this premise. Searle and Von Mises both accept that gold has properties that make it desirable as for jewellery. In the Von Mises extract quoted above, he notes the "natural qualities of the two metals", being gold and silver, as making them "serviceable for the satisfaction of human wants". There may be some evolved neurological basis for why humans value gold, such as that its colour and sheen evoke the sun which we naturally find appealing (Schoenberger 2011), but I don't think that we need such speculative reasons to agree that the properties of gold contribute to its role as jewellery such that it is not arbitrary. It is therefore these intrinsic properties that are critical to there being a desire for it. This is quite different to Bitcoin where it is the beliefs of individuals that are critical. But even if we accept that gold has value because of these intrinsic properties, it does not follow that it has value only because of these. When gold serves as a medium of exchange and store of value it is valuable for reasons other than, or in addition to, its intrinsic properties, namely, that it can be used for exchange and to store value. These other reasons are the same that make Bitcoin valuable.

When money is represented using a physical thing, such as coinage, it is possible that there are both intrinsic and extrinsic properties that are relevant to its function. For example, coins are small and storable, which is a useful property that make them more valuable as a representation of money than, say, bushels of corn which decay over time and require significant space. This is an intrinsic feature of coins which is relevant to its functionality as a representation of money, but it is far from the only factor. The point is that certain representations of money may have intrinsic properties that play some role in its value, but it is the extrinsic properties, particularly the beliefs individuals have about it, that are essential to its function as a representation of money.

6.2 Quantifying and expressing value with money

One historic account along these lines is that of Twentieth Century social philosopher Georg Simmel, who argued that the relationship between value and money depends on complex social networks (Simmel 2004 [1900]). Economic value, on Simmel's account, stems from society as a whole, it is "intrinsic" to society rather than to any particular object. The starting point for Simmel's view is the Kantian notion that concepts of phenomena in the world are synthesised in our minds as a combination of sense experience with the inherent categories of understanding. Inspired by this idea, Simmel argues that value in the economic world is a synthesis of our subjective desires for objects and our encounter with the rest of society through exchange. Exchange always involves a loss of something we give up in order to gain the other thing. The exchange conveys others' willingness to sacrifice, providing information that is synthesised with our own desires to create an intersubjective value of a commodity that can appear to be an objective price. As Dodd describes Simmel's position (Dodd 2016: 28):

> Although both valuation…and exchange are intersubjective processes, they operate in a such a way that value comes to appear as an objective property of things themselves. When two or more objects of desire are compared, the immediacy of our desire translates beyond the dualism of subject and object into something quantifiable.

Making "something quantifiable" is done with money. Money allows us to measure objects against each other, creating a relative ranking of value and it allows us to compare our own desire for an object to someone else's. Money plays the role of a generic idea of value, so it can serve as a common referent for individuals when expressing their subjective view of the value of any particular object. The resulting prices that become social indications of value "appears" to be objective, as Dodd puts it, but is an intersubjective fact that flows from individuals' desire for the object, and beliefs about others' desires, based on their exchanges.

In Simmel's account, money is the quantifying mechanism, a way of taking a feeling of desire and converting it into a form that is capable of communication. The communicative aspect of money is much discussed in the sub field of information economics (Hayek 1945; Grossman and Stiglitz 1976) where price signals serve to convey information about scarcity, quality and desirability. This is consistent with the notion of money as a communicative device, though it is prices, rather than money per se, that does the communicating. In Simmel's account, prices are the outcome of a social process. Money serves to facilitate the "conversation" that members of a society have in expressing their subjective desires, leading to an intersubjective price. In this way, money is a kind of language that people use to communicate about value. This idea is similar to that found in German language theory of the Eighteenth century going back to Leibniz (Gray 1999). Leibniz argued that words are second-order signs that represent our mental concepts of things – the words themselves have no meaning save as signifiers of concepts where the meaning is invested. In the same way, bank notes are second order signs of money. The banknote is not itself meaningful, save as a symbol of the concept of money, or in Simmel's language, money is the "idea" of value. This is surprisingly modern conception of money that is consistent with my view of money as emerging from a society, much like a language does, and can accommodate both physical currencies and Bitcoin. This accommodation is possible because it is the concept that acts as money, rather than any particular symbol of it. This is completely at odds with Searle's view that money is banknotes and computer records are representations of bank notes. Both banknotes and computer records are representations of the "idea" or concept of value, as quantified and expressed using money.

There are two responses Searle might mount to the Simmelian view of money as socially determined. One I will consider in the next section is his argument that mutual beliefs cannot be a foundation of institutions because they imply an infinite regress. But he might also insist that the constitutive rules that underpin money are achieved by speech acts and therefore are consistent with his ontology. Such a response would fail for the following reasons. The Simmel account focuses on individual subjectivity as key to value, while money serves the purpose of quantifying this value. The desires of individuals for something are synthesised with their experiences stemming from exchange in order to determine the value of money. There is no collective intentionality at work; it is the result of the mutualistic beliefs of individuals, ones that interact to form a social result, being a value expressed in money. But, money becomes not just an expression of value, but a thing of value in its own right. That is because the Simmelian mechanism applies equally to money itself, and it is in this regard that the Searlean response is blocked. Money is simultaneously a means of quantifying value, and a store of value itself. The value of money is socially determined by individuals' desire for it and their experience of exchange using it.

This commodity aspect of money, in which it is valuable itself rather than just a concept of value, implies that money has an additional function as a store of value. The storage function of money is critical to the financial system, which serves to mediate and manage this storage function. Shares, derivatives, debt and other financial instruments consist of rules to manage the distribution of quantities of money to different parties under different scenarios. This storage function has no big ontological consequences for the nature of money, other than that money's role as a concept of value can be self-referential.

Money is a way of quantifying and expressing the value of things, but the commodity role of money means we must value it too. Our subjective desire for money is synthesised with our experience of exchange in order to derive a value of money. This is subject to updating as we experience exchange over time. We may be unaware that a currency has fallen in value due to inflation but become aware through exchanges. Indeed, what economists call the "money illusion" reflects delays in belief updating about the value of currencies. Irving Fisher described the money illusion as "the failure to perceive that the dollar, or any other unit of currency, expands or shrinks in value" (Fisher 1928: 4). Governments and businesses are able to buy employee gratitude for a time by granting increases in salaries even when they are below inflation because employees temporarily believe they are really richer. But over time we update our beliefs when we realise that what we can get for our money has become less. When Zimbabwean inflation reached trillions of percent, such that prices were increasing by the minute, we may simply abandon the currency completely in favour of using some other representation of buying power (in

Zimbabwe's case, American dollars or even share certificates of listed companies). If one representation of money loses our confidence, we are likely to seamlessly adopt others, reverting to quotes in terms of kilograms of sugar or any other representation of value. Simmel, Hayek and Leibniz all accept that multiple currencies can be in circulation at any one time, and indeed many different symbols can serve the function of representing money

6.3 The creation of money

Various authors argue that the choice of money in a society emerges spontaneously, rather than through the speech acts Searle requires. One group of writers including David Glasner and Friedrich Hayek have been described as belonging to the "free banking" school (Horwitz 1994). Glasner, for instance, argues that money "did not originate in a deliberate decision taken at a particular moment by a single individual or by an entire community. It emerged as the unintended consequence of a multitude of individual decisions" (Glasner 1989: 6) while Hayek forcefully reject accounts such as Searle's (Hayek 1976):

> …the superstition that it is necessary for our government…to declare what is to be money, as if it had created the money which could not exist without it, probably originated in the naïve belief that such a tool as money must have been 'invented' and given to us by some original inventor. This belief has been wholly displaced by our understanding of the spontaneous generation of such undesigned institutions by a process of social evolution of which money has since become the prime paradigm.

In this extract, Hayek is arguing that money emerges spontaneously in society via "social evolution". This naturally fits Simmel's account, such that we can see money's emergence as a result of the synthesis of subjective value with socially mediated exchanges. While the "free banking" arguments of Hayek were focused on justifying the private issuance of currency and minimising the role of the state, the ontological claims are consistent with mine. We need not follow Hayek in his claim that governments should not be involved in the creation of money in order to accept his ontology of money. Indeed, I have already outlined the regulative role that governments can play to support the robustness of monetary systems – but the important point is that these are not ontologically essential to the existence of money. All that is required is a society, in which a concept for the intersubjective determination of value must exist. That concept is money.

7. Mutual beliefs and the infinite regress

This account of intersubjective beliefs about money is vulnerable to Searle's objection that it implies an infinite regress. Searle expresses his objection as follows (Searle 1995: 24):

> What is the relation between singular and collective intentionality, between, for example, the facts described by 'I intend' and 'We intend'? Most efforts I have seen to answer this question try to reduce 'We-intentionality' to 'I-intentionality' plus something else, usually mutual beliefs. The idea is that if we intend to do something together, then that consists in the fact that I intend to do it in the belief that you also intend to do it; and you intend to do it in the belief that I also intend to do it. And each believes that the other has these beliefs, and has these beliefs about these beliefs, and these beliefs about these beliefs about these beliefs…etc., in a potentially infinite hierarchy of beliefs… in my view all these efforts to reduce collective intentionality to individual intentionality fail.

Arguments from an infinite regress are a common logical device. The general form is to take one moment and show the same considerations apply at a previous moment, which in turn implies a previous moment, *ad infinitum*. For this to work there needs to be a regress formula and a triggering statement (Gratton 1996). The regress formula specifies how a particular statement entails an infinite regress. In Searle's case, the regress formula is something like "every belief I have about your intentions entails a belief you have about my intentions". The triggering statement is the catalyst to spark the regress. In Searle's case, that is a statement like "I believe that you intend…”. There is a further implicit premise to Searle’s account which is that every intention requires a belief about your intention, and vice versa. Collectively, this implies a *reductio ad absurdum*, leading to the conclusion that mutual beliefs of this sort are impossible.

When set out this way it becomes clear that the fault with Searle’s argument lies at the regress formula and the implicit premise. A belief I have about your intentions does not *entail* that you have a belief about my intentions. Further, I can form intentions about joint actions that do not require any belief about your intentions at all. As my earlier Zombie example illustrates, I can even assume another individual has no intentionality whatsoever. All that matters is that I can predict their behaviour. Of course, we often fail to predict others’ behaviours, precisely because we don’t have a complete understanding of their beliefs and intentions. This fact doesn’t stop us from taking the “intentional stance” (Dennett 1989) by trying to imagine what they are thinking in order to predict their behaviour. A complete, infinitely regressive cascade of mutual beliefs is not necessary for me to have a go at imagining what they may be thinking I’m thinking, even if I will get it wrong at times. So the regress formula doesn’t work because my beliefs about your intentions are open to error: there is no entailment of a regress. Further, I don’t need to know your intentions – what I need to do is predict your behaviour, predictions which may be available to me from other sources, such as a social science. Hence the implicit premise is wrong.

Fitzpatrick (2003) argues that Searle’s objection is like taking Zeno’s paradox too seriously. The paradox is that of a runner striving for the finish line. As he moves he can only run half the distance ahead of him (the regress formula). In running each successive half, he never reaches the finish line. This paradox is an interesting test of our intuitions about fractions but it is a useless guide to whether runners will actually reach the finish line.

For my account of money as a concept to quantify and express value to succeed, all that is needed is for individuals to believe that their chosen concept is widely recognised, and that particular representations of that concept, such as dollars, are widely recognised. This does not have to be a full regress of shared beliefs. In fact, as the money illusion makes clear, my knowledge of others' beliefs is easily mistaken and slow to update. My beliefs are subjective and can be imprecise, and vary as I obtain new information about others' behaviour.

My account provides for a process of individual belief updating, which Searle's account does not. Searle needs a declaration for his collective intentionality to change so that an institutional fact can change. A much better explanation is that beliefs held by individuals about particular concepts of money have changed as they experience exchange and synthesise those experiences with their subjective views of value. The intersubjective concept of money is therefore dynamic. All of this can happen in the absence of critical elements of Searle's ontology such as the assignment of an agentive function via collective intentionality. As I have shown above, the response that this is an "intentionality-independent" "fallout" does not succeed.

8. Conclusion

Searle's rules-based account fails because it cannot account for freestanding Y terms and because it cannot account for institutional collapse. I have argued that one of Searle's central examples, money, is in fact a freestanding Y term. I have shown that the institution can emerge spontaneously and results from individuals' need to quantify and express value and exchange it with others. Money exists at the conceptual level and can be used as a store of value that can be allocated to individuals. Bank notes are merely one form of record of such allocations, alongside ledgers, certificates of deposit, electronic records, and a host of other forms of representations of money. This system of quantification, representation and allocation of value is the function of the financial system. Individuals' sense of the value of money is dynamic and subject to updating as they experience others' willingness to exchange goods and services for money. There can be many different concepts of money at play in a society at any one time, and some can collapse as an accepted means of quantifying value as individuals’ exchange behaviour changes.

In the next chapter, I return to the science of studying the financial system and consider in detail one aspect of it: how the act of theorising about it and studying it has an impact on the ontology of it. Understanding the causes and mechanisms of the financial system is clearly a desirable application of social science, and theoretical finance is a research programme that is engaged with the task. In Chapter 5 I take this further to consider how our efforts to design the system are informed by the science of it. What should be clear from this chapter is that the financial system is not something that can be created by declarations and rules. It is something that exists in an equilibrium that only holds conditionally. Our knowledge of the system can help us to shift this equilibrium toward socially desirable goals, or it can accidentally upset the equilibrium. The remainder of this thesis draws out the implications of this relationship between finance theory and the financial system.

Chapter 4
Financial Theory and Performativity

1. Introduction

In this chapter I shift focus partly back to methodology and epistemology in finance, but with a view to understanding how the activities we generally think about as scientific in nature, can also affect social ontology. I have claimed that financial theory is not only about understanding the financial system, but also about changing it. In this chapter and the next chapter I interrogate this causal process and how we can make sense of it.

I argue that there are two general categories of causal impact that theoretical finance can have: when theories are used as the basis for social designs and engineering, and when theories affect individual behaviour and decision-making. The first effect I call "design performativity" and the second I call "theory performativity". I develop these as categories of the general concept of performativity, which is the view that economics, broadly construed, has some effect, of any kind, on economies.

Much of this chapter consists of analysis of performativity. I consider some influential accounts of performativity, particularly those by the sociologists Michel Callon (1998, 2007) and Donald MacKenzie (2008, 2007), which serves as my foil. I argue that these accounts of performativity fail to distinguish between possible interpretations of the concept. I argue that two interpretations are clearly unhelpful, the first I call "strong performativity" and the second "weak performativity". I will provide more technical definitions below, but, roughly, strong performativity is the idea that economics has extremely powerful causal authority over the economy.

Weak performativity is the almost trivial notion that economics has *some* causal impact on societies, whether intentional or unintentional.

I then develop design performativity and theory performativity as more useful notions of performativity. Design performativity is the idea that economists design and engineer certain social institutions. Theory performativity is the idea that theories influence societies, for example by providing normative decision rules and other insights that influence how people behave. I argue that these do provide useful ways of understanding the impact economics has on social ontology. However, I show that even on a generous interpretation of Callon's account, his claims about the impact of theories collapse to a claim of strong performativity. I provide an alternative account of theory performativity which I consider to be a more accurate interpretation of the effect of economic theory on economies.

I dismiss one of the main claims of the performativity theorists, that economics is not a positive science. I argue that there is no way to avoid the conclusion that economics is positively responsive to the features of social institutions, and that the resistance to seeing economics as, at least in part, a positive social science, cannot be sustained. I show that we can recognise that economics is in part a positive social science while still accepting that theories affect the agents within social systems. I show, however, that these effects work through quite specific mechanisms which performativity theorists do not distinguish.

The performativity thesis has been developed primarily by sociologists and has evolved into a subfield of the discipline known as the sociology of scientific knowledge (Ashmore 1989). Apart from Callon and MacKenzie, prominent exponents include Barry Barnes and Bruno Latour. This literature is heavily influenced by philosophy, from JL Austin, who first coined the term "performative" (MacKenzie, Muniesa, and Siu 2007) to the philosophers of science connected to the sociological turn, including Thomas Kuhn (1962) and Paul Feyerabend (1993). Some philosophers have weighed in on the body of work that has been developed by sociologists, picking up on the implications for debates in methodology of economics, particularly Francesco Guala (2007, 1999).Within the sociology literature two approaches have developed. One associated with the Edinburgh school, featuring Barnes and MacKenzie as key figures, has come to be known as the field of Science and Technology Studies. Another, often seen as a sub-discipline, is centred on Paris, with Latour and Callon as key figures, and has developed an approach called Actor Network Theory. I will not go into detail on these particular debates within the sociology literature. My interest is in analysing the concept of performativity that is found in both, in order to draw out useful implications for my account of the methodology of theoretical finance. In the next chapter I will develop the distinction between design and theory in economics, and apply this to theoretical finance in some detail.

2. Performativity

The thesis of "performativity" in the context relevant here "consists in maintaining that economics, in the broad sense of the term, performs, shapes and formats the economy, rather than observing how it functions" (Callon 1998: 2). On this account, economics is not a positive science but is instead a constructivist process that generates the workings of economies and economic agents within them. Economics is *An Engine, Not a Camera* says the title of Donald MacKenzie's study of financial markets (2008).

To make sense of these claims, I will distinguish between four interpretations of performativity:

1. strong;
2. weak;
3. design; and
4. theory.

The first two serve to mark out two extreme positions in the range of possible interpretations of the concept. The latter two are more plausible accounts of performativity as a real phenomenon that results from the work of economists, and financial theorists in particular, and their effect on the world. In the case of theory performativity I provide two interpretations – one that I think provides a more formal interpretation of the concept as it appears in the sociology literature, and one that I argue is more plausible.

3. Strong Performativity

Strong performativity is something of a caricature of the performativity thesis but one that its key authors do not reject, and in some ways, actually endorse. It takes quite literally the claim that economics "performs, shapes and formats" the economy. It is, as Callon puts it, "tantamount to claiming that physics and physicists are able to influence the laws governing the course of the planets" (2007: 313). Indeed, it is even stronger than this, the claim that economists *determine* the laws of economics rather than merely influencing them. Even though this seems to be something of a straw man of what performativity theorists actually claim, by staking out this position we can make some progress toward a more analytically coherent notion of performativity. Strong performativity is consistent with the Searlean account of social ontology discussed in the previous chapter, by which institutions are created through declarations.

To express this version more formally, I will use a general model of social institutions that distinguishes between historic causes and simultaneous causes. There are two types of causes, diachronic and synchronic. The diachronic causes are those that have evolved over time, the history that led to the conditions that made an institution possible. The synchronic causes are the components of an institution that serve to cause the properties of that institution at any moment. So for example, the diachronic causes of an orchestra are the decisions of individuals to form the orchestra, the historic conventions for orchestras that were used, the development of the instruments, the creation of music to play, and so on. But, when an orchestra performs Beethoven's Fifth, it is the synchronic causes, the components consisting of the musicians, conductor, vibrating bows and so on, that cause the orchestra to exist at that moment. The whole that emerges from those parts, together with its set of properties, has both these diachronic and synchronic causes. The components must be in the right organising structure in order for this whole to exist, so the organising structure is a property of the institution (Lawson 1996).

We can put this somewhat more formally. A synchronically emergent whole, W, with a set of properties E, emerges from components $a_1, \dots, a_n$, each of which have their own set of properties $P_1, \dots, P_i$. One of the properties in E is the organising structure, O, that applies to the components. W emerges diachronically from a set of causal conditions $b_1, \dots, b_m$. The concept of emergence has been subject to a great deal of controversy in both physical and social sciences (Bedau and Humphreys 2008). In my account, synchronic emergence is the idea that E is not a member of $P_1, \dots, P_i$, i.e. that the properties of the whole are different to the properties of any component. I will use this model to explain some properties of the different notions of performativity in this and the next three sections.

If we take Callon's and MacKenzie's words quoted above literally, i.e. that economics is an "engine" that "formats" the economy, then the claim is that economists have full control over the causal conditions $b_1, \dots, b_m$ that give rise to W. Moreover, economists are fully able to dictate, or at least predict, E, the properties of the emergent whole W. It also implies that economists are able to specify the properties of components, or at least know them fully in advance, such that $P_1, \dots, P_i$ are fully predicted by them such that E is fully determinable. One way to interpret this claim would be to see neoclassical economics, with individuals cast as *homo economicus* engaged in rational utility optimisation, as somehow being imposed on the world, forcing people to become rational utility maximising optimisers.

Strong performativity implies that economists have full control over the diachronic emergence of the organising structure and properties of the components of an institution. To put it baldly, this implies that economists are somehow omnipotent and all the components in an economy lack any human agency such as the ability to respond dynamically in an unexpected way to the society around them. The mechanisms by which they exert this control are wide. Callon and MacKenzie both begin with declarative Austinian speech acts and extend to all modelling and other theoretical work of economists. All of this is "performative" by which is meant that it shapes the economy.

We can then define strong performativity as follows:

Strong performativity. Economics is strongly performative iff it completely determines the properties of economic institutions and their components.

Notice that this provides a clear distinction between performativity and reflexivity, another concept used to describe the relationship between theories and economies. Reflexivity can be interpreted as a causal relationship from W to $a_1, \ldots, a_n$, in other words, that the whole affects component's behaviour, what we might call downward causation. Reflexivity points to a weakness in my general model because it clouds the difference between diachronic and synchronic causes. If the components are causally responsive to the whole, which is the general claim of reflexivity, the notion of synchronic emergence becomes incoherent. While this objection has merit, it can be safely put aside by seeing synchronic emergence in the moment, while diachronic emergence concerns the dynamics of an institution over time. The whole can be causally bound up in the features of the institution in the next moment, but at any moment the features of the whole result from the components and the organising structure they are in. Formally this means that W in t_2 has a set of causal conditions that includes W in t_1. This technical quirk need not concern us too much, but it is a bigger problem for strong performativity.

Strong performativity proposes a well-defined and complete causal mechanism – from economists to the causal conditions for institutions. The properties of the components in institutions are also fixed. There is no causal mechanism from *W* to the components, which is required for reflexivity. Reflexivity would be a problem for an aspirant omnipotent economist because it would compromise the economist's complete causal autonomy in her job to determine the properties of *W*[5].

The notion that economists "format" the economy or that economics is an "engine" of the economy, has some surprising implications. For one thing, the economy would be endogenous to economics. Theories would be developed and then imposed on the world, with no risk of exogenous factors outside of the theories playing a role. It is this idea that inspires the performativity theorists to dismiss economics' status as an empirical science, claiming that there is no independent reality that economics studies and develops theories about.

This notion of performativity seems like a caricature, but it is one that Callon seems comfortable to accept as a description of his thesis. Writing in a 2007 anthology of articles on performativity, almost a decade after the 1998 book that first outlined his performativity thesis, he reflects on it thus (Callon 2007: 328, parentheses and neologisms are his):

> Caricaturally and generally speaking we could say that the economy does not exist before economics performs it, and that when economic (or economicized) elements are already there it

[5] Note that this is a quite different use of the term "reflexivity" than that found in the sociology of science literature (Ashmore 1989; Woolgar 1988) that briefly became a focus of debate in the late 1980s. That debate considered whether the sociology of scientific knowledge could be applied to itself; whether a sociology of science could become scientific, a means through which science could come to know itself. This is quite different to my sense of reflexivity as a feature of social ontology.

> means that economics (at large) has already been that way. I prefer the risks of overinterpretation of this statement (which can be pushed so far as to become a caricature; some criticise me for saying that economics creates the economy from A to Z!), rather than risking the underinterpretation favoured by the notion of expression (the idea that there are economic practices per se which exist and existed before economics put words to them).

Here Callon makes clear that he would prefer to be seen as advocating for strong performativity than a positive account of economics as an articulation of the world. He embraces this extreme version of his view rather than admit that economists are capable of playing a positive scientific role of observing and theorising about how economies work, of "expressing" them. On this strong account, economies are *tabulae rasae* upon which economists perform, in every respect, the way that they work and what they consist of. He draws his own dichotomy, with economics either omnipotent engineers or inert theorisers and observers, claiming the former interpretation is better than the latter. I doubt he holds that the former claim is *true*, merely that it is better than characterising economists as positive scientists, where "better" is perhaps determined according to political or rhetorical effect. It is difficult to understand why Callon employs this strategy; perhaps he sees it as part of a Hegelian dialectic in which his role is to develop the antithesis to the traditional thesis of economics, in order that we might arrive at some synthesis as an account of the true state of the world.

To take a step toward a more acceptable concept of performativity, which I will argue sits somewhere apart from Callon's dichotomy, I want to start by rejecting this strong performativity. The idea of economists as omnipotent engineers of economic reality is inconsistent with the fact that economists so often fail to predict the course of economies or even to accurately describe mechanisms at work in economies. In case you need convincing, but also because it will be important to the development of my argument regarding weaker forms of performativity to follow, I want to show that this strong concept of performativity is wrong by drawing on two important ideas in the history of economic thought that are directly relevant to the causal interaction of economists with economies: the Lucas Critique and Robert K Merton's self-fulfilling prophecies.

Robert Lucas (1976) famously argued that as soon as policy makers use a specific parameter based on historic data as a policy tool, they change the micro-foundations that had previously generated that data. As Lucas put it (1976: 41):

> Given that the structure of an econometric model consists of optimal decision rules of economic agents, and that optimal decision rules vary systematically with changes in the structure of series relevant to the decision maker, it follows that any change in policy will systematically alter the structure of econometric models.

Lucas's warning was to policy makers who might employ an economic model as a basis for some policy intervention. His point is that the act of a policy intervention, based on the model, changes the world that the model was describing. We cannot consider certain decision rules of economic agents to remain consistent following the implementation of some or other macroeconomic policy intervention. So, for example, if the monetary policy committee cuts interest rates to boost the economy, economic agents may respond by expecting the economy to worsen, given the signal of poor conditions the committee would effectively be making, and so the expected stimulus effect doesn't materialise. When a statistical series is used as a basis for new policy, the structure that generated the statistical series is changed.

The Lucas Critique shows reflexivity and its unpredictable consequences. This problem can be expressed using my formal model as follows. When institutions use a model as the basis for an institution, it synchronically affects the components within the organising structure that the model was describing in the first place. As a result, E of the W is affected, such that the intervention on E is affected by the fact of its measurement. This is a clear example of reflexivity, as I described it above, frustrating economists in their effort to implement policy.

Even when economists try to direct behaviour they often fail. The sociologist Robert K Merton showed this through his self-fulfilling prophecies (1948). His point is, in a sense, the opposite to Lucas's critique – a false belief becoming true of the world, in contrast to a true model becoming false. The examples Merton gives are racism and rumours about the solvency of a bank. His racism example is of how black labourers are willing to work as scab labour during a strike not because of their low moral character, as racists declare, but because they are excluded from unions in the first place. Merton's point is that racist attitudes lead to outcomes that confirm the racist attitude. Similarly, a perfectly solvent bank can be rendered insolvent if a rumour breaks out that it is in fact insolvent (in a liquidity sense). Little can be done to stop the tide of sentiment, including by economists as many bank failures have shown. Merton's point is that societies can't be so easily persuaded by scientists or anyone else. As he put it (Merton 1948: 197):

> [Not] in the social realm, no more than in the psychological realm, do false ideas quietly vanish when confronted with the truth. One does not expect the paranoiac to abandon his hard-won distortions and delusions upon being informed that they are altogether groundless. If psychic ills could be cured merely by the dissemination of truth, the psychiatrists of this country would be suffering from technological unemployment rather than overwork. Nor will a continuing "education campaign" itself destroy racial prejudice and discrimination.

This makes quite clear Merton's view of the efforts of economists to change the false beliefs of the population, such as that a bank is insolvent, are often unsuccessful. But oddly, Merton's self-fulfilling prophecies have been claimed by performativity theorists as evidence of the opposite. MacKenzie (2007: 3) describes Merton's self-fulfilling prophecies "as investigating a version of performativity". Ferraro, Pfeffer and Sutton argue that theories can "become self-fulfilling when institutional designs and organisational arrangements – structure, reward systems, measurement practices, selection processes – reflect the explicit or implicit theories of their designers in the process that transforms 'image into reality'" (Ferraro, Pfeffer, and Sutton 2005: 8,9).

But Merton's point is that economists struggle to "perform" behaviours in society; that they are often ignored. They are not omnipotent economist kings. Rather than examples of performativity, Merton's self-fulfilling prophecies are better thought of as illustrating the failure of performativity. Callon, for one, gets this the wrong way around. He writes (Callon 2007: 324):

> The notion of "prescription" is not very far removed from that of "self-fulfilling prophecy". It is also frequently mobilised to describe the mechanisms through which a conformity between economic theory and economic reality is achieved. Whereas self-fulfilling prophecies imply (similarly) formatted human minds ready to believe in the truth of certain categories or assumptions proposed by economic theories, prescription implies a medium, an intermediate device between theory and behaviour, between economics and the economy. Generally this medium is taken to be institutions and the norms that they impose.

Here Callon gives institutions a strong causal role, able to downwardly influence the components within them, serving as devices to impose norms onto people in line with economic theory. Institutions indeed have causal powers, including sometimes over their components. However, institutions are always vulnerable to difficult-to-control reflexivity through which agents dynamically respond to institutions in ways that cannot be predicted. Callon's interpretation admits of no such reflexivity, whereas the Lucas Critique emphasises the importance of it. Merton's original notion idea of false beliefs being made true, often despite the best efforts of experts to convince the public otherwise, has been morphed by Callon into the notion of "prescription", which he sees as the practice of economists imposing their theories on society. This legerdemain is not convincing. Both the Lucas Critique and Merton's original self-fulfilling prophesies show clearly how the relationship between economic theory and behaviour is not determinative.

The Lucas Critique and Merton's self-fulfilling prophecies provide good counter-arguments to the notion of strong performativity. But this leaves open other relationships between theory and reality. As a next step toward making sense of this causal relationship, let us consider a very weak version of performativity.

4. Weak performativity

A weak interpretation of the performativity notion holds only that there is *a* causal relationship between economics and economic phenomena. This relationship need not be predictable, deliberate or intended, merely that *some* effect exists between economics and the economy. Certainly such a weak notion of performativity would be of limited use. As Donald MacKenzie *et al* say, "to speak at a high level of generality about the 'effects' of economics on economies is a dangerous shortcut" (2007: 6) because it would strip the notion of performativity of any meaningful content. Everything would count as performativity, ranging from economists creating the economy, in the strong sense of the previous section, to the actions of economists that lead to very different outcomes to those intended, as in the Lucas Critique. With such a wide application, it is difficult to see how this weak notion of performativity does any work.

Nevertheless the general idea of causal impact on economies by economists is important to understand. As the argument in this section will show, weak performativity does have some value in that it picks out a subset of causes that can be relevant to the ontology of a social institution, namely those which flow from economics.

We can easily see that economists can form components within an organising structure. Again, given the case of a synchronically emergent whole, W, with properties E, that has emerged from components $a_1, \dots, a_n$ given that a certain O

organising structure applies to the a_is. O is a property of W which emerges diachronically from a set of causal conditions $b_1, \ldots, b_m$.

All we need to make the claim true is to posit that economists are among the members of the set of components $a_1, \ldots, a_n$ and have properties $P_1, \ldots, P_i$ attached to them. Those properties can be many, including the theories and other work they may produce. As components within the organising structure *O*, their effect on the resulting emergent institution *W* depends on the organising structure. Each component causally interacts with the other components, so economists are able to have an effect on others. For example, it may be necessary for a concept such as "the market" to exist so that the people operating in the institution understand how to interpret prices and place bids. Such an understanding can be informed by the economists operating within their midst. These components fit within an organising structure from which the market synchronically emerges with properties (*E*) such as "competitive pricing", contractual agreements, record keeping standards, and so on.

Weak performativity accommodates a perspective according to which economists are not able to determine *E* fully. As components have a reflexive relationship with *W*, with feedback into their behaviour through a dynamic systemic process, *E* can change over time. Economists can influence the components around them, be influenced by them, and be influenced by *E*. All of these are "effects" but they are multi-directional, so economists affect and are affected. It may help to think of examples such as Merton's bank run discussed above. Within the relevant causal factors of a bank run, economists can be active in attempting to convince those around them to trust a bank. But as some historical examples have shown, when regulatory authorities have announced government bailouts of banks, making them substantially safer than before, bank runs have *worsened*, in contrast to authorities' intentions of restoring confidence (Wang 2013). Naturally this causes economists to consider further interventions to restore confidence, showing the dynamic interaction of social reality and social science. It also shows that the effect of economists can be unpredictable and quite different to what they intend.

The notion of weak performativity picks out a part of the components of *W*, that part which reflects the causal impact of economists. The direction of causation is from economists to the rest of the components in *O*. It does not consider the effects *on* economists or other properties of the whole outside of those caused by economists. In the strong version there are no effects on economists because it denies that economics can be a positive science – it has no exogenous content. In this weak version only one direction of causation is recognised – that of economists on economies. It is weak because the effect can be of any sort and intended or not intended, giving rise to what can be called “multiple performativities” (MacKenzie, Muniesa, and Siu 2007). I prefer to call this weak performativity because it is vaguely specified, rather than being specified in several ways.

We can then define weak performativity as follows:

Weak performativity. Economics is weakly performative iff it has an effect on social institutions.

Given that economics is the practice of economists, who are themselves agents in any society, weak performativity is so obvious as to be trivial. The effect referred to in the definition is not constrained in any way. Economists form part of communities, are themselves agents in economies as consumers and play many other roles. In my social ontology, institutions emerge from components, including all relevant individuals, where “relevant” should be understood as having a causal role in terms of the organising structure. To the extent that weak performativity says anything meaningful it is that it picks out a subset of effects in the creation of social institutions caused by economists. But the particular mechanisms through which economists affect economies is unconstrained within this concept of performativity. As already discussed, they may include Lucas Critique-style causes which lead to unexpected outcomes, a kind of counter-performativity, as well as explicit efforts at institutional design, a subset of effects I discuss next, as well as the effects of economists' work in developing theories, which I will discuss in section 6. The task in the remainder of this chapter is to narrow down these mechanisms to develop more coherent notions of performativity.

5. Design performativity

Economists are increasingly often engaged to assist in developing various economic institutions, establishing the causal conditions that will support their emergence. For example, teams of such economists are engaged by the World Bank to assist developing countries to develop institutions such as stock exchanges and central banks, though some quarter of projects fail to do so successfully (Kilby 2000) and others have been engaged in designing auction systems mobile phone spectrum (Binmore and Klemperer 2002)

Such explicit, intentional, institution building projects naturally involve actions by economists that can be described as performative, in that the resulting institution is in part generated by them. But this performativity takes place in the causal components of the institution and is explicit and intentional. It is often driven by legal instruments such as laws and regulations that are enforced by courts that can be explicitly framed for the purpose of creating institutions. "Social engineering" is therefore a better term than "performativity", though for the sake of analysing the performativity concept, let us see this as a special type of performativity, which we may call design performativity:

Design performativity. Economics is design performative iff social institutions emerge from causal components created by economists with the intention of causing the institution.

Design performativity is the result of economics in the engineering stance, about which I will say much more in the next chapter, and is the most constructivist interpretation of performativity. But when it comes to the performativity thesis advanced by Callon and MacKenzie, design performativity is banal. To make the claim that institution builders have an intentional effect on institutions is rather unsurprising. The notion of intentionality presupposes that human beings are able to affect their environment. Institution building is complicated by reflexivity, one reason that the World Bank's projects may fail so often, but the fact that certain individuals are tasked with the creative process and thereby influence outcomes is mundane. More interesting is the fact that economists seem to have become more influential in driving social policy, but this fact does not imbue economics with unique and exclusive powers of performativity.

Callon cites the development of the European Central Bank and the influence of Milton Friedman's monetarist theories as an example of performativity. He writes (Callon 2007: 324):

> We can say that the creation of a European central bank, directly inspired by the monetarist theses of Milton Friedman, helps to make real monetary markets correspond to the descriptions and analyses proposed by theories or models qualified as abstract. Similarly, enforcing incentives inspired by economic theories and their assumptions about human or organisational behaviours causes the behaviours to fit the theory's predictions.

Here Callon is arguing that the theory serves as a kind of blueprint for the creation of the institution. I will discuss the role of blueprints at length in the next chapter, but the point Callon is making is that the theory provides a template for a particular kind of institution.

Similarly, MacKenzie uses the example of Chilean economists trained by Friedman and his University of Chicago colleagues. "Especially under the government of General Pinochet, the 'Chicago Boys' did not simply analyse the Chilean economy; they sought to reconstruct it along the free-market, monetarist lines whose advantages they had been taught to appreciate" (MacKenzie 2006: 16). MacKenzie describes the Chicago Boys example as a “vivid manifestation of a general phenomenon,” continuing (2006: 16, italics as per original),

> The academic discipline of economics does not always stand outside the economy, analysing it as an external thing; sometimes it is an intrinsic part of economic processes. Let us call the claim that economics plays the latter role *the performativity of economics*.

There are, however, three mechanisms at work in these examples that Callon and MacKenzie do not distinguish. The first is design performativity as I have defined it. The Chicago Boys were appointed to positions where they were able to deliberately shape institutions, attempting to control the causal process giving rise to organisational structure and institutions in Chile. The second is weak performativity, in that economics can be “an intrinsic part of economic processes” merely by having *some* causal effect, of any type, including unintended effects. The Chicago Boys certainly occupied powerful positions in Chile and had multiple causal effects on the institutions within it, both deliberate and unintended. Weak performativity, as already discussed, is not saying much. But there is a third meaning of performativity that MacKenzie and Callon seem to have in mind but do not clearly distinguish, which is the notion of economic theorising having an effect on economies, simply by teaching the theory. This is the idea of theories themselves becoming “an intrinsic part of economic processes” in many different ways. The performativity writers do not explicitly distinguish this type of performativity from design performativity but doing so brings much clarity to the notion. I draw this out next.

6. Theory performativity

Economics students behave differently in strategic choice situations from students who have never studied economics. One possible, though disputed, explanation is that the theories they have been exposed to have affected their behaviour to make them more like the rational agents postulated in economic theory[6] (Marwell and Ames 1981). Similarly, the creators of the Black-Scholes model successfully lobbied traders in Chicago to use it to price derivatives they traded on financial markets, which changed prices and market infrastructure, at least until the 1987 market crash (MacKenzie, 2006). In these cases, theories in economics seem to have changed behaviour in the world.

Examples like these pick out a particular type of causal mechanism at work in the idea of performativity that so far has not been made explicit. This is the ability of theories to affect behaviour in an economy simply by being theories that people have encountered and come to know. The effect is quite particular: it is to cause behaviour that is predicted by the theories. This is a special subset of effects, one that excludes unintended and deliberately engineered effects. It is this subset that I am going to characterise as exhibiting "theory performativity".

In the next subsection I will distinguish three types of economic theorising – normative theories, positive theories, and testing models for institutional designs. These distinctions will support my argument that the concept of performativity is best understood as situations where actors in an economy learn from normative or positive theories about the worlds they are embedded in. I will show that this line of argument has important ontological consequences though not the ones claimed in much of the performativity literature.

[6] Though another explanation is that students prone to rational analysis are more attracted to economics in the first place. Questions have also been asked about the gender make up of Marwell & Ames' study groups which may have biased results more than the study of economics (Frank, Gilovich, and Regan 1993). While Marwell & Ames's research may not be definitive, it serves to illustrate the idea that performativity theorists convey, that economic theories influence the behaviour of those exposed to them.

6.1 Three types of economic theory

There is no doubt that economists aspire to make a difference in the world, by shedding light on some economic phenomenon that had not been well understood, or by developing ways of understanding how to achieve specific goals. In the next chapter I will explore at length the nature of economic reasoning, at least as it applies in finance, and argue that financial theorists are involved in both theorising about the world and in explicitly designing and engineering institutions. The primary method economists and financial theorists employ in reasoning about the world is the model, a device that aims to bring underlying phenomena to light, help calculate implications, or serve as blueprints for institutions.

My aim in this section, however, is to focus on how theories and the models associated with them may be performative. To bring that to light, I will take a broad brush approach to categorising economic theory into three different forms: normative theory, positive theory that makes claims about real processes in an economy, and thirdly, institutional design. This three-way distinction has a long history. John Neville Keynes[7] (1904) argued that economics is part science, part art, and part normative (Weston 1994). Keynes's idea of "normative" referred to what ought to be in the moral sense of right or wrong. His "art" was the creation of systems of rules that would provide for a given end, a form of institution building. The distinction between "ought" and "art" is subtle. Saying an agent "should" behave in a certain way could mean she "ought" to, by appeal to a value system, or because she should do so by reference to specific goals. The "art" is developing the

[7] Father of the more famous John Maynard Keynes, referred to elsewhere in this thesis.

institutional structure that makes it the case that certain goals will be reached if one behaves in a certain way. The art supports a different sense of normative, making it the case that specific behaviours will allow individuals to achieve specific goals. This latter case is not obviously normative in a moral sense, but rather the practical advice of the economist, saying that if you want to achieve an end (say, maximise returns for a given amount of risk) then you should behave in a certain way (invest in a diverse portfolio of assets). Keynes' notion of science, of discovering things about the world, is what I conceive of as positive economics. His notion of normative – in the sense of moral oughts – can be conceived of as a special case of desired ends within a broader notion of normative that includes the pursuit of other desires. This interpretation is consistent with Milton Friedman's concept of normative, a category of economics concerning "how any given goal can be attained" (Friedman 1953: 3). Modern portfolio theory is a subset of what Herbert Simon calls normative microeconomics, that part of economics developed to guide individual decisions in attaining given goals (Simon 1959). Keynes's art includes the process of institutional building.

There is one further way economics may affect society. Research programmes usually develop their own set of concepts and linguistic terms that are used in developing a body of theory. This is certainly true of theoretical finance. These sociological aspects of the act of theorising can also have in impact on the world merely by establishing and spreading concepts used to talk about the world. Financial professionals are well-versed in the language of modern portfolio theory, for example with concepts like alpha and beta as terms for risk. By describing financial markets in this way, they may then also be able to affect how it operates. This aspect of economic theory has already been discussed in Chapter 2 in the context of Thomas Kuhn's paradigms in science. One aspect of a scientific paradigm is a consistent way of seeing and describing the world.

In summary, then, there are three ways we may expect theory performativity to happen:

1. Positive theory. By revealing facts about the world, people may learn and change behaviour.
2. Normative theory. By connecting people's objectives to particular choices, choices might be changed.
3. Sociological impact. By disseminating particular linguistic and conceptual frameworks from research programmes, economists can affect the way people see and talk about the world.

These provide a set of possible effects, all of which could result in some form of performativity, but they are effects that are different to the design performativity of the previous section. In the next section I consider how these effects are thought of in the performativity literature.

6.2 Theory performativity in the literature

Callon's concept of performativity is strongly influenced by the Austinian concept of performativity – the speech act ontology by which social facts are brought into existence by declaration, much like the Searlean theory of social ontology discussed in Chapter 3. In describing Donald MacKenzie's work on the development and impact of the Black-Scholes model (2006), Callon explains the effect of the theory on the world as follows (2007: 320 parentheses his):

> What MacKenzie describes with surgical precision is the gradual actualisation of the world of the formula: a formula that progressively discovers its world and a world that is put into motion by the formula describing it; a formula that previously functioned in a paper world, which was perfectly real (for what could be more real than paper or equations?), subsequently functions, after many investments, in a world of computers and silicon, algorithms, professional skills, and cleverly adjusted

> institutions. We could say that the world it supposes has become true, but it is preferable to say that the world it supposes has become actual[8].

Callon here illustrates some important elements of his notion of performativity and its relationship to theory by focusing on models. He describes the model as real in the "paper world". This is consistent with the view of models as abstract simplifications that illustrate theorised mechanisms. The model is a device that shows how the theorised tendency may be true. This truth is deductive, in that, given the specifications of the model's inputs and mechanisms, the consequences follow deductively. What Callon notes in this extract is the gradually increasing correspondence between the world of the Black-Scholes model and the actual world, achieved through a process of "adjusting institutions" and the technical infrastructure of computerised trading and professionalised behaviour so as to allow the model to "become actual". Callon's notion is sociological: the model plays a central role in coordinating individual behaviour, driving institutional design and the development of technology that underpins markets for derivatives. So his vision of performativity, at least as he reveals it here, is one of theoretical economic models as central causal drivers. Models, on this account, are not positive accounts of real facts in the world.

Callon does not have much to say about why some models manage to spark this causal chain and others do not. Obviously the annals of economics are littered with models that no one took seriously, so there must be some reason some models have this performative causal capability and others do not. Callon argues that performativity occurs through successful repetition and a minimum of failures in the

[8] This process was, as MacKenzie (2006) points out, temporary, with the match of theory and reality decaying after the 1987 market crash, at least with respect to pricing of derivatives on the Chicago Board of Exchange.

process of repetition. An institutional process shaped by a model that constantly failed to deliver the expected outcome would likely decay. An unexpected event can be interpreted in terms of categories offered by the model and thereby be managed within the institutional structure based on the model, but some events will fail to do this, leading to crisis. A model can be adjusted in response to unexpected events in order to incorporate them into the model's framework, thereby allowing the real world institutions to do so too, but this is not always possible. Coining an interesting neologism, Callon distinguishes "between performations that manage to produce regularities and repetition and performations that are constantly faced with unexpected events, which they sometimes – only sometimes – absorb for a while" (2007: 326). Crisis awaits those institutions, and models, which fail to manage the unexpected events. We are not told why exactly models can't be adjusted for all unexpected events. Presumably some are so far from the theoretical world of the model, and so incapable of being explained in those terms, that the model fails as a guide to the world. This could perhaps be interpreted as a variety of Kuhnian crisis, with the correspondence between the institution and theory surviving during a period of "normal science" only to fall with an accumulation of empirical disconfirmation. But Callon would likely resist this characterisation given that it would presuppose some empirical process by which the model is tested against the real world, and found wanting. As Callon puts it (2007: 322):

> To predict economic agents' behaviours an economic theory does not have to be true; it simply needs to be believed by everyone. Since the model acts as a convention, it can be perfectly arbitrary. Even if the belief has no relationship with the world, the world ends up corresponding to it. We can thus consider that the famous Black and Scholes formula has no truth value, that it says nothing of real markets, and that it is simply a coordination tool that allows mutual expectations. It constitutes a false but effective representation, and can be seen as pure convention.

This passage makes it clear that Callon's account admits of no positive scientific aspect to economic modelling, which can be "perfect arbitrary". Economists have imposed the model, consistent with the strong interpretation of performativity discussed above. The model "acts as a convention" so somehow it is able to enforce a desired social arrangement. This is an odd idea of how conventions come about. It can be contrasted with David Lewis's, who describes conventions as Nash equilibria from which individuals have no incentive to deviate (Lewis 1969). A Callon convention seems to be blind to the incentives and motives of individuals, other than to adhere to an otherwise arbitrary model created by economists.

Callon's notion has conventions plucked out of thin air by the arbitrary models. He does not allow that the success of some models may be because they pick out a real feature of an economy, and that institutions built on the collective learning from the insights models provide would be more successful. The success of models comes down to the ability of the model to continue to "perform" the world, with that ability in no way determined by whether the model was "right" about the world in the first place.

Callon does not give a clear definition of this notion of performativity, but I think the analysis above provides the basis for the following:

Def: Callon's theory performativity: An economic theory is always performative and generates social structures as long as they can cope with unexpected events.

This differs from strong performativity only in that it allows for decay through a vaguely described process. I will show that this concept of performativity cannot be sustained. The following definition is preferable:

Alternative def: theory performativity: Economic theory is performative iff it changes behaviour in response to normative and positive claims that individuals accept and use in making decisions to reach their goals.

To analyse these two definitions, let me return to my general model of social ontology through emergence. Recall that a whole, *W*, with properties *E*, emerges synchronically from components $a_1, \dots, a_n$, each of which have their own properties

$P_1, \ldots, P_i$, provided that a certain organising structure, O, applies to the a_is. W emerges diachronically from a set of causal conditions $b_1, \ldots, b_m$.

The crucial difference between Callon's performativity and my theory performativity is this: Callon's takes place at the causal conditions $b_1, \ldots, b_m$ whereas mine happens at the components $a_1, \ldots, a_n$ of the whole. For Callon, the economic model specifies the world that must exist so it sets down all the causal conditions for the resulting institution. That institution is then brought into existence by the theory. The theory is the complete cause of the institution. On my account the theories are components in the institution, which is dynamic. Those components respond reflexively to the whole. Therefore economists, and their theories, are part of the dynamic system that is the institution. Their behaviour is changed and theories changed in response to the properties of the institution as a whole. This allows for positive economics – features of the world are exogenous to models. Economics can therefore be responsive to the world, even as that world is affected by theories. On Callon's account, the theory is not *a* cause, it is the entire cause. On my account, the causal process is multidirectional.

On Callon's account, the relationship between the emerged institution and model can only be threatened by unexplainable events, when properties of the whole, E, fail to correspond with the predictions of the model. On Callon's account, such failures disrupt the causal conditions and a decay occurs between the model and the real world. On his account, a failing model is then replaced by an alternative, more successful model, presuming such is available. He is silent as to what happens in the absence of an alternative. On my account, a model that has faced empirical disconfirmation, such as the Black-Scholes model post 1987, can be still be useful as one source of insight, but treated with some circumspection by practitioners.

Callon's account leads to a contradiction. Economists' models are determinative of the world, yet the decay happens when the properties of the world no longer conform to the expectations of the model. A new model then has to accommodate those failures. It follows that the new model must be empirically responsive to the world. The new model is determined by events in the world, rather than having complete control over the organising structure or the causal conditions that allow it to emerge. Empirical factors must be involved. Callon's theory performativity cannot escape this contradiction.

My notion of theory performativity accommodates the involvement of empiricism *and* effects by economists on the world. At the component level, theories, all relevant ones at once, are among the components $a_1, \ldots, a_n$ from which the whole synchronically emerges. For the theories to have a causal effect on the emergent institution they have to affect other components, shifting their behaviour. Recall that the *O* is a property of the *W* – in other words, one of the features of the whole is that it has a specific organising structure for the components. In some cases theories may play an important part as components of an institution in propagating knowledge and conceptual frameworks that may be important for the institution to work. But that does not mean the theories are immune to change.

My account also accommodates one of the key observations of the sociologists, that individual actors in an economy seem to be behaviourally conditioned by some economic theory. For example, if economic theory proposes that agents are "calculative" (Callon 1998: 15) in determining their optimal decisions, consumers may adjust behaviour to be more calculative, because they desire the outcomes normatively specified by the model. But on my account of social ontology this should be no surprise – economists in their interaction with components at the organising structure level can propagate theories and affect behaviour simply by providing normative guidance. They do so merely as a part of the organising structure. Those economists are also simultaneously affected by what other people do and the properties of the institution itself as they develop positive theories. There are also multiple other influences on the components within the organising structure including non-economists affecting each other.

My account provides a more plausible mechanism for economic theory to affect behaviour and also for economies to affect theories. Economists study social institutions. They are, in that sense, affected by the institutions, their organising structures and their historic causes, to the extent that they aspire to make positive claims about these. As Mary Morgan puts it (2012: 401),

> …the process of interaction between the economy and economists has never been one-way. Just as economists as policy advisors have tried and sometimes succeeded in shaping things economic, so sudden and unexpected changes in economic behaviour and events have equally shaped and prompted changes in economics.

Economic models can help to understand mechanisms and processes at work throughout the social world. In the complex web of mechanisms and processes that affect institutions such as the financial system, models are an important epistemic tool. My account allows for multiple models to simultaneously exist, consistent with Dani Rodrik's notion of a "library of models" as the main methodological strategy of economics, as a horizontal science that widens its explanatory mechanisms rather than a vertical science continuously trying to perfect the one description of reality (Rodrik 2015). The contrast between mine and Callon's theory performativity is clear – Callon's demands that the causal conditions in *the* model are made real, whereas mine has no such demand, merely that models become causally involved with the components in the organising structure.

My account allows for economists to be components within organising structures and, as they try to develop knowledge about those structures, also affecting those around them. It also is consistent with a variety of claims about the sociology of science as it applies to economics including the methodological issues highlighted in Chapter 2 of a hard core commitment to a frequentist ontology of risk in traditional theoretical finance. Economists can still be seen to be functioning in Lakatosian research programmes.

It should also be clear that my theory performativity differs from weak performativity as I defined it above. Theory performativity happens specifically because people learn and adjust behaviour in light of the theory. So the effect is specifically that the theory changes behaviour because people believe the theory to be true, or at least partly true, either because they've been convinced by economists or otherwise come to know it. This is also quite different to the notion of a self-fulfilling prophecy which is made true because a claim triggers a behaviour that leads to the claim becoming true. The notion of self-fulfilling prophecies points to causal mechanisms outside of any normative claim by economists, such as the "madness of crowds" that causes irrational behaviour like racism and runs on solvent banks.

Consistent with Callon, economists can affect the causal components that lead to specific assemblages of components and organising structures. But where economists are engaged in this form of social engineering it is an example of *design performativity*, as I've defined it above, and not of theory performativity as I've defined it in this section. Theory performativity happens as economists affect other components within organising structures through their normative and positive work. It is separate from the deliberate shaping of conditions to cause the diachronic emergence of institutions.

There is one respect in which this may not appear to be true, which is when certain knowledge is a precondition for the diachronic emergence of institutions. For instance, a society may have to be educated in various respects such as reading and basic mathematics in order for many types of institutions to exist, and this may be true of some theoretical economic models too. It may also be the case that institutions like those of the financial system require a pipeline of business school graduates taught to understand and speak the language of modern portfolio theory in order for it to function. Such background facts form part of the causal conditions for any institution to emerge. It is important to recognise that economists are in a sense prisoners of these facts. The institutions they design have to take such background facts as given.

7. Conclusion

To be clear about the relationship between economic theorising and social ontology I have set out in this chapter, let me summarise it as follows. Economists can be engaged in social engineering in conjunction with policy makers in order to create social institutions. This is *design performativity* and affects the causal history to the emergence of an institution. Economists are also themselves agents within any economy and their actions have effects on those around them. This is *weak performativity* and follows trivially from the fact that economists and their theories are inevitably components within societies. Economists also develop positive theories about features of the world and normative theories to determine optimal decisions given specific goals. Where these affect the behaviour of decision makers, this is *theory performativity*.

I also rejected the concept of strong performativity. I described this as the claim that economists exercise omnipotent powers over the economy at large, a claim which seems preposterous but one that performativity theorists accept to some extent. When it comes to theory performativity, the standard literature suggests a process by which theoretical models are imposed onto the world so that institutions such as financial markets are shaped to match those models. I argued that this account collapses into strong performativity. It cannot explain how institutions might decay without conceding that economists respond to the institutional failure and attempt to develop new models to accommodate the failure. This inevitably introduces positive empiricism to the task of economists.

The notion of theory performativity that I defended works through three mechanisms. First, individuals learn from positive theories about how the world works, and use those theories to adjust their behaviour in pursuit of their goals. Second, some theory provides normative decision tools to employ in order to rationally achieve their objectives. Third, certain research programmes in economics can develop linguistic and conceptual frameworks that become widely used in a society. This can provide background conditions for certain institutional change to come about, several of which I will discuss in the next chapter.

I distinguished theory performativity from design performativity, which is the explicit effort to design institutions by economists working together with policy makers. From our financial markets to auction systems for public goods, economists have weighed in on the blueprints for their creation.

In the next chapter I will examine in depth the difference between theory and design types of reasoning. I will link theory performativity and design performativity to financial theories, showing how they affect the financial system and how the financial crisis might be expected to affect them. With a clear concept of design and theory performativity in hand, the way is open to achieving this clarity.

Chapter 5

Theorising and Designing in Finance

1. Introduction

The aim of this chapter is to explore the nature of reasoning with models in theoretical finance. I argue that the use of models in theoretical finance serves not only an epistemic purpose, but also an ontological purpose, because models are used both to study the financial system, and to shape financial institutions. The key argument of this chapter, and the culmination of this thesis, is that the epistemic function of theory models is not appropriate when those same models are used as designs for institutions in the financial system.

This continues a clear theme in this thesis, that finance is not only a positive science attempting to make sense of the world, but also a task of social engineering; of designing and creating the institutions of the financial system. In this penultimate chapter, I argue that one reason social scientists were surprised by the financial crisis is precisely the morphing of theoretical models in finance from devices in the employ of a positive science into blueprints for institutions. While models were deemed theoretically adequate and useful to reasoning in finance, they are not adequate as blueprints. We were surprised at the failure of institutions built from theoretical models because we thought their epistemic success as theory models provided us with good grounds to expect their success as blueprints.

The argument has the following structure. First I distinguish the different epistemic requirements we should have of theory models and design models. I do this by introducing the notion of two "stances", the theory stance and the design stance, that financial theorists might take. I also distinguish three categories of models, which I call theory models, blueprints and scale models. The last of these is used in both the theory stance and the design stance, but in different ways.

These new concepts allow me to analyse models in theoretical finance, particularly mean variance optimisation (MVO) and the Capital Asset Pricing Model (CAPM), and related models such as the Black-Scholes model and Value at Risk (VaR) modelling. I argue that these models serve an epistemic role of illustrating and describing real features of the world, when seen from the theory stance. However, I show that the epistemic requirements change when we take the design stance, and these models fail to provide adequate grounds to accept them as blueprints for institutions. The assessment of blueprints occurs through types of experiment, such as testing scale models and simulations. In contrast, theory models are tested for logical coherence and often by the coherence of predictions with data. As a general sound bite that captures much about the distinction, we can say that in theory models simplicity is a virtue while in design models it is a curse. This sound bite is only partly helpful – there are simplifications that are important in the design stance too. In general, modelling in the design stance, particularly when using scale models, is about introducing complexity to test how designs will work in the real world. In contrast, theory models often benefit from idealising and abstracting the world into simple causal processes and inputs in order to support claims about general features of the world, as a variety of thought experiment.

In the previous chapter I distinguished between what I called theory and design performativity. I argued that theory performativity occurs when behaviour is changed because of positive or normative implications of theories and practices related to them. People learn from theory and change their behaviour as a result. In contrast, design performativity is the result of financial theorists explicitly creating institutions in an economy. I argue that such institution-building, particularly in the financial system, treats the theory model as a blueprint, and in so doing attempts to impose the simplifications used in the theory model onto the world.

Theories and models form components within an organising structure from which social institutions emerge, whereas social engineering involves intervening in the historic process that leads to the organising structure. Chapter 4 provided the conceptual background to the analysis in this chapter that draws out just how the models of theoretical finance can be used in social engineering.

In the next step of the argument I examine specifically how the MVO and CAPM have contributed to both theory performativity and design performativity. I argue that both models are useful in the theory stance in two different ways. MVO provides reasons for why investors should diversify their exposures. The CAPM provides reasons why riskier assets should provide higher returns. These are, in a sense I will discuss, positive claims about the world. These claims might be persuasive about not only how the world does work, but about how it should work, specifically in respect of our own financial decisions. In this latter normative role, models in finance provide decision rules to apply when deciding whether to buy or sell assets. Modern portfolio theory is a subset of what Herbert Simon called normative microeconomics; that part of economics developed to guide individual decisions (Simon 1959). Normative models provide reasons for certain decisions, of the form “to achieve X, do Y”. Debates about the normative/positive distinction in economics often concern ethical or ideological values at work in economics, which are described as “normative” considerations (Weston 1994). However I use the term as Milton Friedman does, as a category of economics concerning “how any given goal can be attained” (Friedman 1953: 3), of which ethical goals can be seen as a subset. As discussed in the previous chapter, this guide to decision making results in theory performativity because individuals learn to apply the model to their own behaviour.

I then argue that the models of finance have also been used as *blueprints instead of theory models*. Designers in finance try to design the world to be more like the model world. There are two ways this can be true. The first is to explicitly use the model as a blueprint. The second is to engineer institutions that are designed to make the normative implications of a model true, often by relying on the theory performativity caused by the model as a background condition. In this latter sense, the model works both for theory performativity and design performativity, with the two interrelated and mutually reinforcing the other. However, my analysis allows one to pull apart these two effects, revealing the role of the models of finance in guiding social engineering. By drawing on my analysis of the different types of models in the design compared to the theory stance, I show why it may be problematic to use the models of finance in this way.

I begin by separating the epistemic requirements we should have of theory models and design models.

2. Stances, theory stances, and design stances

It seems impossible to escape intentions when reasoning about models in science generally, because the representational role of models requires an understanding of what agents intend the models to represent (Giere 2010). Consider a rudimentary model, such as one of the solar system commonly found in classrooms that shows the sun in the middle and planets orbiting it. To understand how to interpret such a model, we need to understand what the components of the model are intended to represent, for example that small balls are planets and that the wires holding them in place are not relevant to the narrative supported by the model. Whenever models are used to explain, the audience inevitably either explicitly or implicitly is taking an intentional stance, putting themselves in the shoes of the modeller to understand how to interpret the model. Such forms of explanation relying on speculation about mental states have been shunned in the behaviourist climate that has governed much of economics over the last century, but I think it is reasonable to claim that intentions matter, and that they form a reasonable basis to search out a difference in the types of models that are used in social science.

In this section I develop two separate categories of intentional stance, which I call the theory stance and the design stance. The purpose of doing so is to lay the foundation for analysis of the type of models that may be used in each stance, the epistemic requirements we should have of them, and the standards by which we may assess them. The concept of the "intentional stance" has been developed by Daniel Dennett (1989). By taking the intentional stance, we hope to understand another person, by reasoning that her visible actions are caused by her intentions. We are interested in this because it provides some predictive power: intentions are reasons for action. Reasons provide explanations for why an agent chooses to do what she does (Davidson 1963). I think it is useful to take the intentional stance of the financial theorist and ask how their approach to the world differs depending on whether they intend to design some institution in the world, or to describe and explain some fact in the world.

Dennett uses the term "design stance" in the context of predicting the behaviour of built objects (Dennett 1989). For example, if we wanted to understand the function of an alarm clock, we could either exhaustively investigate its physical properties to understand its mechanisms, or we could reason that it was designed by someone who wanted to build an instrument that sounds an alarm at a set time. The latter approach would be a more powerful epistemic strategy, in that we will get to know the mechanisms and make predictions about the clock more efficiently. If we are to make sense of the things that economists produce, among which are models, we can do so by taking the design stance. In my usage the term "design stance" is somewhat different Dennett's – I am specifying a subcategory of the intentional stance where the intention is to design something in the world, and contrasting it with another subcategory where the intention is to come to know something about the world, which I call the theory stance.

Economists do intervene in the world to develop particular institutions. This became much more prominent towards the end of the last century. Governments increasingly looked to economists to help design various institutions, often based on market mechanisms. Examples range from the distribution of scarce resouces like radio frequencies (Binmore and Klemperer 2002), to the design of labour clearinghouses to match people to jobs (Roth 2002). Financial theorists have similarly been intimately involved in the development of the world's financial markets (Bernstein 1993). Financial theorists were asked not only to provide theoretical insights that may be relevant to the way institutions might work, but also involved themselves in actively designing and then engineering the institutions themselves. I will discuss several examples in this chapter. Economists and financial theorists are not alone in engineering social institutions; entrepreneurs build companies, politicians pass legislation, researchers in many fields are also called on to design institutions that fit their expertise. Philosophers have also often seen themselves as social engineers, at least since Socrates's design of the just constitution in Plato's Republic. But financial theorists draw extensively on the toolkit of the research programme, particularly models, in doing so.

Different designers from different traditions can combine efforts, though these can produce hybrids of inconsistent ideas that fail to work in practice. For example, economists' proposals for the California electricity market in the 1990s were partly overruled, resulting in a market in which wholesale prices could fluctuate according to market forces, but retail prices to consumers were capped according to a scheme implemented by politicians. This hybrid brought utility companies to the verge of bankruptcy (Roth 2002). This example illustrates that designs can be good or bad, and it is worthwhile to carefully assess any design before it is implemented. The assessment of designs is a key part of the work undertaken by economists, as it is by engineers in the case of built objects. This type of assessment would take place in the design stance.

2.1. The difference between models and experiments

My distinction between theory and design stances is somewhat like a distinction that Francesco Guala has made between two “traditions” in economics, that of “builders” and “testers” (Guala 1999, 2007). His “building” tradition is similar to what I describe as the design stance, while his “testing” tradition has much in common with what I call the theory stance. Guala’s focus is on experiments and how they are used in each stance. In contrast, my focus is on models, but there is a sense in which models can be seen as experiments, and experiments as models (Mäki 2005). It is worth spending a moment considering just how models in economics might be like experiments as we commonly understand them in the physical sciences.

The claim that models in economics are like experiments rests on two premises. First, that models in economics are fundamentally similar to models in physics. Second, that experiments in physics have a relationship to models in physics that is analogous to the relationship between experiments and models in economics. I will argue the second premise is valid in the theory stance but invalid in the design stance. In this section I argue that experiments in physics are analogous to models in economics, only in the theory stance and not in the design stance.

The typical experiments of the physical sciences are carefully constructed set-ups created to demonstrate that some predicted outcome is obtained. The similarity between models and experiments becomes clear in Nancy Cartwright’s concepts of nomological machines and socioeconomic machines. She describes the first of these as follows (Cartwright 1999: 50, parentheses are hers):

> What is a nomological machine? It is a fixed (enough) arrangement of components, or factors, with stable (enough) capacities that in the right sort of stable (enough) environment will, with repeated operation, give rise to the kind of regular behaviour that we represent in our scientific laws.

On Cartwright’s account, experiments are nomological machines. They are contrived situations that are set up to show a law working provided certain shielding conditions are met. Cartwright thinks models in economics work much the same way, but now called socioeconomic machines instead of nomological machines (Cartwright 1995b). Much like experiments in the physical sciences, these are shielded worlds in which the anticipated regularities are able to be generated. *Ceteris paribus* clauses do the job of setting up the shielding conditions for the socioeconomic machine to work, for the tendencies to be made clear. In Cartwright's view, there is no difference in the metaphysics of physics and economics. "Economics and physics equally employ *ceteris paribus* laws, and that is a matter of the systems they study, not a deficiency in what they have to say about them" (Cartwright 1999: 151). Note however that the similarity is of experiments in the physical sciences and *models* in economics. It does not necessarily follow that *experiments* in economics are the same as experiments in the physical sciences. The argument I will develop is precisely that the experiments in the design stance, and the blueprints for institutions that are tested, are disanalogous to models in the theory stance and that a failure to recognise this is problematic.

The role of experiment in the design stance is specifically to introduce "real world" properties into a design, to see how it works. As I will explain in the next section, one good set of examples is the use of scale models by engineers. The ambition is to see how the built object might behave in reality, where the shielding conditions of the socioeconomic machine no longer apply. My account is therefore somewhat like the market experiments that Guala (2007) discusses, but quite different to the idea of experiments being just like models that Cartwright and Mäki discuss. The experiments I discuss are precisely *not* like models in that they are not closed worlds.

A final point on the distinction between the theory and design stance is that it is certainly not unique to financial theory, or to economics as a whole. Consider the relationship between physicists and civil engineers. One can work on developing theories about the nature of physical systems, or one can work on designing a particular structure. A bridge designer would draw on Newtonian physics and other theories in developing a design, just as an economist will draw on equilibrium analysis and other models in designing a market. We have come to see engineering and physics as quite specialist fields, and perhaps in time this will be how we come to see the design and theory stances in social sciences, though there are many economists who happily shift between stances depending on the work they are doing.

This provides the grounds I will use to analyse the distinction between the types of models that may be used in each stance in the next section. What should be clear from the last two sections is just how the theory stance and design stance differ according to the intentions of the economist.

3. The types of models in each stance

There is an extensive literature on the nature of reasoning with models across the sciences. I could not hope to do this literature any justice. Instead I want to make some claims about models that I will draw on specifically to make a point about models in theoretical finance and how they are used, on the one hand, to describe, explain and understand the world, and on the other hand, to intervene on it. I will make this point by examining them from the standpoint of the theory stance and design stance.

Models are examples of "surrogate reasoning" that share the strategy of using one thing – a set of mathematical equations, a physical scale model, a graphical representation, among others – to explain another thing, the target object (Mäki 2013). The relationship between model and target is complex and fills much of the discussion around the role of models in science. The idea that models *explain* something about a target object is relatively uncontroversial, though just how this explanation is achieved can be deeply controversial. Explanation in this sense must be broadly construed – it can range from narrative story-telling (Morgan 2012) to experimental illustration (Mäki 2005).

3.1 Theory models

I am going to ignore much of the complexity explored in the literature regarding models and focus firstly on the typical model of the theory stance in economics and then show how this is also true of finance models. This will provide sufficient detail to support my argument regarding the role of theory models in the theory stance compared to the design stance. The theory models that form the bread and butter of economic reasoning are what Dani Rodrik describes as "simplifications designed to show how specific mechanisms work by isolating them from other, confounding effects." He goes on (2015: 12):

> A model focuses on particular causes and seeks to show how they work their effects through the system. A modeller builds an artificial world that reveals certain types of connections among the parts of the whole – connections that might be hard to discern if you were looking at the real world in its welter of complexity.

Rodrik sees models as ways of making causes explicit by simplifying the complexity of the world, developing an "artificial world" but one that still maps on to parts, and the connections between them, in the real world. This is substantially the same as Cartwright's notion of socioeconomic machines that are protected by *ceteris paribus* conditions. The aim of the model is to focus attention on a particular claim about a causal mechanism that exists in an economy. It follows that a good model is one that is able to provide such insight in a straightforward way. So, a theory model is useful precisely because it is simple and yet explanatorily powerful, a claim consistent with Milton Friedman's famous description of the methodology of positive economics, in which he argues for a kind of instrumentalism about models (Friedman 1953: 14):

> Truly important and significant hypotheses will be found to have "assumptions" that are wildly inaccurate descriptive representations of reality, and, in general, the more significant the theory, the more unrealistic the assumptions (in this sense). The reason is simple. A hypothesis is important if it "explains"

> much by little, that is, if it abstracts the common and crucial elements from the mass of complex and detailed circumstances surrounding the phenomena to be explained and permits valid predictions on the basis of them alone.

Friedman's point is that a model can benefit from being simple even if this means very unrealistic assumptions, because it thereby is able to act as a powerful epistemic device in the service of making predictions about the world. Friedman's focus was prediction rather than explanation, but some of his claims about the virtue of simplicity appy equally to models that do explain. Rodrik's view of models sees them as accurately representing certain features of the real world by focusing attention on them using abstraction. Rodrik therefore departs significantly from Friedman's emphasis on predictive success, but agrees with the importance of explaining "much by little", providing tools for economists to understand society. The general causal mechanism that a model illustrates can be expected to be found in reality provided that the *ceteris paribus* conditions that isolate the model world are satisfied, consistent with Cartwright's notion of a socioeconomic machine.

This conception of models works to understand the CAPM and MVO models of finance when they are seen from the theory stance. I will explore this thoroughly in section 4, but this discussion shows how the models might be constructed to provide explanations of the properties of assets and markets.

Rodrik's conception of models is not exhaustive of those used in theorising in economics. While my concept highlights the importance of simplification, others encourage complexity, for example, macroeconomic models run by central banks that attempt to model the causal impact of monetary and fiscal policy on a specific country. These are complex and greater complexity tends to support the accuracy of the model (Hoover 2001). For my purposes, however, the Rodrick notion of theory models provides a sufficient basis for understanding modelling such as that represented by MVO and CAPM.

3.2 Blueprints

In contrast to theory models, models in the design stance are used to support the development of real institutions. Such models provide instructions for the construction of institutions.

Blueprints are designs for physical constructions but they are good metaphors for the process of designing social institutions and many writers on economic institutions borrow the term in describing their projects (for example Sachs 2008; Pearce, Markandya, and Barbier 1989). Similarly, I will borrow the term to describe one type of model that is used in the design stance. In the context of the social world, a blueprint for an institution can consist of a narrative description of the institution to be established, including descriptions of a set of rules and regulations that will be implemented. In both physical and social systems, what matters is that these are instruction guides to assembly. The analogy might be challenged because a physical blueprint seems to depend substantially on diagrams, while the blueprint for a social institution is generally presented as text, though of course diagrams may also be helpful. I don't think this distinction is fundamental in the concept of a blueprint. If we recognise that laws and regulations, and other causal conditions, are essential to engineer the organising structure from which an institution emerges, then the plan for those rules and regulations are like the walls and rafters of a building. It follows then that we can consider an institutional plan; the guide to the conditions to be created and implemented, as analogous to the blueprint; the guide to the walls and rafters to be constructed. The blueprint leaves much unsaid, including the physics that describe the behaviour of the walls and rafters, just like an institutional plan may leave much unsaid, such as the economics that describe the behaviour of individuals in an institution. Blueprints are not an exhaustive description of reality, and many precise details are not included.

Blueprints are a type of model. When the target object is a physical structure, blueprints are a kind of small-scale representation of the structure to be built, in a similar way to how a map is a miniature version of the world. There are similarities and differences between maps and blueprints. They are similar in both employing abstraction, for example, a blueprint for a building represents walls as two dimensional lines like a map might represent roads as two dimensional lines. These representations have to be recognised by the user of the map or blueprint. They are therefore considerably abstracted from the thing to be built or the roads that actually exist. This, however, points to the differences. Maps are about the world and serve the purpose of getting you around it. The London Underground map would make a very poor blueprint were our aim to construct the underground system, but it is very effective at helping us understand how to switch between lines and know where to get off. A blueprint of an underground line could serve as a map of that line after it is built, but it would contain a great deal of unnecessary detail that would make it cumbersome for that purpose.

When it comes to the blueprints for institutions, the "model" I have in mind is an assembly guide that provides instructions for how the institution should be "built". A vivid example comes from an article in the Guardian newspaper by economist Jeffrey Sachs, headlined *Amid the rubble of global finance, a blueprint for Bretton Woods II* (Sachs 2008), and there are many other such blueprints including proposed changes to the way markets for derivatives work after the financial crisis (Financial Crisis Inquiry Commission 2011). Sachs argued for a regulatory overhaul of global financial markets, bringing them under control of institutions like the International Monetary Fund. New regulations would consider the impact of financial markets on climate change, as well as managing their systemic risk and the behaviour of traders and investors in such markets, all intended to "steer the world to safety from the dire threats we face" (Sachs 2008). Such blueprints can be a rough sketch of the proposed institutional structure, such as Sachs' article, but can also include substantial detail, tightly constraining the institution to be built.

What should be clear from this preliminary discussion of blueprints is that they are very different things from theory models. Blueprints are plans for institutions, whereas theory models are usually devices employed in theories about the nature of the world, where they serve to illuminate and sometimes support the claims of such theories.

3.3 Direction of fit

One useful way of separating the categories of models between those used in the theory stance and the design stance, is the direction of fit. The idea of direction of fit comes from Elizabeth Anscombe (1957) in making sense of the relation of thought to the outside world. Mark Platts explains it by distinguishing between beliefs and desires as follows (De Bretton Platts 1997):

> The distinction is in terms of the direction of fit of mental states to the world. Beliefs aim at being true, and their being true is their fitting the world; falsity is a decisive failing in a belief, and false beliefs should be discarded; beliefs should be changed to fit with the world, not vice versa. Desires aim at realisation, and their realisation is the world fitting with them; the fact that the indicative content of a desire is not realised in the world is not yet a failing in the desire, and not yet any reason to discard the desire; the world, crudely, should be changed to fit with our desires, not vice versa.

This provides a good way of thinking about models in the two stances. Platt's beliefs have a relationship to the world that is like a theory model's. A theory model is a claim about a true fact in the world. It is true if it fits the world in the right way by accurately representing something that is true in the world, serving to explain some sort of causal process, or to make predictions. If it has no explanatory or predictive power about what is in the world, then it should be discarded. In contrast, a model in the design stance works like Platt's desires. They aim to change the world, to construct something represented by the blueprint. The direction of fit is from the model to the world.

This idea of direction of fit is clear in the relationship between theory and engineering in the physical world. For example, the philosopher Robin Findlay Hendry distinguishes between “abstract” and “concrete” models when talking about chemical engineering (2014: 221):

> …one might think of a ‘pure’ scientific model as an abstract mathematical object, which is developed as a representation of some part of the world. The model is amended to fit the world, not the other way round. In contrast, engineering disciplines seek to change the world, not just to understand it. Process development in chemical engineering involves not only the construction and refinement of a mathematical model, but the construction and refinement of a concrete model production process, which is then scaled up to a full production process.

The abstract mathematical objects that Hendry describes as “pure” scientific models are representations of things that are in the world, so the direction of fit is from the world to the model. Engineering, however, is about getting the world to fit the model. Hendry makes clear that this can start with a “model production process”, a kind of reduced scale prototype of the ultimate process to be engineered, that is then scaled up.

3.4 Scale models

Generally we think of scale models as useful in the physical world, when designs for bridges, aeroplanes and countless other engineered objects are tested. But there are also important uses for scale models to social scientists and social engineers. Scale models are useful in both the theory stance and the design stance, but the logic differs in each stance and their use is quite rare in the theory stance. There is a sense in which experiments of designs for potential financial institutions can be seen as scale models, which will be focused on in section 5. Here I set out the general case for the epistemic value of scale models to enrich my argument that the epistemic value of theory models is different to what we should require for institutional plans.

Consider first the scale model in the design stance. The step of starting with a scale model of a process allows it to be tested to ensure that it performs as expected. Such scale models and prototypes are, in a sense, *simulations*. They take a design for a target object and simulate it by creating a version of it. This provides a step from the abstract to the real, an intermediate point at which the design, represented abstractly through a blueprint, is made more concrete. The objective of such experiments is to assess whether a design that is depicted in the abstract will perform as expected when it is made concrete (I will say more about experiments in general in section 5, below). For example, scale models of bridges are routinely tested in wind tunnels to see how they will perform in different wind conditions. Wind tunnel testing of bridge designs has been mandatory in the US since the collapse of the Tacoma Narrows Bridge in 1940 shortly after it was opened. The bridge was then the third-longest suspension bridge in the world, but swayed violently in heavy winds, eventually collapsing, because its designers had not considered the resonance of the bridge in strong wind conditions (Miller 2001). Such tests work because the wind tunnel and scale model introduce various properties that are not in the abstract world of the blueprint, properties which may simply be too complex to capture in any other way, or ones that designers simply neglected to think about. The important role of the scale model is that it acts as a simulation of the device that will be engineered and the wider environment it will function in.

Scale models are also used in the theory stance, though their application is much more limited, particularly in social science. A model of the solar system, such as a classroom mobile hung from the ceiling, is a very rough scale model[9]. It provides a model of the solar system that allows users to see the order of the planets and (depending on the model) the shape of their orbits. It is an alternative means of representation to mathematical models or models based on two dimensional drawings.

In economics, there are some examples of scale models being used, though most models are mathematical or based on diagrams. Bill Phillips created the Phillips Machine in the early 1950s to illustrate the macroeconomy. Using a series of pipes and chambers and coloured water, it showed the flow of money in Hicks's interpretation of Keynes's General Theory (Moghadam and Carter 1989). It was primarily used as a teaching device, but it also worked as a calculative device to determine how changes to national income, interest rates and exchange rates affected the economy. It did, however, exhibit unintended properties, such as occasional leaks in which red water would spill to the floor, much to the amusement of students who were witnessing the demonstration (Morgan and Boumans 2004). Apparently some professors tried to explain the leaks as a valid analogy to the leak of money into the illicit economy, but these were surely *ad hoc* and useless for determining any

[9] It is very rough indeed. If we were to maintain proportions, as a "true" scale model does, the solar system model would use up considerable space. If the earth was the size of a 2.3cm wide marble, the sun would have to be 2.5m across. If we were representing the distance between them accurately, they'd have to be spaced 270 metres apart (Scherrer 2016).

quantities, given that the leaks were an unintended property of the model. The physical form of the Phillips Machine was, however, not essential and because of uncontrollable properties like leaks, other forms were preferable. The Hicks model could also be represented using a set of differential equations that could more reliably calculate the model's outcomes, which was surely preferable if the objective is to determine the value of variables in the model rather than illustrate the model to students. The example shows that scale models can help make models understandable in a pragmatic sense, but are likely to be less accurate than a mathematical equivalent when it comes to measuring variables. Sometimes a scale model can be useful to investigate theoretical implications of mechanisms and inputs. Indeed, Phillips did learn new relationships through the construction of his model, leading to the famous Phillips Curve that showed the relationship between inflation and unemployment (Morgan 2012). But this was because the physical model allowed him to reason through the deductive consequences of the starting conditions and mechanisms of the model, rather than because it introduced anything new to the theoretical world of the model. It should be clear then that the logic of such scale models is the same as other theoretical models of economics in that they support deductive reasoning from starting parameters and mechanisms.

In contrast, the reasoning supported by scale models of the design stance is not deductive. It is inductive, "seeing if it works" by testing out a version of the target object. Scale models allow for the introduction of real world complexities to the system being tested. The Phillips Machine did face real world complexities, such as the tendency to leak, but this in no way provided insight into whether the claims of the Hicks model were true of the real world. It was merely a failure of the model. However, a bridge that collapses in a wind tunnel might be the successful discovery of unexpected properties of the thing being designed. This isn't always true because scale models can contain features, most obviously a reduced size, that make them unlike the target object. A failure can occur because of the difference between the scale model and the target object, rather than because of the similarities. Reasoning through which of these is the case in the event of a failure is a crucial part of reasoning with scale models.

The deductive nature of a scale model such as the Phillips Machine points to a general point about scale models used in the theory stance: when they work as intended, they are closed systems that lead to some outcome with deductive necessity. It follows that a scale model in the theory stance should contain *no properties that are not implied by the model a priori.* Sometimes a scale model can have unintended properties such as leaks, but this represents a failure of the analogy of the scale model and the theory model. A scale model in the design stance is quite different in that it deliberately *contains properties that are absent from the design.* This follows from the fact that a scale model is built from a blueprint, but uses materials and operates in an environment with properties that cannot be fully anticipated by the designer. The additional properties introduced by the materials and environment are useful in order to test how the design might function in the real world. The point of such a scale model is that it is a step closer to the real world from the abstract world of the blueprint. This allows us to test for unexpected properties of the design that may emerge as a result of the properties and features of the materials and environment. This would be a useful result of the model in the design stance. It follows that complexity is generally a virtue for a a model in the design stance, but simplicity is generally a virtue for a model in the theory stance. This is because models in the design stance – blueprints and scale models – need to be tested for unexpected properties that outside the "world of the model", whereas models in the theory stance are a complete world. The scale model in the design stance takes the abstract model of the blueprint and introduces complexity by creating a version in the world. When models in the theory stance are illustrated using scale models, they are just as closed as their abstract versions.

Can we use scale models in social engineering of financial institutions the way we do in engineering bridges or aircraft? The answer is yes, though the issues are more complex. The experiments used in the testing tradition discussed by Guala (2007), such as small-scale market experiments, have a similar logic to that of scale models in physical engineering. An auction can be modelled perfectly to yield optimal outcomes in a computer simulation, but then encounter unexpected issues when created in reality. Guala considers an auction for telecommunications licences where it was extremely difficult to control the privacy of the bidders. He quotes Richard Cramton, an economist who assisted one of the bidding teams (Cramton 1995: 287 quoted in Guala 2007: 148):

> It was common for a bidder that did not need to bid, because it was the current high bidder, to pretend to place a bid, so as to conceal its identity. These pretend bids were not always successful before round 18, because a bidder could not ask for written confirmation of the pretend bid. Almost all bidders asked for written confirmation for their bids. To get a written confirmation, the bid assistant would have to walk across the room in public view. In round 18, the FCC announced, "Beginning with this round, you may go into the bidding booth and request from the bidding assistant a confirmation of your actions regardless of whether you bid, exercise a proactive waiver, or do not submit a bid." Even this met with limited success, since the sheet on which the written confirmation was printed was folded differently depending on whether it was a real bid or a fake bid.

This illustrates how the computer simulation was unable to capture aspects of the real auction, such as the challenges involved in maintaining the confidentiality of bids in practice (I will say more about the role of computer simulations in general in the next section). For the social engineer in the design stance, such detail is important, whereas in the theory stance it is not. In the theory stance, the model of the auction works as an idealisation that deductively supports a claimed tendency. While we may see the model in the theory stance as a thought experiment, it is an experiment that tests a simplified world for deductive implications. In contrast, experiments in the design stance test out the applicability of the model to the world, dealing with confounding factors by developing the institution to manage them where possible. Institutional plans are tested to find out if they are likely to work "in the wild". Miller (2001) argues that scale models of financial markets can be created and studied in a laboratory setting in order to test for unexpected properties of an institutional design. As I will discuss in section 5 below, it is far more complex to assess whether such experiments in finance work as scale models in the same way that scale models work in the physical sciences. They do introduce properties that are absent from the theoretical world, but these may not be properties shared by the target object, so the epistemic value of such simulations is far more difficult to determine. In the auction system described in the extract above, there were clear learnings from the functioning of the auction. It is not clear that any experiment would have revealed such properties in advance.

The abstract models of the two stances, theory models and blueprints, are distinct. As we have already seen, theory models in economics are devices of deductive reasoning that show how a particular outcome follows from a set of starting conditions and mechanisms. In contrast, a blueprint provides no comparable deductive argument. It is a guide to construction. Here the direction of fit discussed above is the decisive feature that separates these types of models. A blueprint does not make any claim about something being true in the world. Its validity is not determined by whether it corresponds to some true feature of the world. Rather, it provides a guide to change the world, to make something true of the world. The direction of fit underpins the fundamental difference – a theory model is an argument, while a blueprint is not.

In summary, models of the theory and design stances are fundamentally different in the following respects:

	Theory stance	**Design stance**

Target object	General properties of financial phenomena	Specific institutions that are proposed to be created
Direction of fit	World to model	Model to world
Abstract models	Theory models - present a deductive argument about processes that are claimed to exist in the world	Blueprints - contain no argument, but guide construction of a thing in the world
Scale models	Illustrate the world of the model and support deductive reasoning about consequences	Devices for experiments that test for properties not contained in the model

3.5 Computer simulations

I mentioned the use of computer simulations of auctions above as part of the process of designing an auction system. It is worth considering what role in general computer simulations might play in the design stance.

The role I have described for scale models is a type of simulation in that a (small) version of the object being designed is created and tested. Similarly, an auction system can be modelled on a computer and tested by running simulations of it. Indeed, the role played by scale models and wind tunnels have been somewhat superseded by computational fluid dynamics, which uses computers to deal with the complex mathematics of wind dynamics rather than scale models (Baker 2007). The growth in computing power has led to computer-based simulations where complex mathematical properties can be computed without the need for scale models (Frigg and Reiss 2009). Some have argued that computer simulation is a fundamentally different method in science (particularly Humphreys 2004), but such simulations don't appear to introduce any new issues than those already considered in respect of other idealised models. This is true in finance, which often uses Monte Carlo simulations to determine complex statistical outcomes from inputs, a technique used in many applications including the stress tests of banks. Such simulations can also include random variables in order to determine what effect they may have on the outputs.

But a computer simulation is like a theory model in that it is forever trapped by its inputs and mechanism specifications, even if some of those are specified as random. There is no content in the simulation that is not fully determined by its inputs and starting conditions. A scale model, therefore, still plays a role in the design stance that a computer simulation cannot substitute for, because, as in the example of the Tahoma Bridge discussed above, engineers may simply neglect to test for some or other feature. That's why today, despite the development of computational wind dynamics, in wind engineering, scale models and wind tunnels still play a critical role. Computers are important in scale model testing, being used to help make sense of the data from such tests as well as to calibrate the scale model so that it has features that are appropriately similar to the target object (Baker 2007), but the scale model still performs its function of introducing properties that are absent from the model. Computer simulations are fundamentally the same as simple abstract theoretical models in that deductively necessary conclusions follow from the initial conditions, with the only difference being that computers are needed to solve the mathematics in determining the possible outcomes. There will inevitably still be possible properties of the target object that such simulations would have neglected to consider. Scale models provide a further level of investigation, and one that is not constrained by what the researcher may think to include in the simulation.

4. Are MVO and CAPM models of the theory stance or the design stance?

The Mean Variance Optimisation framework and Capital Asset Pricing Model work as simplified models to make positive claims about the world and provide normative guidance to decision-makers, functions that make the models useful in the theory stance. However, they have also been used as blueprints, with financial institutions created to fit models. The latter role is problematic, principally because we have no grounds to believe the models are good blueprints, as opposed to being good at providing normative or positive insights about the world. I argue that we confuse supportive evidence for models as theory, with supportive evidence for models as blueprints. These roles require different kinds of evidence.

In what follows I disentangle the different impact that CAPM and MVO have, particularly to separate theory performativity and design performativity. The former reflects the fairly mundane role of theory in providing lessons that are then used in institutional design. The latter reflects the more problematic role of using models as blueprints for institutions.

4.1 MVO and CAPM as theory models

As discussed in Chapter 2, MVO provided a model that showed why portfolios should be diversified, rather than placing all of your investments into whichever security was anticipated to deliver the highest return (Markowitz 1952). While portfolio managers had always believed in not “putting all of your eggs in one basket”, until MVO this was folk wisdom rather than a scientific claim. MVO reduced the complex financial world to a simple one with few features: the expected return of all available assets, the variance of those returns, and the correlations between them. In that simplified model world, and given the assumption of risk aversion, and that volatility was equivalent to risk, the result is that decision makers would choose “efficient” portfolios that do not take on unnecessary risk for a given amount of return.

As quoted in Chapter 2, Markowitz initially described his work as a positive description of investment behaviour as well as a normative device: “Various reasons recommend the use of the expected return-variance of return rule, both as a hypothesis to explain well-established investment behaviour and as a maxim to guide one’s own action” (Markowitz 1952: 87). Markowitz was steeped in the neoclassical economics of the University of Chicago and took it for granted that economic decision makers were rational. Given that foundation, it was simple to see MVO as a positive description of the decisions investors were in fact making. In this sense, then, Markowitz’s model had a world-to-model direction of fit. Markowitz did not think that decision makers were literally conducting the statistical analysis contained in MVO, but that the model provided a good explanation behaviour that would be rational.

The CAPM was developed from MVO by assuming that the maximally possible diversified portfolio, one containing the universe of investible assets, was the most efficient. At the level of theory, this implied that the decision to add any asset to a portfolio was purely a function of whether the expected returns from that asset were sufficient to compensate for the risk that asset contributed to the portfolio, given a risk appetite. Returns thus exist in equilibrium with risk, measured as variance of returns. As a theoretical model this had clear positive implications. It implies that highly volatile assets are also those that will deliver the highest returns over the long run. The simple model world provided this as an unambiguous positive prediction. The extensive empirical testing of the CAPM led by Fama and French has focused on corroborating or falsifying this result with mixed outcomes (Fama and French 1992, 2004). However, the CAPM also has clear normative implications, in providing a way to calculate what return should be expected in order to justify an investment decision.

The normative implications of MVO and CAPM have a model-to-world direction of fit at a microeconomic behavioural level, counselling the individual investor on their best strategy. I described this normative model-to-world fit in the previous chapter as a part of theory performativity, as a consequence of individuals learning certain behaviours and believing they will help to achieve certain goals. This is different to the model-to-world fit of design performativity, where the world of the model is imposed in the real world through institutional engineering. Such a macro-level fitting of the world to the model involves creating the institutional structures that lead to the anticipated outcome. This differentiation of micro and macro levels of model-to-world fit is subtle but important. Individual decision rules do not treat the model as a blueprint. Instead the model works as a calculative device. Individual decisions can be made on the basis of the mathematics of MVO and CAPM. In contrast, design performativity is achieved when the model forms a blueprint for the structure of particular institutions.

4.2 CAPM and MVO as performative

I will argue that there are three distinct ways that modern portfolio theory, particularly the exemplar models of theoretical finance, can affect the design of financial institutions:

1. Positive learning from theory models which motivates certain behaviours, a form of theory performativity. I will discuss the development of using "risk adjusted returns" as a measure of investment success, inspired by the lessons of MVO and CAPM, as an example. Institutional engineers can draw on this learning directly, while such social learning can form background conditions that allow some institutions to function;
2. Normative behaviour as individuals use models as calculative devices to make optimal decisions to achieve certain goals. The financial advisory industry uses modern portfolio theory as the basis for advising clients how

to structure their investment portfolios while, for a period at least, traders used the Black-Scholes model to calculate derivatives prices;

3. Models serving as blueprints to design institutions so that they work just like the model world. I will discuss the development of index tracking funds, and the use of CAPM as a device to determine regulated prices, as examples.

I will discuss these examples, below. I will argue that the third type of effect is problematic.

4.2.1 Theory performativity

First, consider market behaviour that emerged through a process of theory performativity. The CAPM and MVO have had a dramatic impact on the way investors think about their portfolios and the expectations they have regarding professional fund managers. The MVO implied that the optimal investments, those that fall on the efficient frontier, were ones that minimised risk *and* maximised returns. It followed that investment returns are not the only matter to consider when assessing the performance of portfolios. We should expect that those portfolios with the highest risk are also the ones with the highest returns, but high risk portfolios can also have highly negative returns, especially in the short run. A fund manager who simply pursues the highest risk strategy would not be demonstrating skill when they subsequently come out with the highest return performance but might instead be demonstrating recklessness.

This idea was a natural consequence of MVO and was given formal expression by William Sharpe (1975, 1966) and what has come to be known as the Sharpe ratio. This calculates "excess" returns per unit of risk, where excess return is understood as the return over and above the risk free rate, and risk is understood as the standard deviation of the portfolio. Returns that are high and stable will have a better Sharpe ratio than returns that are high and volatile. The Sharpe ratio is one of several that can be used to examine returns from a risk-adjusted standpoint which are now commonly used by investment advisors to assess investment managers (Feibel 2003). These formal methods have their theoretical grounding in MVO and CAPM, but the general idea that returns are just as important as risk has become mainstream across the investment markets. This is the result of positive theory performativity as individuals learnt from the theories that returns are not all there is to investing and that risk mattered. It also led to normative theory performativity as investment decision makers used the models, and the derivative methods for calculating risk-adjusted returns, in order to assess professional fund managers and make decisions about which to use.

Some of the implications of the positive claims of modern portfolio theory were not comfortable for financial markets professionals in the 1950s and 1960s. The Efficient Markets Hypothesis, proposed by Eugene Fama (1970), holds that securities prices reflect all available information at any time, which suggests that there is no benefit to be had in researching stocks and attempting to learn more about the market. Financial analysts saw their task, in part, as the job of identifying mispriced securities, using the tools provided by Benjamin Graham and David Dodd in their influential book *Security Analysis* (Graham and Dodd 1962 [1934]). Such mispricing arose when the "intrinsic" value of the stock, as determined from techniques like discounted cash flow analysis, differed from the price. But modern portfolio theory implied this was a fool's errand as the price properly discounted all available information.

The EMH and random walk theories also suggested a lot of randomness in day-to-day securities price movements, which challenged the "chartists" or "technical analysts" who used graphical representations of price movements in an effort to anticipate future movements (MacKenzie 2006). But as MacKenzie outlines, the resistance from the investment profession waned over time and the implications of the models began to be accepted particularly during the 1970s. Thus the arguments that modern portfolio theory put forward about the nature of market prices began to win over its audience.

4.2.2 Theory performativity as a background condition

This acceptance of the CAPM by market participants can be seen as a background fact that allowed for the design of new institutions. Acceptance of such theory can be a necessary causal condition for certain institutions to be possible. For example, widespread understanding of the CAPM was a necessary precondition for a type of institutional infrastructure called index tracking funds.

The CAPM worked on the basis that the market portfolio was the maximally diverse portfolio and was also “efficient”. All assets could be assessed on the basis of whether they contribute or lessen the risk of the portfolio, with the “required return” determined as the necessary equilibrium condition to justify the purchase decision. The beta of a specific stock was a function of its covariance with the market as a whole, and formed the coefficient to the equity risk premium, so investors could determine how much return they should demand to buy the stock. At the time, however, the market portfolio was an abstract concept. There were market indices, but assembling a portfolio based on such an index would require buying every constituent investment. Index funds solved the problem by creating a form of fund that made the idea of investing in the market as a whole feasible.

The first index fund was launched by Wells Fargo in 1971, the design of which was helped by Sharpe, among others (Bernstein 1993). Such index funds took the lessons of CAPM to heart, replacing active stock picking in favour of assembling a portfolio of stocks that mechanically matched an index. This required technical innovations in order to allow a fund to mechanically buy and sell instruments so as to track an index dynamically. While the theory performativity drove the market place to understand the value of index-tracking funds, the institutions still had to be designed in order to be engineered into reality. The theory provides normative grounds for why such products might be valuable, but it is not a blueprint for the design of such products. The background models remain in a state of world-to-model fit.

Of course, the theory performativity was not all encompassing. Index funds were met with derision when first introduced, with some professional investors claiming it was absurd to blindly invest in stocks that could well be losers (MacKenzie 2006). But even if outperformance were possible by conducting exhaustive analysis, which some financial theorists argued was still theoretically possible, it is still arguable whether the outperformance would cover the cost of such analysis. Gradually market participants, schooled in the CAPM, accepted the value of passive instruments. The use of index tracking has expanded rapidly since, and now represents a major component of the total investment market.

Does the rise of index investing represent theory performativity or design performativity? The early work of Markowitz and Sharpe was largely undertaken in the theory stance. The normative implications were clear for investment decision makers, so we should not be surprised that the theories affected those decisions. As a result of CAPM, investors became much more concerned about the risk in their portfolios, measured by beta, and not just the returns. The idea of "risk-adjusted" returns became prominent – a portfolio which outperformed, but did so only by assuming a much higher beta, no longer won the plaudits it would have done before. This approach to measuring investment performance directly followed from the positive implications of CAPM (MacKenzie 2006).

These impacts of CAPM are examples of theory performativity. To the extent that the model changes the world, it is as a result of its role as a component in society, a background fact that ensures people understand certain concepts. Individuals came to accept that the world worked as the theory described and so stimulated demand for passive index funds. Switching to the design stance, this theory performativity provides the background environment in which new institutions can be designed.

But these models also became a form of blueprint serving as designs for the new passive instruments.

4.2.3 MVO and CAPM as blueprints

But can we use the theory models as blueprints directly? The answer is yes, in three distinct ways: overt blueprinting of institutions; engineering of idealising assumptions into institutions; and market participants' engineering of model outputs into reality. In this subsection I will explicate the first two of these, which is the use of models as blueprints by institutional designers. In the next section I will consider how market participants can manipulate the world in order to comply formally with a model while still pursuing different ends. Both are a form of design performativity that I will claim is dangerous and led to fundamental weaknesses in financial institutions.

In the first case, the CAPM can be literally made true by imposing it on the world. Stewart Myers argued that it could be used to determine what prices were reasonable for a regulated entity (Myers 1972). Subsequently, governments in the United States and United Kingdom appointed bodies to regulate prices that utilities can charge and therefore the rates of profit they can earn, where the CAPM has been used as the model (MacKenzie 2006). In doing so, the CAPM is playing a model-to-world direction of fit. The regulators determine an equity risk premium, the appropriate beta as an indication of the riskiness of the particular business model, and the risk-free rate, and back out the required return. However, instead of this being the return the investor should insist on in seeking out securities to buy, it is turned around and imposed on the investor as the "fair" return they should be allowed to earn. The model is providing a blueprint for the relationship between risk and return which is being implemented in the world. In using CAPM for regulated pricing, the model is literally made true.

In the theory stance, the simplifications of the CAPM do not detract from its positive account of a feature of the world. When the CAPM is blueprinted in the way Myers advocates, the simplifications are ignored. Beta is treated as not just an abstract simplified representation of risk, but as risk in reality. What are the grounds for doing so? We know that the mathematics of MVO and CAPM works to prove that investors should choose diversified portfolios and should price assets according to the risk they represent. But that mathematics works by relying on idealisations, particularly of risk and return. When CAPM is used as a blueprint for institutions, the idealisations that work in the model world are assumed to work in the real world. We are proceeding as if risk really is captured by beta, financial returns fully specify return expectations, and post hoc data fully determines expectations. In the theory stance, this may be appropriate because it allows for the model to point to claimed features of the real world. But in the design stance the model is now being used as a blueprint, which implicitly assumes the idealisations are true. It is as if we have taken Newtonian physics and assumed its frictionless plains and point masses are true and engineered a design on that basis. We have not even tested a scale model.

When institutions show weaknesses that stem from the failure of reality to match idealised assumptions in the theory model, additional efforts can be made to attempt to make the idealisations true in the world. The Black-Scholes model is another piece of modern portfolio theory that has caused both theory and design performativity. Theory performativity flowed from the spread of the theory in the financial markets. As MacKenzie notes, “gradually, the financial markets changed in a way that made the Black-Scholes Merton model’s assumptions, which at first were grossly at variance with market conditions, more realistic” (MacKenzie 2006: 256). This was achieved in part because Black and Scholes circulated daily closing prices for options based on their model to all market participants in Chicago, so propagating the model as a calculative device for market participants to use.

However, Black-Scholes also supported institutional blueprinting in that it is often used by derivative market operators to manage credit risk, requiring market participants to post security or settle obligations at the end of each day as if the model's result was the correct value for the derivative. The model provided a blueprint for the value of derivatives that was "made true" by constructing the market in order to fit the model. It was taken as granted that risk was equivalent to variance and the institution was designed on that basis. The assumption was treated as part of the blueprint.

Another consequence of the blueprinting phenomenon is to try and engineer idealising assumptions into reality. Because an outcome follows deductively from a model's specifications, including its idealising assumptions, social engineers who consider the outcomes desirable can attempt to engineer the idealising assumptions into reality. For example, the basic Black-Scholes model assumes that markets are liquid with zero transaction costs. So market infrastructure was designed such that there are no transaction costs in trading derivatives (though the cost of trade is usually built into the price via the implied volatility or interest rate assumed by the seller).

Similarly, the Efficient Markets Hypothesis argues that prices reflect all available information. A critical assumption of the hypothesis is that new information is rapidly priced into markets. This assumption has led to institutional designs that try and facilitate the rapid distribution of information, on the basis that efficient markets are desirable. For example, the London Stock Exchange operates a "Regulatory News Service" (RNS) through which all listed firms are required to announce anything that may have a material impact on share prices. The RNS is designed such that all market participants receive news simultaneously (Cai, Keasey, and Short 2006). For the same reasons, most financial markets include strict prohibitions against trading on inside information, backed by criminal sanction. There are other normative considerations that motivate such prohibitions, including that it is unfair if market participants have different levels of access to information (Brudney 1979) though this sense of unfairness has been challenged as ill-founded because insider trading can promote economically useful activity (Carlton and Fischel 1983).

Unfairness may or may not be a good motivation for such information distribution mechanisms, but what is not challenged is that such distribution does drive market efficiency. This idea takes the EMH as a blueprint, and information distribution mechanisms become ways of "making true" the assumptions about information distribution that the EMH depends on for its deductive conclusions. Whether devices such as the RNS successfully achieves this standard of information dissemination is another question. The EMH model works in theory, but it may not work in practice. To assess whether it will work as a blueprint for institutions is a very different question to whether it explains something about pre-existing markets.

These examples all illustrate how the model can be used as a design to develop institutions. It takes a theory model and treats it as a design rather than a model that sheds lights on a pre-existing feature of the world. Governments effectively make it the case that risk and return are related to each other in exactly the way that CAPM suggests, or market participants all have access to the same level of information. It is left unsaid that CAPM and MVO idealise the world and contain ontological commitments to a world that generates stable probability distributions, among other idealising assumptions.

Imagine we were to take this approach in designing physical objects. We would take an idealised model in physics, such as one including frictionless planes and point masses, and use it as a blueprint for a real physical system. However, in the physical world, the impossibility of frictionless planes and point masses would probably render the exercise impossible – our designs would fail immediately. But in the social world, one can implement the idealised model by assuming that its idealisations are true, or can be made true by regulating them into existence. In section 5 I will make clear what the problems are in doing so.

4.2.4 CAPM and MVO as perverse design performativity

If we know an institution is built on the basis of a theory model, one response is to manipulate the world such that when the model is applied to it, we get an outcome that we desire. In such cases, the model forms a blueprint for institutional design, but only because market participants know that market oversight applies theories based on the model. Theory performativity at one level leads to design performativity at the level of market participants. In so doing, the participants are using the model as a blueprint.

For instance, agents in a market can effectively game the regulators by reshaping their activities so as to fit the institution that has been built from a theory model blueprint. This is best explained through examples, of which I will discuss two.

Central banks use Value at Risk (VaR) models to determine how much market risk particular banks face, which then makes those banks use VaR models to control their own risk. VaR models are an application of the theoretical approach motivated by MVO. Typically, correlations between assets are determined by examining the historic correlations between asset prices to determine a variance for a particular portfolio now, much like MVO (Danıelsson 2002). Roughly this works as follows: if a bank's portfolio of securities is split equally between three assets: bonds, equities and property, the variance of the portfolio can be determined by looking at the historic correlations between the three assets, even if the bank had not owned them historically. The probability of loss can then be determined at various confidence levels by extrapolating from the standard deviation and mean return and assuming a Gaussian distribution. So, for example, a bank can be told that it cannot hold a portfolio that implies a risk of daily loss greater than 5% of its capital, with a 99.9% confidence level. There was little effort made to positively confirm that VaR modelling successfully indicated the bankruptcy risk of banks. Danıelsson's 2002 paper argued that it was particularly poor at analysing risk in crises under the title "The Emperor has no Clothes: Limits to Risk Modelling", which rang particularly true when the financial crisis struck (Danıelsson 2002).

The model did, however, provide a blueprint of a sort. Banks tried to influence the volatility of assets in the market place by issuing more highly rated bonds such as collateralised debt obligations that were assumed to be less volatile, and selling these to themselves and other clients that needed to meet lower VaR targets (Treanor 2008). In this way, banks deliberately engineered the world in order to fit the regulatory limits imposed on them. The direction of fit was, at least in part, from model to the world, with the banks using the model as their blueprint to shape the characteristics of the financial markets.

Banks' efforts to fit the investment market to match VaR models represent a type of behaviour that sees market actors manipulating features of the market, such as the variance of asset prices, to fit the model in the way they desire. This need not be a problem. If the model is valid as a positive claim about the world, making the world such that the anticipated outcome (in this case, low market risk) materialises would be a useful application of the theory. It would be comparable to using interest rates to control inflation, having been convinced by Keynesian models. But when central banks used VaR models as the basis for managing risk, they created an environment in which banks used VaR models to shape investment instruments in order to meet regulatory requirements. The problem was, as Danıelsson had argued, that the VaR models were not about a real mechanism in the economy, so were invalid theory models. As we discovered when the financial crisis struck, the fact that banks all were in line with the VaR limits that had been imposed on them meant nothing when the reality of the risks in the system emerged. VaR was particularly problematic – as a theory model it was misguided, and as a blueprint for institutional risk management it was a disaster.

The second example is along the same lines. Derivative markets use models based on normal distributions of historic prices to set margins that investors must pay at market close every day to manage credit risk in the market (Longin 1994). Throuhg a days' trading, counterparties to derivatives end up with implied gains and losses to their positions. For example, if an investor holds an option to sell an asset on the strike date at a set price, then a fall in the spot price today implies that he is exposed to credit risk. So the market place requires the counterparty to the option to post additional margin to cover that credit risk. These are also based on calculations of the probability of loss based on historic volatility. But, if you wanted to avoid that additional margin requirement, you could simply manipulate the spot price ensuring that it did not close lower than the strike price. In certain thinly traded stocks this was quite possible. The market was engineered as if the theory model was a blueprint, but the blueprint was completely blind to the impact of price manipulation of thinly traded stocks. Market participants can and did manipulate their margin obligations by deliberating manipulating the prices of underlying stocks in order to lower margin requirements (Theobald 2010). Again, the world was made to fit the model, albeit in a frankly illegal way.

In such cases, the model functions as a blueprint for market actors to reshape the market, but only because regulators and other institutions have used the model as a valid basis for the design of the institution, be it a banking risk management system or the detail of derivatives markets. The example again illustrates that trying to impose a theory model onto a real institution is a fraught exercise.

In this section the two examples show how design performativity is achieved as actors effectively manipulate reality to get what they want while the institution still works according to the blueprint contained in the theory model. It shows how real behaviour can depart from the behaviour assumed in the model. It is an obvious consequence of using theory models for design rather than to make claims about certain features of the world in a positive theory stance. When we start shaping the world to fit our models, participants have an incentive to function in formal compliance with the institutions while busily pursuing ends that could undermine the whole system. Had we been able to simulate such institutions in a way analogous to a physical scale model, we might have detected these in advance. In the next section I will consider whether it is possible to do so.

5. The problem with blueprinting theory models

The methods by which we assess theories and models in finance are complex, as has been discussed throughout this thesis. Models like CAPM and MVO can be assessed as positive theories by how well they shed light on real phenomena in the financial system. They can also be assessed for their normative force by examining the outcomes of decisions that are motivated by the models. Such assessment can be based on coherence of predictions with data, the internal logical coherence of the models, the rhetorical success in communicating insight into financial markets, and so on, all of which is routine work by theorists in the theory stance. At least part of the success of the CAPM and MVO in establishing themselves as exemplar models of the research programme comes down to their success in such assessments.

The problem is that those standards of assessment are not appropriate for assessing blueprints. And if theory models are going to be used as blueprints, then their success as theory models is not relevant. In the design stance, there are different standards we should apply in assessing blueprints. In particular, we should have a much greater role for experiments to test institutional designs.

In both stances, tests and experiments can be useful, but for different reasons. In the theory stance, experiments provide evidence to support the tendencies or mechanisms that the model proposes exist in the world. Mostly, such experiments are about assessing data, determining whether the data matches the predictions of such models. In the design stance, experiments help to calibrate a design, detecting unexpected properties before the full institution is created. Experiments in the theory stance are primarily deductive, while in the design stance they are inductive. One puts the model into the wild, even a carefully controlled wild, just to see how it copes.

Reiss (2008) argues that most experiments in economics are about testing theory, which implies that much of the experimental work undertaken is done in the theory stance. The objective of such work is to show that outcomes predicted by models actually obtain. Such experiments include classroom market places that test out whether particular behaviours and outcomes predicted in a model do in fact occur. Such experiments are laboratory science and it is not obvious how we might use them to learn about how an institution might work in the real world. For that, we need an applied science.

Theory models allow for deductive thought experiments to understand consequences of certain inputs and mechanisms. Experiments involving such models are carefully set up so that the *ceteris paribus* conditions have analogous shielding conditions in the experiment. Classroom market places are, in a sense, scale models of real market places. But such scale models are more like the Phillips Machine discussed in Section 3.4 – they are models of the theory model, rather than models of the institution to be engineered. Such classroom market places assist in reasoning through the logical implications of the model or to illustrate the model in a different format. In contrast, scale models in the design stance are inductive, where the ambition is to see if a design works before implementing it at full scale.

The assumptions and *ceteris paribus* clauses that achieve the simple world of the theory model ensure that the world of the theory model is disanalogous to the real world in important ways. As I have emphasised in this thesis, the MVO and CAPM have a core ontological commitment to risk as equivalent to the volatility of market returns and the world being such that it generates stable probability distributions. This core allows both models to "reveal" certain truths, such as that investors are better off diversifying their portfolios. Yet the use of these models as blueprints has meant that this notion of risk has become entrenched in the financial markets, both through theory performativity as market actors have absorbed the language of the models, and through design performativity as institutions have built this measure of risk into their institutional structure. While the theory model of MVO "works" by showing how a diverse portfolio is desirable when we see risk in a certain way, it does not follow that we can think of banks as risky or not risky by seeing their portfolios in the same way.

Using the models of modern portfolio theory as blueprints has meant bank regulators use variance to manage market risk and derivatives markets use it to manage credit risk. However, this aspect of the theory models is fundamentally disanalogous to risk in the world. The financial system is vulnerable to all sorts of risks, such as the impact of natural disasters, terrorist attacks, managers defrauding institutions, insurance companies misselling policies, and many others. When we design the financial system on the basis that variance is a sufficient concept of risk, we are making a profound error. It is an idealisation that helps theory models to illustrate and explain certain tendencies, but, just like we can't design frictionless planes into reality, we cannot design variance as equivalent to risk.

When theory models are used as the basis for institutional structures, there usually has been little empirical investigation of whether such institutions so created *are capable* of behaving as expected in the model. What we should want is to be able to test the models in the same way we might test whether a bridge built from a blueprint is capable of behaving as expected in a wind tunnel. The aim should be to see how institutions behave "in the wild" after being created from the blueprint of a theory model, it is the inductive approach of "seeing if it works".

The problem is that the role of the scale model in physical engineering is very difficult to achieve in social engineering. There are experiments that have been used to support design work in economics. For example, Roth (2002) describes experiments testing different algorithms for worker clearinghouses to match workers to employers. Several different cities and institutions used different algorithms allowing for tests of which institutional structure was effective. This is a kind of randomised trial rather than a scale model, but it clearly is one way we might test designs that does have epistemic power.

Randomised trials work to some extent in financial markets. There is variation in different markets around the world, providing a natural randomised trial, although the fundamental structure of stock exchanges and other securities exchanges is remarkably consistent worldwide. That consistency is driven by the standardisation efforts of the International Monetary Fund and global regulation forums such as the Bank of International Settlements in Basel and the International Accounting Standards Board. Stringent regulation of securities exchanges has substantially narrowed the possibility of variation. But variation still exists, even if only in the background conditions such as the type of society in which they function, providing empirical insight into what features are still able to vary *despite* standardisation.

So if randomised trials are difficult to use in finance, can scale models work better? Miller (2001) argues extensively for further use of experimental economics in the design of financial markets. Together with Charles Plott and other researchers at Caltech in the 1980s, Miller developed several approaches to experimental set ups to test markets. However, it is questionable whether these experiments really work as scale models in the design stance, or work as illustrative models in the theory stance. For instance, one Caltech experiment sought to draw out the lemons effect[10] of differing quality goods in a market and the way sellers signal quality. The goods in question were athletic shoes, and the quality differentiator was the number of stripes on the shoe which sellers could add at a cost. Any seller, whether of quality or inferior shoes, could add stripes, but the consumer surplus was large enough to cover the cost of the stripes only if the purchased shoes really were superior quality. The point of the experiment was to see how long it would take for the market to settle on a certain number of stripes to act as a stable signal of quality. The experiment is

[10] This refers to the effect discussed in Akerlof's seminal paper *The Market for Lemons: Quality Uncertainty and the Market Mechanism* (Akerlof 1995) which showed how a market will collapse if there is information asymmetry between sellers and buyers, such as when sellers know whether their product is high quality (a "peach") or low quality (a "lemon") but buyers are unable to tell the difference.

analogous to the Spence model of education signalling which represents education choices as an effort to separate talented from ordinary students, with an equilibrium number of degrees reached where the cost of education meets the signalling benefit (Spence 1973). As Miller tells it, it could take some time for equilibrium to be reached in the Caltech experiment, sometimes longer than the three hours allotted. Some interventions accelerated the process, such as allowing for stripes to be drawn in different colour chalk, depending on whether it was a superior or inferior quality of shoe.

As discussed in the context of scale models and wind tunnels above, the success of a scale model depends on it being analogous to the target object in the right ways. It is important to understand what features of a scale model are important and which not, in order to interpret the results. For instance, some features of a model of a bridge are exactly analogous to the target object, such as wind resonance, and others approximately analogous, such as tensile strength. But the size of the model bridge is obviously disanalogous, and a failure born from the disanalogous features, rather than the analogous features, clearly should not be interpreted as implying that there is a flaw in the target object's design. Much work is required to interpret the results of such scale model tests in order to understand what they imply for the performance of the target object.

In market experiments, the process of understanding the analogous and disanalogous features is much more complicated, largely because the subjects in such experiments know that it is not a real institution and the subjects are not the same as the agents who would act in the envisaged institution. The Caltech signalling experiment involved students, who may be very different from real agents in a real market. The incentives they face are quite different to ones faced by real agents. For instance, where monetary incentives apply they are usually far smaller than those available in the real world, and other incentives, such as professional reputation, simply can't be replicated. The students may face different incentives, such as getting out of the classroom as quickly as possible, or impressing their peer group. It is hard to see such an experiment as useful in testing a real market design. Instead the Caltech experiments seem to play a role along the lines of the Phillips Machine – illustrating the model in action and exploring some of the results to learn more about the theoretical model. The interventions in the experiment are an attempt to force the experiment to deliver more model-like outcomes, rather than an effort to introduce properties that are not in the theoretical model. This isn't always the case, however. One "discovery" from the Caltech experiments was that the market reached equilibrium quickly if sellers were immediately identified as selling superior quality shoes or inferior quality shoes after the sale and before the next round. This insight provides a justification for using feedback mechanisms in online auction sites such as eBay, in which buyers can identify whether the goods were described accurately, so calibrating buyers' understanding of the signals in the market place. This is arguably a reading of the experiment from the design stance, seeing it as a scale model, though the lesson is so general as to be questionable as a test of a specific institutional structure in a way analogous to a scale model.

It is difficult to create scale models or other experiments in the design stance when it comes to the financial system, though we should try. If we adopt a design stance approach to experiments, we might be able to shift the research questions, from ones that are set within the framework of theory models, to ones that deliberately try to step outside of them. Often the best scale models we have are actual implementations of institutions, perhaps on small scales where possible. Implementations on a small scale, such as a small stock exchange with only a few instruments listed, are more analogous to larger scale implementations than a classroom experiment. As Miller's experiments illustrate, learning from variations in actual institutions provides a way to achieve the same results as a scale model approach. Small scale classroom-based experiments of simulated markets might be pedagogically useful, and occasionally provide justifications for specific institutional interventions, but don't have the comparable epistemic strength as scale models in the physical sciences. But what I consider clear is that, despite the difficulties of such design-stance experimenting, using the epistemic tools of the theory stance is no substitute. The success of a model in the theory stance gives us no grounds to believe it will be successful as a design.

Theory models are capable of teaching us much about the world, learnings which we can take into consideration in developing our institutions. But if the institutions we design are to be robust in the real world, rather than the model world, we need to understand how the idealisations of the theory world differ from the real world. We need to design institutions that treat the world as it actually is, rather than trying to force it to behave like the world of the model. While testing our designs is difficult, we should try.

6. Conclusion

Key institutions in the financial system have been designed as if theory models are good blueprints. In many cases, designers went to lengths to try and force the simplifications of models into the world, such as regulations controlling the flow of information in stock exchanges or stripping out transaction costs in derivatives markets. In other cases, designers simply assumed that the simplifications of theory models were true, such as the equivalence of risk and variance. Testing and experiments of designs were not pursued, while the success of models in the theory stance was considered adequate reason to use the models as blueprints.

I have developed the following argument in this chapter. By separating out two "stances" of social scientists I have distinguished between theory models and blueprints as two different types of models with different epistemic functions. The exemplar models of theoretical finance work as theory models in that they are simplified models of financial markets that serve to explain, understand and describe tendencies that actually exist in the world. Such models are tested for their logical coherence and also, in complex ways, against data. Blueprints, while also simplified depictions of the world, play a quite different epistemic role as guides to the engineering of things to be built in the world. Those designs can be tested by the use of scale models and simulations, including computer simulations and other experiments such as randomised trials. Scale models introduce complexity by adding properties, such as those of the materials used in construction or the environment in which the design will be implemented, that are not capable of being included in a computer simulation. A computer simulation is bound by the starting conditions, being its programming and inputted parameters.

The theory models of theoretical finance are used as blueprints in the design of institutions. Governments can use them as the basis to dictate what returns are reasonable for companies. Market designers are motivated by them to implement certain institutional structures such as simultaneous information dissemination. Banks can use them to engineer financial markets to fit the models used by regulators. Such institutional arrangements impose the model on to the world, attempting to make the world like the model. Institutions are designed that treat financial risk as if it is nothing but market volatility, which gives agents an incentive to create instruments that allow them to formally comply, while in reality heightening risk on a wider interpretation. The problem is that the tests of theory models, such as their logical coherence and fit to the data, are not good tests of design. There is no test of theory models that introduces complexity in the way the test of a scale model does, precisely because theory models are shielded worlds. *The features that make a theory model useful are precisely the features that make it a poor blueprint*. This aspect of the reasoning employed in theoretical finance was underappreciated. While models were accepted within the theory stance, often with good supportive evidence, we assumed that was a good reason to accept them in the design stance as blueprints. But the tests we should apply to blueprints were seldom if ever undertaken.

While it is difficult, we should aim to test designs in ways analogous to scale models and through other experimental strategies such as randomised controlled trials. Moreover, we should learn from experience, specifically the implementation of institutions. Where possible, we should implement new designs on a small scale and test them before doing so on a much larger scale. Doing so may move us closer to a robust financial system more able to withstand risk, using a much more expansive notion of the concept. Certainly, the creation of financial institutions assuming the equivalence of variance and risk was an error.

Chapter 6

Conclusion: So What Now?

I will not summarise the arguments that I have made in this thesis here (section 6 of Chapter 1 did that), but in conclusion I will make some slightly speculative comments about the impact that these arguments might have on the practice of theoretical finance. As this is a work of applied philosophy, I'm concerned that the philosophy I have presented is continuous with practice. I will also point to some opportunities for further research.

1. A shifting conception of risk

My original view back in 2008, that the financial crisis would have a profound impact on financial theory was correct. However, the impact that I expected is not the one that has occurred. Theoretical finance, at least the form of modern portfolio theory, has not been cast on to the scrap heap of conjectures refuted by evidence. But the status of the research programme has been diminished.

One argument I have presented in this thesis is that the key models and theories in finance, like Mean Variance Optimisation, the Capital Asset Pricing Model and the Black-Scholes Model for option pricing, are useful, but in a narrow sense. They contribute to shedding light on tendencies and mechanisms that may be at work in an economy and its financial markets. In that respect, they sit alongside other useful theories, models and research programmes such as behavioural finance and multi-factor models of asset returns. They are also useful because they provide normative guidance for individuals; a way of thinking about what investments they should make. But rather than a decision rule that determines optimal investment decisions, they are best seen as guidance. They work by idealising the world into simple inputs and mechanisms in a way that shines a light on particular causal relationships, providing reasons to expect particular outcomes in equilibrium.

Despite this usefulness of modern portfolio theory, the research programme has taken reputational damage from the financial crisis. Arguably it has lost the status it once held as a highly reliable guide to what is in the world and what decisions we should make. It has been diminished to merely one guide among many. Given the crisis and the substantial losses experienced by investors and the public at large, trust in the models has been damaged. It is difficult to make any claim about what the right level of trust should be, as it differs for each instance of application to particular research questions, but it is not difficult to see that the level of trust placed in the models of modern portfolio theory was unjustifiable. Entire banking systems and capital markets were regulated and operated as if modern portfolio theory was an entirely reliable way of understanding financial markets.

That general point is, I think, increasingly widely accepted. It is having an appropriate impact on the way regulators work. For example, the Bank of England has shifted its focus in banking regulation from statistical analysis to the use of simple heuristics as indicators of risk, guided by the evolutionary psychologist Gert Gigerenzer (Aikman et al. 2014). The Bank is making an explicit effort to lessen its reliance on the frequentist concept of risk at the heart of portfolio theory and instead apply itself to Knightian uncertainty, the unknown unknowns, that may be present in financial systems. Aikman and his colleagues are eager to ensure that this has practical application (Aikman et al. 2014: 4):

> The central premise… is that the distinction between risk and uncertainty is crucial and has received far too little attention from the economics and finance professions to date. The shift from risk to uncertainty can turn what we think we know about decision making upside down. Decision rules that attempt to achieve ever greater precision can become increasingly

> imprecise; rules that attempt to weight optimally all the relevant information can sometimes generate poorer results than those based on simple averages or those that deliberately disregard information. Taking uncertainty seriously forces us to recognise that, in some circumstances, there are potential benefits to more simplicity over greater complexity.

This extract offers support for one aspect of the models of portfolio theory while taking it away from another aspect. It supports the positive interpretation of models as useful idealisations of the world, providing insight via their simplicity. It dismisses, however, the normative role of models as calculative devices; as decision rules that provide for precise outcomes. This echoes my claim in Chapter 5 that simplicity is a virtue and complexity is a vice when it comes to models to be used for institutional engineering. In place of the normative models of modern portfolio theory, the Bank is turning to the use of heuristics. So it has reduced, if not abandoned, its use of Value at Risk (VaR) modelling, the technique based on Markowitz-style analysis of correlations and variance, and instead developed a range of heuristics as indicators for the health of banks and the rest of the financial system. For instance, the Bank now focuses on banks' growth rates, because a bank that is growing too fast might be moving into increasingly risky activities in order to expand. This simple measure turns out to be a far better predictor of bank failures than complex VaR-based models.

Gigerenzer's approach can be contrasted with the bounded rationality school of behavioural economics pioneered by Herbert Simon (Simon 1959). Rather than dismissing human decision making as beset with biases and cognitive limits as Simon does, Gigerenzer emphasises that our decision-making tools are honed by evolution to be optimal in certain environments (Gigerenzer 2015). Rather than relying on the apparent certainty of the statistical measures of modern portfolio theory, Gigerenzer argues that we are better off using simple measures that are generated by easily understood mechanisms. If humans are given the right information in the right way then they are remarkably good at understanding risk and uncertainty.

Of course, by definition Knightian uncertainty is unknowable, so we cannot know whether a regulatory approach based on heuristics will provide long-term stability in ways that regulation based on frequentist risk has not. Gigerenzer argues that the heuristics approach is more valuable the less certain we are about the future, largely because it avoids the false sense of security that precise statistical frequencies appear to provide (Gigerenzer and Brighton 2009). Whatever the merits of that argument, which would take another thesis to assess, it is clear that modern portfolio theory, with its faith in volatility as an adequate conception of risk, has lost the status it held until 2008. The heuristics approach provides an alternative way of dealing with uncertainty in the world, and shows how we can think about risk in a way that escapes the conceptual framework that modern portfolio theory had instilled into the research programme.

This is part and parcel of the theory performativity I have identified and developed in this thesis. Modern portfolio theory guided thinking about how we should understand risk in the financial system. It influenced decision makers and gave them a set of tools. It provided a conceptual framework to use in communicating with each other and making sense of the world. In that respect it changed the world by causing market participants to behave in particular ways and created background facts that allowed for new financial innovations such as the tracker funds discussed in the last chapter. But institutions are fluid and new theories are always capable of becoming causally relevant to institutions, changing them as modern portfolio theory did once. New, or even old theories that gain new prominence, accumulate in the library of tools available to social scientists. I expect that the models of modern portfolio theory are going to be less frequently borrowed from that library.

The theory of social ontology I have argued for in this thesis underpins a vigorous, positive social science, one that is able to develop knowledge about the institutions of the financial system, knowledge about their properties, the organising structures that govern their components, and the causal background from which the institutions emerge. Financial theory can and should understand itself in this context, as a component within institutions that has a causal consequence itself.

The thesis I have presented, however, goes further than the role of theoretical finance as a social science. I have also focused on the role it has played in the engineering of institutions. In particular, I argued that its models were used as blueprints and not just idealised tools of reasoning and explanation, and that this led to weaknesses in institutions and our knowledge of them.

2. A new approach to design

A key purpose of this thesis was to bring to bring to light how models have directly shaped the institutions that existed (and often still do exist) within the financial system. If a model depicts a simple world in which certain desirable outcomes follow deductively from the mechanisms and inputs of the model, then we might try to make the actual world deliver those outcomes by ensuring that it works like the model. Models became blueprints to use in shaping institutions; the world was made to fit the model.

This approach to social engineering is fundamentally flawed. The deductive logic that applies in theory models works because of simplifications. The MVO framework depicts markets as highly stylised, with agents interested only in maximising returns and minimising risks, with those defined in narrow statistical terms. The real world is massively more complex. Theoretical finance deals with this complexity by employing idealisations that can be used in a simplified model to tell a story about the way the world works. These simplifications make it possible to point to specific theorised tendencies or mechanisms. Yet fundamentally the test of such models is whether they really do represent something true in the world. They are therefore assessed by how well they fit the world.

When such models are instead used as blueprints, the world must be forced to be more like the abstract model, tamed of its wildness so that the deductive neatness of modern portfolio theory could be applied to financial institutions. Regulators and practitioners employed all sorts of devices, from regulating information dissemination to specifying bank VaR limits, to make the world more like their models. They created central banks and banking systems that depended on risk being equivalent to volatility and tried to demand that it in fact was.

The financial crisis was a stark unravelling of this imposed design. Institutions decayed in ways that the conceptual framework of modern portfolio theory simply couldn't make sense of. The financial system, which had been warped to fit the models of modern portfolio theory, had no way of recognising risks that could not be understood in terms of historic volatility. There was no way to understand the complete collapse of a financial market, as occurred in the trading of collateralised debt obligations, for example. No probability distribution for market returns can possibly accommodate a collapsed market. A conception of risk that considers only volatility fundamentally depends on the existence of markets and the prices and price movements that they produce. Its internal metaphysics depends on the world being one where stable probability distributions exist. Central banks and many investment firms could not see the crisis coming because they were engineered according to the dictates of modern portfolio theory. When the true risks materialised, the result was system-wide collapse.

The analysis I have presented argues for an approach to institutional design that emphasises the use of tests and experiments. These are harder to perform than the otherwise analogous testing of designs in physical engineering. Scale models, understood as small-scale implementations of an institutional design, are difficult to use in social science because it is more difficult to understand whether failures of the scale model should be interpreted as failures of the design that would apply to the full-scale institution, or artefacts of the disanalogous features of the scale model. Such small scale implementations are not the only test – randomised controlled trials are another. More work has to be done, however, to develop methods of testing institutional designs before they are implemented.

While conceding that testing of institutions is hard in finance, the idea that we *should* test them is surely a healthy form of epistemic modesty. Designs of institutions in the financial system have too often been undertaken as if modern portfolio theory provides a definitive and sufficient way to understand the system. When reality has not conformed to expectations, we have intervened to try and make it more like the model, to tweak the causal mechanisms of institutions so that they are more like the socioeconomic machine captured in the model. Such an approach puts the theory on too rarefied a pedestal. It might be brought down to earth if we change the research objectives in theoretical finance from seeking out confirmation of models, to testing for failures of institutional design.

To some extent, the vision of design that I am advocating is similar to evidence-based policy making, an approach to public policy that has developed over the last two decades (see Nutley, Davies, and Smith (2000), Cartwright and Hardie (2012)). This is only to an extent, because evidence-based policy is focused on interventions in the public realm that are designed to promote the public good through public institutions. In the financial system, private financial institutions must capably manage the deontic obligations the members of a whole society have to each other, though public institutions like central banks, public transaction infrastructure, and the legal system are crucial. While an efficient and capable financial system is certainly a public good in a macro sense, it is the private obligations at a micro-level that have to be managed, often far from the direct agency of any public sector entity. Nevertheless, the literature on evidence-based policy provides some useful insights as to how to think about evidence in social engineering.

While randomised controlled trials are often considered the "gold standard" of public policy evidence, in fact there are many problems with extrapolating from a situation in which an institutional intervention or design is shown to work, to another (Cartwright and Hardie 2012). For example, a particular policing strategy that is successful in one town cannot be assumed to work in the next town, because it may depend on unidentified factors that are not consistent between the towns, such as employment levels. Similar considerations apply in the approach to financial institutions that I have advocated. What we need is a theory of evidence, a way of understanding what evidence should "count" in determining whether a particular institutional design will work the way we want it to. This applies to randomised trials as well as scale models, where much work has to be done to understand just what the performance of the model tells us about the design. While I have sketched out some ideas and problems with how we can use experiments in building financial institutions, developing a comprehensive theory of evidence is a clear opportunity for further research.

3. A final word

In the introduction to this thesis I wrote that consideration of the foundations of a discipline is an opportunity to enrich current practice. I hope that the arguments I have made make some small contribution to doing so. In particular, I hope that my account of how financial institutions are affected by reflexivity and the shifting properties of their components, will give practitioners and policy makers new ways of understanding the properties of financial institutions. By seeing them as emergent, I hope to enrich the understanding of how background conditions affect the properties and functioning of financial institutions, including the causal properties institutions should possess in order to improve their robustness.

The task of theorising about the financial world remains as important as ever. With the diminished status of portfolio theory, there is a clear need for new theories that can fill the gaps that cannot be explained by existing theory. I have argued that the notion of risk at the heart of modern portfolio theory is a distinct weakness, and we need to build theories that deal with risk and uncertainty in different ways. Such work will further knowledge of how the financial system works, but will also inform future institutional designs that might prove to be more robust. And that would make a better world.

Bibliography

Ahonen, Guy. 1990. 'A "Growth of Knowledge" Explanation to the Response to Chamberlin's Theory of Monopolistic Competition', *Research in the History of Economic Thought and Methodology*, 7: 87-103.

Aikman, David, Mirta Galesic, Gerd Gigerenzer, Sujit Kapadia, Konstantinos V Katsikopoulos, Amit Kothiyal, Emma Murphy, and Tobias Neumann. 2014. 'Taking Uncertainty Seriously: Simplicity Versus Complexity in Financial Regulation', *Bank of England Financial Stability Paper*. Bank of England: London.

Akerlof, George. 1970. 'The Market for “Lemons”: Quality Uncertainty and the Market Mechanism.'. *The Quarterley Journal of Economics.* 84.3: 488-500.

Anscombe, Gertrude Elizabeth Margaret. 1957. *Intention* (Harvard University Press, Cambridge MA).

Ashmore, Malcolm. 1989. *The Reflexive Thesis : Righting Sociology of Scientific Knowledge* (University of Chicago Press: Chicago).

Austin, J. L. 1962. *How To Do Things With Words* (Harvard University Press: Cambridge).

Backhouse, Roger E. 2012. 'The rise and fall of Popper and Lakatos in economics', *Philosophy of Economics (Handbook of the Philosophy of Science)*, 13: 25-48.

Banz, Rolf W. 1981. 'The Relationship Between Return and Market Value of Common Stocks', *Journal of Financial Economics*, 9: 3-18.

Basu, Sanjoy. 1977. 'Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis', *The Journal of Finance*, 32: 663-82.

Bedau, M., and P. Humphreys. 2008. *Emergence: Contemporary Readings in Philosophy and Science* (MIT Press: Cambridge, MA).

Bernstein, Peter L. 1993. *Capital Ideas: The Improbable Origins of Modern Wall Street* (Simon and Schuster: New York).

Bhandari, Laxmi Chand. 1988. 'Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence', *The Journal of Finance*, 43: 507-28.

Bhaskar, Roy. 2014. *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences* (Routledge).

Bianchi, Marina, and Hervé Moulin. 1991. 'Strategic Interactions in Economics: the Game Theoretic Alternative.' in Mark Blaug and Neil de Marchi (eds.), *Appraising Economic Theories* (Edward Elgar: Aldershot).

Binmore, Ken, and Paul Klemperer. 2002. 'The Biggest Auction Ever: the Sale of the British 3G Telecom Licences', *The Economic Journal*, 112: C74-C96.

Black, Fischer, Michael C Jensen, and Myron S Scholes. 1972. 'The Capital Asset Pricing Model: Some Empirical Tests.' in Michael C Jensen (ed.), *Studies in the Theory of Capital Markets* (Praeger Publishers: New York).

Blaug, Mark. 1992. *The Methodology of Economics: Or, How Economists Explain* (Cambridge University Press: Cambridge).

Blume, Marshall E, and Irwin Friend. 1973. 'A New Look at the Capital Asset Pricing Model', *The Journal of Finance*, 28: 19-34.

Bondt, Werner FM De, and Richard H Thaler. 1987. 'Further Evidence on Investor Overreaction and Stock Market Seasonality', *The Journal of Finance*, 42: 557-81.

Brentani, Christine. 2004. *Portfolio Management in Practice* (Elsevier Butterworth-Heinemann: Oxford).

Bridgman, Percy Williams. 1927. *The Logic of Modern Physics* (Macmillan: New York).

Brudney, Victor. 1979. 'Insiders, Outsiders, and Informational Advantages Under the Federal Securities Laws', *Harvard Law Review*, 93: 322-76.

Cai, Charlie X, Kevin Keasey, and Helen Short. 2006. 'Corporate Governance and Information Efficiency in Security Markets', *European Financial Management*, 12: 763-87.

Caldwell, Bruce J, and Terence Hutchison. 1984. *Appraisal and Criticism in Economics: A Book of Readings* (Allen & Unwin Boston: Boston).

Callon, Michel. 1998. *The Laws of the Markets* (Blackwell Publishers: Oxford).

———. 2007. 'What Does it Mean to Say That Economics is Performative?' in Donald A MacKenzie, Fabian Muniesa and Lucia Siu (eds.), *Do Economists Make Markets? On the Performativity of Economics* (Princeton University Press: New Jersey).

Carlton, Dennis W., and Daniel R. Fischel. 1983. 'The Regulation of Insider Trading', *Stanford Law Review*, 35: 857-95.

Cartwright, Nancy. 1995a. 'Causal Structures in Econometrics.' in Daniel Little (ed.), *On the Reliability of Economic Models* (Kluwer Academic Publishers: Dordrecht).

———. 1995b. '"Ceteris Paribus" Laws and Socio-Economic Machines', *The Monist*, 78: 276-94.

———. 1999. *The Dappled World: A Study of the Boundaries of Science* (Cambridge University Press: New York).

———. 2002. 'The Limits of Causal Order, From Economics to Physics.' in Uskali Mäki (ed.), *Fact and Fiction in Economics: Models, Realism and Social Construction* (Cambridge University Press: Cambridge).

Cartwright, Nancy, and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing it Better* (Oxford University Press: Oxford).

CFA Institute. 2016. 'CFA Program', https://www.cfainstitute.org/programs/cfaprogram/Pages/index.aspx CFA Institute, Accessed 9 July 2013.

Childers, Timothy. 2013. *Philosophy and Probability* (Oxford University Press: Oxford).

Coleman, Thomas S. 2011. *A Practical Guide to Risk Management* (CFA Institute: Charlottesville, VA).

Cramton, Peter C. 1995. 'Money Out of Thin Air: The Nationwide Narrowband PCS Auction', *Journal of Economics & Management Strategy*, 4: 267-343.

Danıelsson, Jón. 2002. 'The Emperor Has No Clothes: Limits to Risk Modelling', *Journal of Banking & Finance*, 26: 1273-96.

Davidson, Donald. 1963. 'Actions, Reasons, and Causes', *The Journal of Philosophy*, 60: 685-700.

De Bretton Platts, Mark. 1997. *Ways of Meaning: An Introduction to a Philosophy of Language* (MIT Press: Cambridge).

De Marchi, Neil. 1991. 'Introduction.' in Neil De Marchi and Mark Blaug (eds.), *Appraising Economic Theories : Studies in the Methodology of Research Programs* (Elgar: Aldershot, UK).

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. 2009. 'Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?', *Review of Financial Studies*, 22: 1915-53.

Dennett, Daniel. 1989. *The Intentional Stance* (MIT Press: London).

———. 1999. 'Faith in the Truth.' in Wes Willams (ed.), *The Values of Science: The Oxford Amnesty Lectures 1997* (Westview Press: Michigan).

Dietrich, Franz, and Christian List. 2016. 'Mentalism Versus Behaviourism in Economics: a Philosophy of Science Perspective', *Economics and Philosophy*, 32: 249-81.

Dodd, Nigel. 2016. *The Social Life of Money* (Princeton University Press, New Jersey).

Douglas, George Warren. 1969. 'Risk in the Equity Markets: An Empirical Appraisal of Market Efficiency', *Yale Economic Essays*, 9: 3-45.

Eatwell, John, Murray Milgate, and Peter K Newman. 1992. *The New Palgrave Dictionary of Money & Finance* (Macmillan Press Limited: London).

Engelen, Ewald. 2011. *After the Great Complacence: Financial Crisis and the Politics of Reform* (Oxford University Press: Oxford).

England, Bank of. 2014. 'Damaged and Mutilated Banknotes', Bank of England, Accessed 31 January 2015. http://www.bankofengland.co.uk/banknotes/Pages/damaged_/default.aspx.

Fabozzi, Frank J, Sergio M Focardi, and Caroline Jonas. 2010. *Investment Management After the Global Financial Crisis* (The Research Foundation of CFA Institute: Charlottesville VA).

Fabozzi, Frank J, Francis Gupta, and Harry M Markowitz. 2002. 'The Legacy of Modern Portfolio Theory', *The Journal of Investing*, 11: 7-22.

Fama, Eugene F. 1970. 'Efficient Capital Markets: A Review of Theory and Empirical Work', *The Journal of Finance*, 25: 383-417.

———. 1995. 'Random Walks in Stock Market Prices', *Financial Analysts Journal*, 51: 75-80.

Fama, Eugene F, and Kenneth R French. 1992. 'The Cross-Section of Expected Stock Returns', *Journal of Finance*, 47: 427-65.

———. 1993. 'Common Risk Factors in the Returns on Stocks and Bonds', *Journal of Financial Economics*, 33: 3-56.

———. 2004. 'The Capital Asset Pricing Model: Theory and Evidence', *Journal of Economic Perspectives*, 18: 25-46.

Feibel, Bruce J. 2003. *Investment Performance Measurement* (Wiley: Hoboken, NJ).

Ferraro, Fabrizio, Jeffrey Pfeffer, and Robert I Sutton. 2005. 'Economics Language and Assumptions: How Theories can Become Self-Fulfilling', *Academy of Management Review*, 30: 8-24.

Feyerabend, Paul. 1993. *Against Method* (Verso: New York).

Financial Crisis Inquiry Commission. 2011. *The Financial Crisis Inquiry Report: Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States* (Public Affairs: New York).

Fisher, Irving. 1928. *The Money Illusion* (Adelphi Company: New York).

Fitzpatrick, Dan. 2003. 'Searle and Collective Intentionality', *American Journal of Economics and Sociology*, 62: 45-66.

Frank, Robert H., Thomas Gilovich, and Dennis T. Regan. 1993. 'Does Studying Economics Inhibit Cooperation?', *Journal of Economic Perspectives*, 7: 159-71.

Friedman, Milton. 1953. 'The Methodology of Positive Economics.' in Milton Friedman (ed.), *Essays in positive economics* (University of Chicago Press: Chicago).

Frigg, Roman, and Stephan Hartmann. 2012. 'Models in Science'. In *Stanford Excyclopedia of Philosophy*. Accessed 19 June 2015. http://plato.stanford.edu/archives/fall2012/entries/models-science/.

Giere, Ronald N. 2010. 'An Agent-Based Conception of Models and Scientific Representation', *Synthese*, 172: 269-81.

Gigerenzer, Gerd. 2015. *Risk Savv: How to Make Good Decisions* (Penguin Books: London).

Gigerenzer, Gerd, and Henry Brighton. 2009. 'Homo Heuristicus: Why Biased Minds Make Better Inferences', *Topics in Cognitive Science*, 1: 107-43.

Gillies, Donald. 2000. *Philosophical Theories of Probability* (Routledge: London).

Glasner, David. 1989. *Free Banking and Monetary Reform* (Cambridge University Press: Cambridge).

Graham, Benjamin, and David Dodd. 1962 [1934]. *Security Analysis: Principles and Technique* (McGraw Hill: New York).

Gratton, Claude. 1996. 'What is an Infinite Regress Argument?', *Informal Logic*, 18.

Gray, Richard T. 1999. 'Buying Into Signs: Money and Semiosis in Eighteenth-Century German Language Theory.' in Martha Woodmansee and Mark Osteen (eds.), *The New Economic Criticism: Studies at the Intersection of Literature and Economics* (Psychology Press: London).

Grossman, Sanford J, and Joseph E Stiglitz. 1976. 'Information and Competitive Price Systems', *The American economic review*, 66: 246-53.

Guala, Francesco. 1999. *Economics and the Laboratory: Some Philosophical and Methodological Problems Facing Exerimental Economics*, PhD Thesis, London School of Economics and Political Science: London.

———. 2007. 'How to Do Things with Experimental Economics.' in Donald A MacKenzie, Fabian Muniesa and Lucia Siu (eds.), *Do Economists Make Markets?* (Princeton University Press: Oxford).

———. 2016. *Understanding Institutions: The Science and Philosophy of Living Together* (Princeton University Press: New Jersey).

Hacking, Ian. 1979. 'Imre Lakatos's Philosophy of Science', *British Journal for the Philosophy of Science*, 30: 381-402.

Hájek, Alan. 2012. 'Interpretations of Probability'. In *The Stanford Encyclopedia of Philosophy*. Accessed 15 April 2016. https://plato.stanford.edu/archives/win2012/entries/probability-interpret/

Hands, D Wade. 2001. *Reflection Without Rules: Economic Methodology and Contemporary Science Theory* (Cambridge University Press: Cambridge).

Hart, Keith. 2001. *Money In An Unequal World : Keith Hart and His Memory Bank* (Texere: London).

Haugen, Robert A. 1995. *The New Finance: The Case Against Efficient Markets* (Prentice Hall: New Jersey).

Hausman, Daniel M. 1994. 'Why Look Under the Hood.' in Daniel M Hausman (ed.), *The Philosophy of Economics: An Anthology* (Cambridge University Press: New York).

Hayek, Friedrich A. 1976. 'Denationalisation of Money', *London: Institute of Economic Affairs*: 20-35.

Hayek, Friedrich August. 1945. 'The use of Knowledge in Society', *The American Economic Review*, 35: 519-30.

Heckerman, David. 1998. 'A Tutorial on Learning With Bayesian Networks.' in Michael I Jordan (ed.), *Learning in Graphical Models* (Springer: Dordrecht).

Hendry, Robin Findlay. 2014. *Kinetics, Models, and Mechanism* (Walter de Gruyter: Berlin).

Holton, Glyn A. 2004. 'Defining Risk', *Financial Analysts Journal*, 60: 19-25.

Hoover, Kevin D. 2001. *Causality in Macroeconomics* (Cambridge University Press: Cambridge).

Horwitz, Steven. 1994. 'Complementary Non-Quantity Theory Approaches to Money: Hilferding’s Finance Capital and Free Banking Theory', *History of Political Economy*, 26: 221-38.

investing.com. 2016. 'investing.com', Accessed 29 August 2016. http://uk.investing.com/equities/sabmiller-historical-data.

Jeffrey, Richard C. 1992. *Probability and the Art of Judgment* (Cambridge University Press: Cambridge).

Johnstone, DJ. 2013. 'The CAPM Debate and the Logic and Philosophy of Finance', *Abacus*, 49: 1-6.

Kahneman, D, and A Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica.* XLVII, 263-291.

Keynes, John Maynard. 1921. *A Treatise on Probability.* (Macmillan: London)

Keynes, John Neville. 1904. *The Scope and Method of Political Economy* (Macmillan: London).

Kilby, Christopher. 2000. 'Supervision and Performance: The Case of World Bank Projects', *Journal of Development Economics*, 62: 233-59.

Knight, Frank H. 1921. *Risk, Uncertainty and Profit* (Hart, Schaffner and Marx: New York).

Koller, Tim, Marc Goedhart, and David Wessels. 2005. *Valuation: Measuring and Managing the Value of Companies* (Wiley: New Jersey).

Koopmans, Tjalling C. 1947. 'Measurement Without Theory', *The Review of Economic Statistics*, 29: 161-72.

Kuhn, Thomas S. 1962. *The Sructure of Scientific Revolutions* (University of Chicago Press: Chicago).

Lakatos, Imre. 1968. 'Criticism and the Methodology of Scientific Research Programmes', *Proceedings of the Aristotelian Society*, 69: 149-86.

———. 1970. 'Falsification and the Methodology of Scientific Research Programmes.' in Imre Lakatos and Alan Musgrave (eds.), *Criticism and the Growth of Knowledge* (Cambridge University Press: Cambridge).

———. 1978. 'History of Science and its Rational Reconstructions.' in Imre Lakatos (ed.), *The Methodology of Scientific Research Programs* (Cambridge University PRess: Cambridge).

Lakatos, Imre, John Worrall, and Gregory Currie. 1980. *The Methodology of Scientific Research Programmes: Volume 1: Philosophical Papers* (Cambridge university press: Cambridge).

Lakonishok, Josef, Andrei Shleifer, and Robert W Vishny. 1994. 'Contrarian Investment, Extrapolation, and Risk', *The Journal of Finance*, 49: 1541-78.

Latsis, John. 2015. 'Quine and the Ontological Turn in Economics.' in Stephen Pratten (ed.), *Social Ontology and Modern Economics* (Routledge: New York).

Latsis, Spiro J. 1972. 'Situational Determinism in Economics', *British Journal for the Philosophy of Science*: 207-45.

———. 1976. *Method and Appraisal in Economics* (Cambridge University Press: New York and Melbourne).

Lawson, Tony. 1996. *Economics and Reality* (Routledge: New York).

———. 2012. 'Ontology and the Study of Social Reality: Emergence, Organisation, Community, Power, Social Relations, Corporations, Artefacts and Money', *Cambridge Journal of Economics*, 36: 345-85.

———. 2015. 'A Conception of Social Ontology.' in Stephen Pratten (ed.), *Social Ontology and Modern Economics* (Routledge: London).

Levy, Haim. 2011. *The Capital Asset Pricing Model in the 21st Century: Analytical, Empirical, and Behavioral Perspectives* (Cambridge University Press: Cambridge).

Lewis, David. 1969. *Convention: A Philosophical Study* (Harvard University Press: Cambridge, MA).

———. 1980. 'A Subjectivist's Guide to Objective Chance.' in William L Harper, Robert Stalnaker and Glenn Pearce (eds.), *IFS: Conditionals, Belief, Decision, Chance and Time* (Springer: Dordrecht).

Lintner, John. 1965. 'The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets', *The review of Economics and Statistics*, 47: 13-37.

Loewer, Barry. 2004. 'David Lewis's Humean Theory of Objective Chance', *Philosophy of Science*, 71: 1115-25.

Longin, Francois M. 1994. 'Optimal Margin Levels in Futures Markets: A Parametric Extreme-Based Method', *London Business School Institute of Finance and Accounting Working Paper*, 192.

Lucas, Robert E. 1976. 'Econometric Policy Evaluation: A Critique.' in Karl Brunner and Allan H Meltzer (eds.), *The Phillips Curve and Labor Markets* (Carnegie-Rochester Conference Series on Public Policy: Amsterdam, North Holland).

MacKenzie, Donald A. 2008. *An Engine, Not a Camera: How Financial Models Shape Markets* (MIT Press: Cambridge, MA).

———. 2017. "A Material Sociology of Markets: the Case of 'Futures Lag' in High-Frequency Trading." In *Auguste Comte Memorial Lecture*. 27 February 2017. London School of Economics: London.

MacKenzie, Donald A, Fabian Muniesa, and Lucia Siu. 2007. *Do Economists Make Markets? On the Performativity of Economics* (Princeton University Press).

MacKinlay, A Craig. 1995. 'Multifactor Models do not Explain Deviations From the CAPM', *Journal of Financial Economics*, 38: 3-28.

Mäki, Uskali. 2001. *The Economic World View: Studies in the Ontology of Economics* (Cambridge University Press: Cambridge).

———. 2005. 'Models are Experiments, Experiments are Models', *Journal of Economic Methodology*, 12: 303-15.

———. 2013. 'Contested Modeling: The Case of Economics.' in Ulrich Gähde, Stephan Hartmann and Jörn Henning Wolf (eds.), *Models, Simulations, and the Reduction of Complexity* (De Gruyter: Berlin).

Manirakiza, Jean Guillaume. 2014. "The Role of Land Records in Support of Post-Conflict Land Administration: A Case Study of Rwanda in Gasabo District." In. Enschede, The Netherlands: University of Twente.

Markowitz, Harry. 1952. 'Portfolio Selection', *The Journal of Finance*, 7: 77-91.

———. 1959. *Portfolio Selection: Efficient Diversification of Investments* (Yale University Press: New Haven).

———. 1991. 'Foundations of Portfolio Theory', *The Journal of Finance*, 46: 469-77.

Markowitz, Harry M, G Peter Todd, and William F Sharpe. 2000. *Mean-Variance Analysis in Portfolio Choice and Capital Markets* (John Wiley & Sons: New Jersey).

Marwell, Gerald, and Ruth E. Ames. 1981. 'Economists Free Ride, Does Anyone Else?: Experiments on the Provision of Public Goods, IV', *Journal of Public Economics*, 15: 295-310.

Merton, Robert K. 1948. 'The Self-Fulfilling Prophecy', *The Antioch Review*, 8: 193-210.

Miller, Merton H. 1999. 'The History of Finance', *The Journal of Portfolio Management*, 25: 95-101.

Miller, Merton H, and Myron Scholes. 1972. 'Rates of Return in Relation to Risk: A Reexamination of Some Recent Findings.' in Michael C Jensen (ed.), *Studies in the theory of capital markets* (Praeger: New York).

Miller, Ross M. 2001. *Experimental economics: how we can build better financial markets* (John Wiley and Sons Limited: New Jersey).

Mises, Ludwig von. 1934. "The Theory of Money and Credit." (Cape: London).

Moghadam, Reza, and Colin Carter. 1989. 'The Restoration of the Phillips Machine: Pumping up the Economy', *Economic Affairs*, 10: 21-27.

Morgan, Mary S. 1992. *The History of Econometric Ideas* (Cambridge University Press: Cambridge).

———. 2012. *The World in the Model: How Economists Work and Think* (Cambridge University Press: Cambridge).

Morgan, Mary S, and Marcel Boumans. 2004. *Secrets Hidden by Two-Dimensionality: The Economy as a Hydraulic Machine* (Stanford University Press: New Haven).

Myers, Stewart C. 1972. 'The Application of Finance Theory to Public Utility Rate Cases', *The Bell Journal of Economics and Management Science*, 3: 58-97.

Noko, Joseph. 2011. 'Dollarization: The Case of Zimbabwe', *Cato Journal*, 31: 339-365

Nutley, Sandra M, Huw TO Davies, and Peter C Smith. 2000. *What Works? Evidence-Based Policy and Practice in Public Services* (MIT Press: Cambridge, MA).

Pearce, David William, Anil Markandya, and Edward Barbier. 1989. *Blueprint for a Green Economy* (Earthscan: London).

Pierre, Jacob. 2014. 'Intentionality'. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/win2014/entries/intentionality/. Accessed 14 May 2016.

Popper, Karl R. 1959. *The Logic of Scientific Discovery* (Hutchinson: London).

Pratten, Stephen. 2007. 'Ontological theorising and the assumptions issue in economics', *Contributions to Social Ontology*, 15: 50.

Quine, Willard V. 1948. 'On What There Is', *The Review of Metaphysics*, 2: 21-38.

Ramsey, Frank P. 1990. 'Truth and Probability.' in David Hugh Mellor (ed.), *Philosophical Papers* (Cambridge University Press: Cambridge).

Rapoport, Anatol. 2013. *Decision Theory and Decision Behaviour: Normative and Descriptive Approaches* (Springer Science & Business Media: Dordrecht).

Reiss, Julian. 2008. *Error in Economics: Towards a More Evidence–Based Methodology* (Routledge: Oxford).

———. 2013. *Philosophy of Economics: A Contemporary Introduction* (Routledge: London).

Rodrik, D. 2015. *Economics Rules: Why Economics Works, When It Fails, and How To Tell The Difference* (Oxford University Press: Oxford).

Roll, Richard. 1977. 'A Critique of the Asset Pricing Theory's Tests Part I: On Past and Potential Testability of the Theory', *Journal of financial economics*, 4: 129-76.

Roth, Alvin E. 2002. 'The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics', *Econometrica*, 70: 1341-78.

Roy, Andrew Donald. 1952. 'Safety First and the Holding of Assets', *Econometrica*, 20: 431-49.

Rubinstein, Mark. 2002. 'Markowitz's "Portfolio Selection": A Fifty-Year Retrospective', *The journal of finance*, 57: 1041-45.

Ruppert, David. 2004. *Statistics and Finance: an Introduction* (Springer: New York).

Sachs, Jeffrey. 2008. 'Amid the Rubble of Global Finance, a Blueprint for Bretton Woods II', *The Guardian*, 20 October 2008.

Samuelson, Paul. 2013 (1947). *Foundations of Economic Analysis: Enlarged Edition* (Harvard University Press: Cambridge, MA).

Savage, Leonard. 1954. *The Foundations of Statistics* (John Wiley & Sons: New York).

Scherrer, Deborah. 2016. *Solar System Scale Model*. (Stanford University: California). http://solar-center.stanford.edu/activities/Scale-Model/Drive-By-Science-Scale-Model.pdf

Schmidt, Reinhard H. 1982. 'Methodology and Finance', *Theory and Decision*, 14: 391-413.

Schoenberger, Erica. 2011. 'Why is Gold Valuable? Nature, Social Power and the Value of Things', *Cultural Geographies*, 18: 3-24.

Searle, John R. 1969. *Speech Acts. An Essay in the Philosophy of Language* (Cambridge University Press: Cambridge).

———. 1975. 'Indirect Speech Acts'. In Peter Cole and Jerry Morgan (eds.), *Syntax and Semantics Volume 3: Speech Acts* (Academic Press)

———. 1990. 'Collective Intentions and Actions.' in Philip R Cohen, Jerry Morgan and Martha E Pollack (eds.), *Intentions in Communication* (MIT Press: Cambridge, MA).

———. 1995. *The Construction of Social Reality* (Allen Lane: London).

———. 2003. 'Reply to Barry Smith', *American Journal of Economics and Sociology*, 62: 299-309.

———. 2007. 'Biological Naturalism.' in Max Velmans and Susan Schneider (eds.), *The Blackwell Companion to Consciousness* (Blackwell Publishing: Malden, MA).

———. 2010. *Making the Social World* (Oxford University Press: Oxford).

Sharpe, William F. 1964. 'Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk', *The Journal of Finance*, 19: 425-42.

———. 1966. 'Mutual Fund Performance', *The Journal of Business*, 39: 119-38.

———. 1975. 'Adjusting For Risk in Portfolio Performance Measurement', *The Journal of Portfolio Management*, 1: 29-34.

Simmel, Georg. 2004. *The Philosophy of Money* (Routledge: London).

Simon, Herbert A. 1959. 'Theories of Decision-Making in Economics and Behavioral Science', *The American Economic Review*, 49: 253-83.

Skinner, Burrhus Frederic. 1953. *Science and Human Behavior* (Simon and Schuster: New York).

Smit, JP, Filip Buekens, and Stan Du Plessis. 2011. 'What is Money? An Alternative to Searle's Institutional Facts', *Economics and Philosophy*, 27: 1-22.

Smith, Adam. 1776. *The Wealth of Nations* (The Electric Book Company).

Smith, Barry C. 2003. *John Searle* (Cambridge University Press: Cambridge).

Spence, Michael. 1973. 'Job Market Signaling', *The Quarterly Journal of Economics*, 87: 355-74.

Stambaugh, Robert F. 1982. 'On the Exclusion of Assets from Tests of the Two-Parameter Model: A Sensitivity Analysis', *Journal of Financial Economics*, 10: 237-68.

Theobald, Stuart. 2010. 'Single Stock Futures Debacle.' in Anton Harber and Margaret Renn (ed.), *Troublemakers: The Best of South Africa's Investigative Journalism* (Jacana Media: Johannesburg).

Thomasson, A. 2003. 'Foundations for a Social Ontology', *Protosociology*, 18: 269-90.

Treanor, Jill. 2008. 'Toxic Shock: How the Banking Industry Created a Global Crisis', *The Guardian*, 8 April 2008.

Turner, Adair. 2009. *The Turner Review: A Regulatory Response to the Global Banking Crisis* (Financial Services Authority: London).

Van der Auweraert, Peter. 2007. "Property Restitution in Iraq." In *Symposium on Post-Conflict Property Restitution*. Arlington, VA: United States Department of State, Bureau of Population, Refugees, and Migration; Bureau of Intelligence and Research; Office of the Coordinator of Reconstruction and Stabilization

Viskovatoff, Alex. 2003. 'Searle, Rationality, and Social Reality', *American Journal of Economics and Sociology*, 62: 7-44.

Von Mises, Richard. 1957. *Probability, Statistics, and Truth* (Courier Corporation: North Chelmsford MA).

Wang, Chunyang. 2013. 'Bailouts and Bank Runs: Theory and Evidence from TARP', *European Economic Review*, 64: 169-80.

Weston, Samuel C. 1994. 'Toward a Better Understanding of the Positive/Normative Distinction in Economics', *Economics and Philosophy*, 10: 1-17.

Woolgar, Steve. 1988. *Knowledge and Reflexivity* (Sage: London).