London School of Economics and Political Science

Essays on Information and Career Concerns in Organizations

Kara Alette Contreary

A thesis submitted to the Department of Economics for the degree of Doctor of Philosophy, June 2012

For my parents.

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 other than where I have clearly indicated that it is the work of others.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 27,222 words in total.

Statement of Inclusion of Previous Work

I confirm that chapter 2 is the result of previous study for the degree of MRes I undertook at the London School of Economics and Political Science. That degree was awarded in 2008.

Abstract

The aim of my thesis is to investigate the role of information and career concerns in organizations. To that end, I submit three papers, each of which addresses a unique aspect of a firm's organizational problem. In the first chapter I investigate the incentives of a firm to reveal strategic information to the market in order to make its leader more conservative as regards early decisions. The firm may do so to achieve coordination between different levels of the firm's hierarchy or to improve adaptation to the firm's environment. I give conditions on employees' career concerns that make the firm voluntarily disclose information concerning its strategic decisions. In chapter 2 I ask why rational voters would knowingly re-elect a politician who has expropriated public funds. In this model, the presence of non-strategic ('impressionable') voters means that even welfare-minded politicians occasionally raid the public purse in order to increase their chances of re-election. Being aware of this dynamic, rational voters opt to reward politicians whose misbehavior is solely due to career concerns. Chapter 3 changes tack somewhat to analyze the optimal decision-making protocol for a committee when one member of the committee is overconfident. I show that overconfidence leads an uninformed committee member to respond to his private information, which causes a better-informed member to stop using her private information. This leads to a loss of efficiency under majority rule, and changes the optimal voting rule for the committee to unanimity.

Acknowledgements

It is impossible to spend six years at the LSE without racking up a truly humbling intellectual debt. I have been incredibly fortunate in my professors and my peers, and I am grateful for the opportunity to have learned and taught in such an extraordinary environment.

I wish to thank my supervisor, Gilat Levy, for her guidance, encouragement and insight. The shortcomings of this work persist despite her involvement, and are due entirely to dimensions on which I could not deliver on her good counsel.

Many thanks also to Francesco Nava, for his enthusiasm, advice and unfailing candor. I appreciate his high standards and his faith that they could be met.

For letting me free-ride on their brilliance for the last five years, I am indebted to the Economic Theory Group. Special thanks amongst the faculty go to Erik Eyster, Philipp Kircher and Balázs Szentes. Amongst my fellow students, I have no words equal to the generosity, inspiration and moral support of Patrick Blanchenay, Can Celiktemur, Sebastian Kodritsch, Elena Manzoni, Ines Moreno de Barreda, Stefan Penczynski and Jan Zapal.

I have profited immensely from the atmosphere in EOPP, and am grateful for the resources they've given me. Most especially, I'm grateful for the access I've had to truly stellar people. Fernando Aragon, Michael Best, Konrad Burchardi, Jon DeQuidt, Jas Ellis, Hannes Mueller, Oliver Pardo y Reinoso, Dimitri Szerman—I cannot think of a more open, challenging, and fun collection of people to spend time with both in and out of the research lab.

There is, in addition, a group of individuals who deserve exceptional thanks. They have been my family in London, and I owe my sanity (such as it is) to them. Radha Iyengar, Barbara Richter, Thomas Schelkle, Dan Stein, and especially Tara Mitchell: thanks for all the frozen turkeys, the tipsy dinner parties and ordinary afternoon coffees. Thanks for being a sounding board for ideas, and other times for talking about anything but work. I will miss seeing you every day, but I'm already looking forward to our reunions, wherever they may be.

Finally, nothing remains but to thank the people who made everything possible. I'm fairly certain they have no clue what I've been up for for six years, but they humored my nerdy inclinations and for that I have only this to say: Thanks, Mom and Dad.

Contents

List of Tables

List of Figures

Introduction

In this thesis I set out to address some questions surrounding the role of information and career concerns in organizations.

In chapter 1 I ask when a firm would discourage its leader from changing firm strategy when his information changes. Firms that care about both adaptation and coordination might choose to make leaders more conservative as a way to make the leader's final decision more predictable for followers. This is especially useful for the firm if followers have strong reputation concerns that cause them to inefficiently disregard public information provided by the leader. I compare two different types of leader conservatism—strategic conservatism that arises from reputation concerns and non-strategic conservatism—and provide some empirical predictions for how one might distinguish between reputation concerns and rote conservatism.

In chapter 2 I present a two-period political agency model with re-election. There are two parts: a benchmark with only rational voting and an extension with campaign spending. I find that, if electoral results are sensitive to campaign activity, good politicians (those who wish to maximize voter welfare) can be induced to behave corruptly. Rational voters realize this, and are willing to forgive corruption in politicians they believe to be good. A preliminary welfare analysis shows that, in some instances, voters may prefer to leave the option of campaign spending open for the information it provides, and also may prefer that good types act corruptly in the interest of re-election.

Chapter 3 analyzes a 3-person committee that must decide between two candidate policies. Two members receive an informative signal about the most appropriate policy. The third member is uninformed, but suffers from overconfidence that leads him to think his noisy signal is informative. When he votes according to his signal, adding noise to the outcome, the lesser-informed of the other two agents strategically ignores her signal under majority rule, effectively trading bad information for good. Because of this behavior, majority rule ceases to be the preferred voting rule. Social welfare is instead maximized by selecting a voting rule that requires unanimity, and in some cases requires unanimity for the *a priori* more likely policy. This is never the case in the absence of overconfidence. Thus, bias in one member's self-assessment leads to a reversal of preferences regarding the constitutional voting rule.

Chapter 1

Career Concerns and the Corporate Hierarchy

1.1 Introduction

This paper asks when a firm would reduce its leader's incentive to adapt firm strategy in response to new information. I consider an environment where information arrives in stages and is dispersed among individual agents, and a firm that cares about both coordination between agents and adaptation to the firm's environment. Coordination can be easily achieved by designating a single person the *leader* and instructing other agents (the *followers*) to coordinate on his 'mission statement', or initial strategic decision. However, if incoming information prompts the leader to change strategy, any actions followers have taken in support of the initial decision will be worthless. This undermines the followers' incentives to coordinate with the leader, particularly when their own private information disagrees with the mission statement.

The potential loss from the leader's commitment problem is compounded by reputation concerns amongst the followers, which motivate them to contradict the mission statement too often. The reason this is attractive to followers is that the leader may ultimately prove them right by changing directions, which is good news about their ability. This points to a simple means of making the leader more effective at coordination: reduce his responsiveness to new information.

I capture the basic problem for the firm in a simple model with two agents, a leader and a follower, in which actions are taken over three periods. In the first stage, the leader observes an initial signal of the firm's environment, which he relays to the follower in the form of a mission statement. Next, the follower must decide whether to hew to the leader's initial projection or to take a different action altogether. She also receives information about the state of nature and may conclude that the leader is likely to change his mind, in which case following the mission statement is unwise. Finally, the leader receives a second signal which fully reveals the state of the world. Based on this knowledge, he may opt to reverse his initial decision in order to match the true state. The firm's decision is whether to take actions that make the leader less likely to change his mind in the final period. That is, whether to engender 'conservatism' in the leader.

I find that when the leader is inclined to favor adaptation over coordination, firms which care sufficiently about coordination want their leaders to be conservative. On the other hand, when the leader is inclined to favor coordination over adaptation, conservative leaders are favored by firms that care sufficiently about adaptation. Importantly, neither case requires significant disagreement between the leader and the firm on priorities regarding adaptation and coordination.[1]

I analyze two substitutable types of conservatism—strategic and non-strategic—and ask when such conservatism is beneficial for the firm. Strategic conservatism arises from reputation concerns, and is activated by the firm disclosing information to the labor market that reveals something about the leader's talent. Revealing the leader's mission statement makes it reputationally costly for the leader to reverse himself, since the mission statement is based on his private information. If more able leaders have better private information, the market for leaders will use evidence of reversals in forming judgments about a leader's talent.

Non-strategic conservatism does not depend on the incentives facing the leader, but instead relies on a personality type (like overconfidence) or workload (overloading the leader's portfolio) to achieve the same end. Overconfident leaders are more apt to simply think they got it right to begin with; overworked leaders may not have time to reconsider early decisions before they become irreversible.

Both methods have been proposed in the literature, however to my knowledge no work has yet considered how the necessary behavioral assumptions operate when applied to the organization as a whole. The firm's choice to manipulate the leader depends on how his incentives interact with those of other agents within the firm.

[1]When I speak of 'the firm', I have in mind the board of directors or some other group of stakeholders who can influence firm policy on strategic disclosure and hiring practices.

1.2 Related Literature

The most closely related paper is Bolton, Brunnermeier and Veldkamp (2011, hereafter BBV). BBV examines the benefit of hiring an overconfident leader when coordination and adaptation matter. Overconfidence remedies the time-consistency problem from new information tempting leaders to revise their initial decisions in a way that followers dislike. The structure of this model borrows heavily from the BBV setup; they too feature a two-level hierarchy whose agents receive a sequence of signals and act according to information available at the time. Whereas they consider only overconfidence, I also consider the effects of reputation concerns.

There is a substantial literature on inefficiency due to agents' interest in manipulating perceptions of their ability, starting with Fama (1980), Lazear and Rosen (1982) and Holmström (1999—original 1982). Scharfstein and Stein (1990) and Ottaviani and Sørensen (2000) demonstrate that agents mimic the behavior of others for reputational reasons. In Kanodia, Bushman and Dickhaut (1989), Boot (1992), Zweibel (1995) and Prendergast and Stole (1996), decision-makers act either too conservatively or too aggressively, depending on which gives a better impression of their talent. Leaver (2009) shows that regulators are overly lenient in order to prevent unfavorable information being revealed to the market.

I follow Prendergast and Stole (1996) in asserting that agents at different points in their careers may have opposite incentives regarding conformism, and propose that this may affect coordination within a firm. Levy (2004) demonstrates that, rather than herding on public information, agents may be too eager to contradict it (anti-herding). This is more likely true early in an individual's career. I find that conservative leadership reduces followers' incentive to anti-herd, and may even result in herding on the mission statement. My results for the leader's behavior are comparable to those of Ferreira and Rezende (2007), which considers how reputation concerns help a leader commit to a particular project, which in their case increases a follower's willingness to invest in it.

In addition, a number of papers highlight the role of leaders' personal attributes.[2] The most relevant are those characteristics that address a commitment problem, such as a managerial bias toward particular activities. Such biases imply commitment to implement innovations in those areas, which gives workers incentives to put effort into the manager's pet projects (see Rotemberg and Saloner, 2000; Hart and

[2] For a discussion of why an organization would want to have 'leaders' at all, see Komai, Stegeman and Hermalin (2007).

Holmström, 2002; and Van den Steen, 2005). Another branch of this literature is explicitly concerned with overconfidence.[3] Englmaier (2004) finds that firms want to delegate cost-cutting R&D decisions to overconfident managers. This is similar to my result that firms may prefer overconfident leaders. Blanes i Vidal and Möller (2007) show how overconfidence can be helpful in a situation similar to that of Ferreira and Rezende (2007). However, in their case overconfidence is beneficial because it *increases* responsiveness to information; in my case, overconfidence helps because it *decreases* responsiveness to information.

Outline

The plan of the paper is as follows. Section 1.3 introduces the model and section 1.4 discusses some relevant benchmark results. Section 1.5 analyzes strategic conservatism that arises from publishing the leader's mission statement, and shows when that is favorable for the firm. Section 1.6 covers non-strategic leader conservatism and then applies the same behavioral assumptions to the follower. Section 1.7 discusses empirical predictions and section 1.8 concludes. All proofs are contained in the appendix.

1.3 Model

I consider a firm comprised of two agents, a leader L (he) and a follower F (she). The firm operates in an environment characterized by a state $w \in \{A, B\}$. *Ex ante* the two states are equally likely. Each player has private information about the state of the world. In addition there is an evaluator E, which serves as a proxy for the external labor market. E's only action is to form a belief about the leader's ability to correctly perceive the state of the world. The game proceeds as follows:

Period 1: The leader observes a private signal $s_1 \in \{A, B\}$ and takes an initial action $a_1 \in \{A, B\}$. The leader's accuracy is $Pr\left(s_1 = w \mid w\right) = t$, where t is known only to the leader but it is common knowledge that t comes from the distribution H with support on $\left[\frac{1}{2}, 1\right]$. a_1 is interpreted as the 'mission statement', and is best understood as the direction in which the leader thinks the firm should go.[4] Action a_1 is always observable by the follower, but the firm can choose to publicize the

[3] For empirical evidence of overconfidence in leadership, see Malmendier and Tate (2005, 2008) and Camerer and Lovallo (1999).

[4] This could be, for example, entering a new market or undertaking a merger.

mission statement, in which case the market, in the form of the evaluator E, will also observe the leader's initial choice. I assume that if the firm's policy is to disclose a_1, then the same action is observed by both F and E. That is, the leader cannot send contradictory signals to different parties.[5]

Period 2: The follower observes a_1 and receives a private signal $s_F \in \{A, B\}$. Similar to the leader, $Pr(s_F = w \mid w) = p$, where p is the follower's private information. It is common knowledge that p is drawn from the distribution G with support on $\left[\frac{1}{2}, 1\right]$. The signals s_1 and s_F are conditionally independent. Based on a_1 and s_F, the follower chooses her action $a_F \in \{A, B\}$.

Period 3: Finally the leader learns the true state of the world w. After seeing a_F and w, the leader takes a final action $a_2 \in \{A, B\}$.

For the duration, consider the case of $s_1 = A$. This is without loss of generality due to the prior $Pr(w = A) = \frac{1}{2}$. Results are symmetric for the alternative assumption $s_1 = B$.

Payoffs

The firm's profit has two components: adaptation to the state of the world and coordination between the two players. The relative importance of these two components is captured by a parameter $\gamma \in [0, 1]$ such that the firm's total payoff is

$$V = \gamma I + (1 - \gamma) I^F$$

where I is an indicator that takes value 1 if $a_2 = w$ (adaptation) and I^F is another indicator that takes value 1 if $a_F = a_2$ (coordination). I assume that the leader cares about both adaptation and coordination (not necessarily in the same way as the firm),[6] but that the follower simply cares about matching the leader's action at the end of play (i.e. coordinating with the leader).

In addition, it may be the case that both the follower and the leader care about their reputations. That is, they care what the prevailing estimate of their 'type' p or t is. The follower cares about her reputation within the firm, so the leader's assessment of the follower's type is relevant. This task does not affect the leader's

[5]Assuming to the contrary that contradictory messages are possible would not change anything, as contradictory messages would never be sent in equilibrium.

[6]For example, if coordination mostly affects short-term benefits, the firm may have a longer time horizon that puts greater weight on adaptation.

payoff directly, but I assume the leader tries his best to guess the follower's type. The leader, on the other hand, cares about his reputation in the external market, or E's belief about his type. The leader's utility is then

$$U^L(a_1, a_2; a_F, w) = \gamma^L I + \left(1 - \gamma^L\right) I^F + \lambda\tau$$

where τ is E's posterior estimate of t given the actions observed. That is, $\tau = E[t \mid a_1, a_F, a_2]$ if the mission statement is published and $\tau = E[t \mid a_F, a_2]$ if not. Utility for the follower is given by

$$U^F(a_F; a_1, a_2, w) = I^F + \mu\pi$$

where π is L's updated belief about F's type p having observed her action and the true state of the world, so that $\pi = E[p \mid a_1, a_F, w]$.

Strategies

The leader's strategy is to pick a_1 in period 1 (a function $\zeta : s_1 \to \{A, B\}$), and to pick a_2 in period 3 (a function $\rho : (a_1, a_F, w) \to \{A, B\}$). Likewise, the follower's strategy is to pick a_F, that is, a function $\sigma : (a_1, s_F, p) \to \{A, B\}$.[7]

E conjectures ρ and σ, observes (a_1, a_F, a_2) and forms beliefs τ about the leader's type using Bayes' rule wherever possible. L conjectures σ, observes (a_1, a_F, w) and forms beliefs π about the follower's type, again using Bayes' rule where possible.

With regard to the decision component of the player's payoffs, it is immediately verifiable in period 3 whether the follower correctly anticipated the leader's action, but I assume that at the time the leader's ability is evaluated, the state of the world is not yet verifiable. This captures the fact that leaders often have to take actions which have consequences in the long run, but that their reputations matter in the short run.[8] So I explicitly assume that the firm can contract on the state of the world with the leader, but that the leader's reputation concerns matter in the interim, as he may avail himself of an outside option before the results of his actions are fully realized.

In short, the leader and follower differ in three ways:

[7]Technically, $\zeta : \{A, B\} \times \left[\frac{1}{2}, 1\right] \to \{A, B\}$, $\rho : \{A, B\}^3 \to \{A, B\}$, and $\sigma : \{A, B\}^2 \times \left[\frac{1}{2}, 1\right] \to \{A, B\}$.

[8]This is the same assumption made in Ferreira and Rezende (2007).

1. Order of moves: The leader is responsible for the ultimate decision taken, and also gets to provide an indication at the beginning of how he is likely to choose. The follower has to take an action in between, before all the information is in, and her decision may be rendered null by the leader.

2. Decision component of payoff: The leader's bonus depends on both adaptation and coordination; the follower's bonus is for coordination only (i.e. by coming up with an idea to suit the firm's new direction, getting a qualification that is only useful if the firm undertakes a specific project...).

3. Reputation locus: Leaders care more about their reputation outside the firm (perhaps because of legacy interests or because headhunters approach them with opportunities), whereas followers care more about their reputation inside the firm (perhaps they are interested in promotion within the firm or get most of their job offers through current and former colleagues).

1 is crucial to my results whereas 2 and 3 are not, provided μ is large enough. If the order of play or information structure were different, this would be a different game. On the other hand, nothing would change if the follower also cared about the leader's final action matching the state of the world as long as her reputation concerns are strong enough. Furthermore, if the follower cared about her reputation outside the firm rather than inside her incentives would be somewhat different (since E never knows the true state of the world), but Proposition 2 of Levy (2004) predicts qualitatively similar results.

The solution concept is perfect Bayesian equilibrium, at times subject to refinement to rule out unreasonable equilibria. The refinement I consider is Banks and Sobel's (1987) Divinity Criterion, which requires out-of-equilibrium beliefs to place relatively more weight on types that gain more from deviating from a proposed equilibrium. Occasionally D1 from Cho and Kreps (1987) is also used to ensure uniqueness of pooling equilibria.[9]

1.4 Benchmark

There are two relevant benchmarks for this model. The first benchmark is the case of no reputation concerns, in which players do not act strategically to affect perceptions

[9] D1 fixes an equilibrium outcome and then eliminates beliefs that put positive probability on a defection by type t if whenever t weakly prefers to defect, there is another type t' that strictly prefers to defect.

of their talent. In this case, $\lambda = \mu = 0$. In the second benchmark the follower has career concerns ($\mu > 0$), and so departs from the non-strategic benchmark in order to affect her reputation within the firm. I first consider the effect of follower career concerns and then in section 1.5 discuss how reputation concerns for the leader can be useful in curtailing the follower's inefficient signaling behavior.

1.4.1 No reputation concerns

First consider the benchmark case in which neither player has reputation concerns ($\lambda = \mu = 0$). Recall the assumption that $s_1 = A$. In period 1, L truthfully reveals his signal to F, since he has no reason not to. So $a_1 = A$. Using backwards induction, the action taken by the leader in period 3, once the follower's action has been taken, depends on the relative importance of adaptation and coordination. Take the example of $a_F = A$. The leader's choice between actions A and B boils down to comparing

$$\gamma^L Pr\left(w = A \mid w\right) + \left(1 - \gamma^L\right) \cdot 1 \gtreqless \gamma^L \left(1 - Pr\left(w = A \mid w\right)\right) + \left(1 - \gamma^L\right) \cdot 0$$

which simplifies to

$$Pr\left(w = A \mid w\right) \gtreqless 1 - \frac{1}{2\gamma^L}.$$

Since $Pr\left(w = A \mid w\right) \in \{0, 1\}$, the leader will match the true state of the world if $\gamma^L \geq \frac{1}{2}$ and will match the follower if $\gamma^L < \frac{1}{2}$. The latter case is somewhat difficult to describe as 'leadership' since the leader will ultimately choose whichever action the follower takes. In this case, the leader is more like an advisor who provides early information to the follower but ultimately makes good on whatever the follower decides.

Knowing this, F compares her expected payoff from taking action A or B. Since she only cares about matching the leader, the relevant comparison is

$$Pr\left(a_2 = A \mid \eta\right) \gtreqless Pr\left(a_2 = B \mid \eta\right)$$

where $\eta \equiv (a_1, s_F, p)$ is the information available to the follower in period 2. In equilibrium, the follower's strategy $\hat{\sigma}$ is characterized by a cutoff $(a_1, \hat{s}_F, \hat{p})$. The follower chooses $a_F = A$ if $Pr\left(a_2 = A \mid a_1, s_F, p, \zeta\right) \geq Pr\left(a_2 = A \mid a_1, \hat{s}_F, \hat{p}, \zeta\right)$ and chooses $a_F = B$ otherwise.[10] In this case, since $\hat{\zeta}\left(s_1\right) = s_1 = A$, the comparison is

[10]This is Lemma 1 of Levy (2004).

$Pr\left(a_2 = A \mid A, s_F, p\right) \geq Pr\left(a_2 = A \mid A, \hat{s}_F, \hat{p}\right)$. Since this is always the case, from here onwards I will drop the argument a_1 from the follower's cutoff strategy.

For values of γ^L greater than or equal to $\frac{1}{2}$, the leader will match the true state of the world. This implies the follower should take action A if and only if

$$\frac{Pr\left(w = A \mid \eta\right)}{Pr\left(w = B \mid \eta\right)} \geq 1. \tag{1.1}$$

If $\gamma^L < \frac{1}{2}$, the follower knows that the leader will match her action, so she is indifferent between her two actions choices. This indifference admits a number of tie-breaking rules, including one which selects the most likely state of the world given η. This is the most efficient equilibrium. Let $\bar{t} \equiv E\left[t\right]$.

Proposition 1. When $\lambda = \mu = 0$ and $\gamma^L \geq \frac{1}{2}$, the unique equilibrium involves $a_1 = s_1$, $a_2 = w$, and $\hat{\sigma} = \left(B, \bar{t}\right)$. When $\gamma^L < \frac{1}{2}$, there exists an equilibrium in which $a_1 = s_1$, $a_2 = a_F$, and $\hat{\sigma} = \left(B, \bar{t}\right)$.

When the leader cares more about adaptation, his final decision reflects the true state of the world, but when he cares more about coordination, his final decision is to match the follower. Knowing this, the follower makes her best guess about the true state of the world given her information. In the first instance this is a strict best response, whereas in the second it is by assumption.

The firm's payoff

The firm's payoff in this case is

$$\hat{V} = \begin{cases} \gamma + (1-\gamma)\left(\int_{\frac{1}{2}}^{\bar{t}} \bar{t} g\left(s\right) \mathrm{d}s + \int_{\bar{t}}^{1} s g\left(s\right) \mathrm{d}s\right) & \gamma^L \geq \frac{1}{2} \\ \gamma\left(\int_{\frac{1}{2}}^{\bar{t}} \bar{t} g\left(s\right) \mathrm{d}s + \int_{\bar{t}}^{1} s g\left(s\right) \mathrm{d}s\right) + (1-\gamma) & \gamma^L < \frac{1}{2}. \end{cases}$$

When $\gamma^L \geq \frac{1}{2}$, the leader matches the true state, so the firm is assured its adaptation payoff γ. Coordination only happens when the follower also chooses the correct state, which happens either when the mission statement is correct and is followed or when the follower is talented enough to follow her own signal and that signal is correct. Conversely, when $\gamma^L < \frac{1}{2}$, coordination is assured, but the firm only gets the adaptation payoff if the follower chooses the correct state in period 2.

Unsurprisingly, this payoff is increasing in $\bar{t}$. The firm would like its leader to be high ability in expectation, so that the follower can confidently follow his lead. This motivates the follower's interest in having a good reputation within the firm. If leaders are promoted from within the firm from the ranks of followers, then followers who are thought to have higher ability are more likely to be promoted.[11] It also motivates the leader's interest in having a good reputation outside the firm in section 1.5. If all firms want leaders of high ability, then having a reputation for talent is likely to generate better outside offers.

1.4.2 Follower reputation concerns and anti-herding

Suppose now that the follower cares about the leader's assessment of her ability ($\mu > 0$). In this case, Levy (2004) shows that 'anti-herding' arises, in which the follower contradicts prior information too often relative to the benchmark.

Let the follower's reputation function be denoted $\pi(a_F; w, \sigma)$. Her reputation also depends on the mission statement a_1, but having established that a_1 is always A I will omit it from the notation from the outset. The value $\pi(a_F; w, \sigma)$ is the leader's expectation of F's type after observing her action and the true state of the world, having conjectured some strategy σ. For example, by Bayes' rule,

$$\pi(a; a, \sigma) = \int_{\frac{1}{2}}^{1} s \frac{(s \Pr(\sigma(s_F, p) = s_F) + (1-s)\Pr(\sigma(s_F, p) \neq s_F))\, g(s)}{\int_{\frac{1}{2}}^{1} (s \Pr(\sigma(s_F, p) = s_F) + (1-s)\Pr(\sigma(s_F, p) \neq s_F))\, g(s)\, \mathrm{d}s} \mathrm{d}s$$

The equilibrium in this case again involves a cutoff strategy $\sigma' = (s'_F, p')$, where $a_F = A$ if and only if $\Pr(w = A \mid s_F, p) \geq \Pr(w = A \mid s'_F, p')$. Knowing that with $\lambda = 0$ the leader acts as in section 1.4.1, the follower's condition for choosing $a_F = A$, equation (1.1), becomes[12]

$$\frac{Pr(w = A \mid \eta)}{Pr(w = B \mid \eta)} \geq \begin{cases} \frac{1+\mu[\pi(B;B,\sigma')-\pi(A;B,\sigma')]}{1+\mu[\pi(A;A,\sigma')-\pi(B;A,\sigma')]} & \gamma^L \geq \frac{1}{2} \\ \frac{\pi(B;B,\sigma'')-\pi(A;B,\sigma'')}{\pi(A;A,\sigma'')-\pi(B;A,\sigma'')} & \gamma^L < \frac{1}{2} \end{cases} \tag{1.2}$$

Looking at (1.2), Levy (2004) shows that there cannot be an equilibrium in which $s'_F = A$, so that in equilibrium only a follower with $s_F = B$ would ever contradict the leader. This follows from the monotone likelihood ratio property. Furthermore,

[11] Note that this implicitly assumes that the firm cannot commit to promoting followers based on any other criteria (like in-job performance).

[12] Details for the follower's problem in all cases are given in the appendix.

the right hand side of (1.2) increases with p', whereas the left hand side decreases with p', giving a unique cutoff with $s_F' = B$ and $p' < \bar{t}$.

Proposition 2. When $\lambda = 0$, $\mu > 0$, and $\gamma^L \geq \frac{1}{2}$, there exists a unique equilibrium in which $a_1 = s_1$, $a_2 = s_2$, and $\sigma' = (B, p')$, where $p' < \bar{t}$. p' decreases in μ. If $\gamma^L < \frac{1}{2}$, there exists a unique equilibrium with $a_1 = s_1$,$a_2 = a_F$ and $\sigma'' = (B, p'')$. For all $\mu \geq 0$, $p'' < p'$.

This is the 'anti-herding'—a follower of type $p \in \left[p', \bar{t}\right]$ (or $p \in \left[p'', \bar{t}\right]$) chooses A in the absence of career concerns. But because taking action B is a signal of confidence in her ability, when she cares about her reputation she will excessively contradict the mission statement. The case of $\gamma^L < \frac{1}{2}$, in which the follower needn't worry about matching the leader because the leader will match *her*, is equivalent to the case of $\mu \to \infty$, in which decision concerns become vanishingly unimportant relative to reputation concerns.

The firm's payoff

The firm's payoff becomes

$$V' = \begin{cases} \gamma + (1-\gamma)\left(\int_{\frac{1}{2}}^{p'} \bar{t} g(s)\,\mathrm{d}s + \int_{p'}^{1} s g(s)\,\mathrm{d}s\right) & \gamma^L \geq \frac{1}{2} \\ \gamma\left(\int_{\frac{1}{2}}^{p''} \bar{t} g(s)\,\mathrm{d}s + \int_{p''}^{1} s g(s)\,\mathrm{d}s\right) + (1-\gamma) & \gamma^L < \frac{1}{2} \end{cases} \tag{1.3}$$

which, compared to (1.4.1) represents a loss of

$$\hat{V} - V' = \begin{cases} (1-\gamma)\int_{p'}^{\bar{t}} \left(\bar{t} - s\right) g(s)\,\mathrm{d}s & \gamma^L \geq \frac{1}{2} \\ \gamma \int_{p''}^{\bar{t}} \left(\bar{t} - s\right) g(s)\,\mathrm{d}s & \gamma^L < \frac{1}{2}. \end{cases}$$

This loss decreases in p' (up to a point), indicating that one way for the firm to address the anti-herding problem would be to decrease μ, if possible. This would involve a higher bonus for matching the leader's action.[13] Depending on the follower's intrinsic career concerns, this may be prohibitively expensive (although the trade-off involved in incentive pay is not something I model here). Therefore I take μ as given and consider an alternative remedy: publishing the leader's mission statement.

[13]However this is only effective if $\gamma^L \geq \frac{1}{2}$.

1.5 Publishing the Mission Statement

Now the firm considers publishing the leader's mission statement, so that the outside market will know whether or not the leader changes his mind from period 1 to period 3. If $\lambda > 0$, this has the effect of making the leader less willing to revise his initial decision in the face of new information. In this model, reputation concerns serve the same purpose as in Ferreira and Rezende (2007). In both cases, the leader's reluctance to change his mind makes his final action more predictable for the follower.

Again, proceed by backwards induction.

1.5.1 The leader's problem

Consider now the case of $\lambda > 0$, so the leader has active reputation concerns. The evaluator is concerned with distinguishing two 'types' of leader: $s_1 = w$ and $s_1 \neq w$. That is, a leader who was initially correct or who was initially incorrect. Bayes' rule implies that $E\left[t \mid s_1 = w\right] \equiv \overline{\tau} > E\left[t \mid s_1 \neq w\right] \equiv \underline{\tau}$. Keep in mind that by assumption $s_1 = A$, so that if $w = A$ it means the leader was initially correct.

Recall that the evaluator never learns the true state of the world. But because the agents' strategies are responsive to their information, their actions provide information to E on the true state of the world, and therefore on how likely it is the leader was initially correct. If the mission statement is not published, the evaluator has four possible observations (a_F, a_2): (A, A), (B, A), (A, B), and (B, B). If the mission statement *is* published, he knows (a_1, a_F, a_2).[14] Using conjectures about the follower's and the leader's strategies, E forms a belief about the true state of the world based on the actions observed, and evaluates the leader based on those beliefs.

Let the evaluator's belief about w be $r\left(a_F, a_2\right) \equiv \Pr\left(w = A \mid a_F, a_2, \sigma, \rho\right)$, where σ and ρ are E's conjectures of the players' strategies. The leader's strategy takes the form $\rho\left(a_F, w\right) = Pr\left(a_2 = A \mid a_F, w\right)$, and σ is a cutoff (s_F, p) for the follower. When the action pair (a_F, a_2) occurs on the equilibrium path, $r\left(a_F, a_2\right)$ must be derived from Bayes' rule. By way of example, if the mission statement is published

[14]Unless necessary for clarity, I will suppress the a_1 argument for concision.

and $s_F^* = B$,

$$r\left(A,A\right) = \frac{\bar{t}\rho\left(A,A\right)\left[1-\int_{p^*}^{1}\left(1-s\right)g\left(s\right)\mathrm{d}s\right]}{\bar{t}\rho\left(A,A\right)\left[1-\int_{p^*}^{1}\left(1-s\right)g\left(s\right)\mathrm{d}s\right]+\left(1-\bar{t}\right)\rho\left(A,B\right)\left[1-\int_{p^*}^{1}sg\left(s\right)\mathrm{d}s\right]}$$

The numerator is the probability of observing the action pair (A, A) following a mission statement of A given that the true state of the world is indeed A. The probability of a correct mission statement is $\bar{t}$, the probability the follower also chooses A is $\int_{\frac{1}{2}}^{p^*} g\left(s\right)\,\mathrm{d}s + \int_{p^*}^{1} sg\left(s\right)\,\mathrm{d}s$ with the first term representing the probability of a type who never contradicts the mission statement, and the second term the probability of a high type getting a correct signal. $\rho\left(A, A\right)$ is the probability of the final action $a_2 = A$ when A is the true state and the follower also chose A. Similarly for the denominator.

The remaining values of $r\left(a_F, a_2\right)$ can be found in the appendix.

The two choices available to the firm are: publish the mission statement or do not publish.

1.5.1.1 Not publish

If the mission statement is not published, then E forms beliefs only based on F's action a_F and L's final action a_2. Not being able to observe whether or not L changed his mind between his initial observation and his final choice means that equilibrium behavior is unchanged from the case in section 1.4.

Lemma 1. If a_1 is not revealed to E, then in any equilibrium that satisfies Banks and Sobel's Divinity Criterion, $a_2 = w$ if $\gamma^L \geq \frac{1}{2}$ and $a_2 = a_F$ if $\gamma^L < \frac{1}{2}$.

The intuition behind this is that when the mission statement is not published, there is no evidence of policy reversals. So the only reason for a leader not to take his preferred action in period 3 is that the evaluator assigns a bad reputation to that action, thinking it likely comes from an incorrect leader. But once a_F and w are realized, both types $s_1 \neq w$ and $s_1 = w$ face exactly the same payoffs, so such beliefs are unreasonable and cannot be part of an equilibrium.

Given this behavior by the leader, the follower also acts as she does in the benchmark. The firm's payoff therefore remains the same as in the benchmark scenario.

If, however, the firm chooses to publish L's mission statement, E can see whether or not the leader changes his mind, and this substantively changes the leader's problem.

1.5.1.2 Publish

Now let E observe the leader's mission statement. There are four histories (a_F, w) at which the leader must make his choice: (A, A), (B, A), (A, B), and (B, B).

In period 3, the leader compares the payoff of taking action A versus action B. Let $\triangle\tau \equiv \overline{\tau} - \underline{\tau}$. Given a_F and w, the leader's comparison is laid out in the following table

$$\begin{array}{lrcl} (a_F, w) & U^L(A \mid a_F, w) & & U^L(B \mid a_F, w) \\ (A, A) & 1 + \lambda r(A, A)\triangle\tau & \gtreqless & \lambda r(A, B)\triangle\tau \\ (A, B) & 1 - \gamma^L + \lambda r(A, A)\triangle\tau & \gtreqless & \gamma^L + \lambda r(A, B)\triangle\tau \\ (B, A) & \gamma^L + \lambda r(B, A)\triangle\tau & \gtreqless & 1 - \gamma^L + \lambda r(B, B)\triangle\tau \\ (B, B) & \lambda r(B, A)\triangle\tau & \gtreqless & 1 + \lambda r(B, B)\triangle\tau \end{array} \tag{1.4}$$

The next lemma establishes that a leader who was initially correct has no reason to reverse himself if the follower agrees with him.

Lemma 2. In any equilibrium which satisfies the Divinity Criterion, $\rho(A, A) = 1$.

That is, if both adaptation and coordination concerns favor sticking with the initial assessment, then the leader does so for sure. This is an intuitive result, as reversing his initial decision is a signal that he was incorrect, which is bad news about his ability. He would never take a decision which is both objectively and strategically bad. How the other types of leader behave depends on the values of λ and γ^L. This behavior is covered by the next proposition. When $\gamma^L < \frac{1}{2}$, the equilibrium may not be unique for some values of λ, so I provide the results which give a lower bound on the benefit of publishing the mission statement for the firm (details are in the appendix).

Proposition 3. When the mission statement is published, a perfect Bayesian equilibrium in period 3 exists. When $\gamma^L \geq \frac{1}{2}$ the equilibrium is unique. Given $\sigma^* = (B, p^*)$, $\rho^*(a_F, w)$ takes the following form for the leader's possible histories:

1. $\gamma^L \geq \frac{1}{2}$

$\rho^*(A,A) = 1$

$\rho^*(B,A) = 1$

$$\rho^*(A,B) = \begin{cases} 0 & \lambda \leq \frac{2\gamma^L - 1}{\triangle\tau} \\ \rho \in (0,1) & \frac{2\gamma^L-1}{\triangle\tau} < \lambda < \frac{2\gamma^L-1}{r^p(A,A)\triangle\tau} \\ 1 & otherwise \end{cases}$$

$$\rho^*(B,B) = \begin{cases} 0 & \lambda \leq \frac{1}{\triangle\tau} \\ \rho \in (0,1) & \frac{1}{\triangle\tau} < \lambda < \frac{1}{r^p(B,A)\triangle\tau} \\ 1 & otherwise \end{cases}$$

2. $\gamma^L < \frac{1}{2}$

$\rho^*(A,A) = 1$

$$\rho^*(B,A) = \begin{cases} 0 & \lambda \leq \frac{1-2\gamma^L}{(1-r^p(B,A))\triangle\tau} \\ 1 & otherwise \end{cases}$$

$\rho^*(A,B) = 1$

$$\rho^*(B,B) = \begin{cases} 0 & \lambda \leq \frac{1}{\triangle\tau} \text{ and } \gamma^L \in \left[\frac{r^p(B,A)}{2}, \frac{1}{2}\right) \\ & \text{ or } \lambda \leq \frac{1-2\gamma^L}{(1-r^p(B,A))\triangle\tau} \text{ and } \gamma^L < \frac{r^p(B,A)}{2} \\ \rho \in (0,1) & \frac{1}{\triangle\tau} < \lambda < \frac{1}{r^p(B,A)\triangle\tau} \text{ and } \gamma^L \in \left[\frac{r^p(B,A)}{2}, \frac{1}{2}\right) \\ & \text{ or } \frac{1-2\gamma^L}{(1-r^p(B,A))\triangle\tau} < \lambda < \frac{1}{r^p(B,A)\triangle\tau} \text{ and } \gamma^L < \frac{r^p(B,A)}{2} \\ 1 & otherwise \end{cases}$$

where the evaluator's beliefs in the pooling equilibria are

$$r^p(A,A) = \frac{\bar{t}\left(1 - \int_{p^*}^1 (1-s)\, g(s)\, \mathrm{d}s\right)}{\bar{t}\left(1 - \int_{p^*}^1 (1-s)\, g(s)\, \mathrm{d}s\right) + \left(1-\bar{t}\right)\left(1 - \int_{p^*}^1 s g(s)\, \mathrm{d}s\right)}$$

and

$$r^p(B,A) = \frac{\bar{t}\int_{p^*}^1 (1-s)\, g(s)\, \mathrm{d}s}{\bar{t}\int_{p^*}^1 (1-s)\, g(s)\, \mathrm{d}s + \left(1-\bar{t}\right)\int_{p^*}^1 s g(s)\, \mathrm{d}s}.$$

When players randomize, they do so according to the following expressions:

$$\rho^*(A,B) = \frac{\bar{t}}{1-\bar{t}} \frac{1 - \int_{p^*}^1 (1-s)\, g(s)\, \mathrm{d}s}{1 - \int_{p^*}^1 s g(s)\, \mathrm{d}s} \frac{\lambda\triangle\tau - \left(2\gamma^L - 1\right)}{2\gamma^L - 1} \tag{1.5}$$

$$\rho^*(B,B) = \frac{\bar{t}}{1-\bar{t}} \frac{\int_{p^*}^1 (1-s)\, g(s)\, \mathrm{d}s}{\int_{p^*}^1 s g(s)\, \mathrm{d}s} \left(\lambda\triangle\tau - 1\right). \tag{1.6}$$

Some comparative statics are immediately discernible. $\rho^*(A,B)$ and $\rho^*(B,B)$

weakly increase in λ. With strong career concerns, leaders are loath to admit they were wrong, preferring to pool with leaders who were correct in period 1.

In addition, $\rho^*(A,B)$ and $\rho^*(B,B)$ increase in $\bar{t}$. As the overall pool of leaders gets better, it becomes even more tempting for leaders who were incorrect to stick to their initial mission statements. This is because the evaluator attributes higher probability to the mission statement being correct, which makes the leader less willing to spontaneously own up that it was wrong.

Furthermore, $\rho^*(A,B)$ decreases in γ^L. As adaptation concerns become more important to the leader, he is less willing to trade the adaptation payoff for the coordination and reputation payoff. In contrast, $\rho^*(B,B)$ does not depend on γ^L. For a leader with history (B,B), adaptation and coordination concerns point in the same direction—B. The only reason for him to choose action A is to manipulate his reputation by claiming to have been correct in period 1.

Finally, both $\rho^*(A,B)$ and $\rho^*(B,B)$ decrease in p^*. This means that as followers become more reluctant to contradict the mission statement (that is, they more often follow the leader), it becomes easier to extract at least some of the leader's final information. The intuition is straightforward: a contradiction by the follower decreases the evaluator's belief that the state of the world is truly what the leader predicted. Therefore, the leader gets a lower reputational payoff from sticking to the mission statement. This is particularly true when only very good types of followers would ever contradict the mission statement. Since the reputational benefit from sticking to his guns is lower, it is less likely to outweigh the payoff from getting the decision correct. This means that an incorrect leader will reveal his true signal more often.

The next section analyzes the implications of the leader's problem for the follower.

1.5.2 The follower's problem

If the firm publishes the leader's mission statement, the leader may fail to revise his initial decision even when he receives information that it was incorrect. The leader's behavior is then described by Proposition 3. The follower now chooses A if and only if

$$\frac{Pr\left(w=A\mid\eta\right)}{Pr\left(w=B\mid\eta\right)} \geq \begin{cases} \frac{1-\rho^*(A,B)-\rho^*(B,B)+\mu[\pi(B;B,\sigma^*)-\pi(A;B,\sigma^*)]}{1+\mu[\pi(A;A,\sigma^*)-\pi(B;A,\sigma^*)]} & \gamma^L \geq \frac{1}{2} \\ \frac{-\rho^*(B,B)+\mu[\pi(B;B,\sigma^{**})-\pi(A;B,\sigma^{**})]}{\rho^*(B,A)+\mu[\pi(A;A,\sigma^{**})-\pi(B;A,\sigma^{**})]} & \gamma^L < \frac{1}{2}. \end{cases}$$

The monotone likelihood ratio again implies that there can be no equilibrium in which $s_F^* = A$, since the right hand side would be less than 1 and the left hand side greater than 1 for all possible values of p^*. So the follower's cutoff strategy will be of the form $\sigma^* = (B, p^*)$. The question of interest concerns how p^* (p^{**}) compares to p' (p''), the benchmark cutoff when the mission statement is not published. If $p^* > p'$, then the anti-herding problem has been mitigated by activating reputation concerns in the leader. Fewer types of follower will inefficiently contradict the mission statement in order to signal their ability.

Proposition 4: When the mission statement is published, a PBE in the whole game exists. The leader's behavior is as described in Proposition 3, and the follower's behavior involves a cutoff (B, p^*) if $\gamma^L \geq \frac{1}{2}$ and (B, p^{**}) if $\gamma^L < \frac{1}{2}$. Both p^* and p^{**}decrease in μ and increase in λ. For all parameter values, $p^* \geq p'$ and $p^{**} \geq p''$.

Note that in this case, it is possible for the follower to start herding rather than anti-herding on the leader's signal, that is, for p^* $(p^{**}) > \bar{t}$. This could happen, for instance, if μ were sufficiently small.

I have shown that exposing the leader to market evaluation lessens the anti-herding problem. It does so by giving the leader incentives to stick with his original action choice more often, which makes his final action more predictable for the follower. The follower must therefore be more certain that the leader will discover he was initially wrong before she is willing to contradict the mission statement. However, there are costs to this tactic. When $\gamma^L \geq \frac{1}{2}$, this improvement in coordination comes at the expense of adaptation. The follower's reluctance to knowingly take an inefficient action in order to signal her type is won at the cost of making the leader reluctant to take the correct action given his knowledge of the state of the world. When $\gamma^L < \frac{1}{2}$, publishing the mission statement improves adaptation at the expense of coordination since the leader no longer matches the follower for sure. Whether or not this trade-off is worthwhile for the firm depends on how highly it values coordination versus adaptation.

1.5.3 The firm's payoff

The following results consider a fixed change from p' to p^* in the follower's strategy and ask when a firm would choose to publish the leader's mission statement given the direct effects of parameters of interest. That is, I consider a fixed benefit and ask when the cost is worthwhile for the firm.

In the first instance, consider a leader who personally favors adaptation over coordination.

1.5.3.1 $\gamma^L \geq \frac{1}{2}$

The firm's payoff becomes

$$\begin{aligned} V^* &= \bar{t}\left[\gamma+(1-\gamma)\left(1-\int_{p^*}^{1}(1-s)\,g(s)\,\mathrm{d}s\right)\right] \\ &\quad +\left(1-\bar{t}\right)\left(1-\int_{p^*}^{1}sg(s)\,\mathrm{d}s\right)\left(\left(1-\rho^*(A,B)\right)\gamma+\rho^*(A,B)(1-\gamma)\right) \\ &\quad +\left(1-\bar{t}\right)\int_{p^*}^{1}sg(s)\,\mathrm{d}s\left(1-\rho^*(B,B)\right) \end{aligned}$$

I want to know when V^* is greater than V', the benchmark payoff from section 1.4.2. Note that if the equilibrium is a pooling one $\left(\lambda \leq \frac{2\gamma^L-1}{\triangle\tau}\right)$, where $\rho^*(A,B) = \rho^*(B,B) = 0$, then $p^* = p'$ and the difference between V^* and V' is 0. Otherwise there will be a non-zero difference in payoff.

The condition for $V^* > V'$ is

$$\frac{\gamma}{1-\gamma} < \frac{\int_{p'}^{p^*}\left(\bar{t}-s\right)g(s)\,\mathrm{d}s+\left(1-\bar{t}\right)\left[\rho^*(A,B)\left(1-\int_{p^*}^{1}sg(s)\,\mathrm{d}s\right)-\rho^*(B,B)\int_{p^*}^{1}sg(s)\,\mathrm{d}s\right]}{\left(1-\bar{t}\right)\left[\rho^*(A,B)\left(1-\int_{p^*}^{1}sg(s)\,\mathrm{d}s\right)+\rho^*(B,B)\int_{p^*}^{1}sg(s)\,\mathrm{d}s\right]}. \tag{1.7}$$

Assumption 1. $\int_{p'}^{p^*}\left(\bar{t}-s\right)g(s)\,\mathrm{d}s$ is non-negative.

Assumption 1 holds either when μ or $G\left(\bar{t}\right)$ is large enough.[15] It puts a limit on the extent to which talented follower types $\left(p > \bar{t}\right)$ start to herd on the mission statement, ignoring their private information.

In this case publishing the mission statement trades adaptation for coordination. Inequality (1.7) captures this trade-off. The numerator of the right hand side is the increased probability of getting the coordination payoff $(1-\gamma)$. The denominator is

[15] If assumption 1 is violated it is still possible to prove a similar result to Proposition 5. For example, in Lemma 5a, as long as $\int_{p'}^{p^*}\left(\bar{t}-s\right)g(s)\,ds + 2\bar{t}\left(1-\int_{p^*}^{1}(1-s)\,g(s)\,ds\right)\frac{1-\gamma^L}{2\gamma^L-1} > 0$, there will exist some subset $\left[\underline{\lambda}, \frac{1}{\Delta\tau}\right] \subset \left(\frac{2\gamma^L-1}{\Delta\tau}, \frac{1}{\Delta\tau}\right]$ such that for all $\lambda \in \left[\underline{\lambda}, \frac{1}{\Delta\tau}\right]$ there exists $\gamma_1' \geq 0$ such that $\gamma \in [0,\gamma_1']$ implies the firm (weakly) prefers to publish the mission statement. In that case, for $\lambda \in \left(\frac{2\gamma^L-1}{\Delta\tau}, \underline{\lambda}\right)$, no firm would choose to publish.

the decreased probability of getting the adaptation payoff γ. Whenever the benefit from increased coordination outweighs the cost of decreased adaptation, the firm prefers to publish the mission statement.

As long as the numerator of the right hand side is positive, some types of firm will choose to publish. If the numerator is positive for $\rho^*(A,B) = \rho^*(B,B) = 1$, then it is positive for all λ by assumption 1 and equations (1.5) and (1.6). Otherwise the firm only ever publishes if career concerns are not too strong.

Proposition 5. Let assumption 1 hold and let $\gamma^L \geq \frac{1}{2}$. For all $p^* > p'$, there exists $\gamma_\lambda > 0$ such that the firm prefers to publish the mission statement if and only if $\gamma \in [0, \gamma_\lambda]$ as long as one of the following is true:

1. $\int_{p'}^{p^*} \left(\bar{t} - s\right) g(s)\, \mathrm{d}s + \left(1 - \bar{t}\right)\left(1 - 2\int_{p^*}^{1} sg(s)\, \mathrm{d}s\right) > 0$ and $\lambda > \frac{2\gamma^L - 1}{\triangle\tau}$

2. $\int_{p'}^{p^*} \left(\bar{t} - s\right) g(s)\, \mathrm{d}s + \left(1 - \bar{t}\right)\left(1 - 2\int_{p^*}^{1} sg(s)\, \mathrm{d}s\right) \leq 0$ and $\lambda \in \left(\frac{2\gamma^L - 1}{\triangle\tau}, \bar{\lambda}\right)$, where $\bar{\lambda} \in \left(\frac{2\gamma^L - 1}{\triangle\tau}, \frac{1}{r^p(B,A)\triangle\tau}\right)$ and is unique.
   In either case, if $\gamma^L \in \left[\frac{1}{2}, \frac{1 + r^p(A,A)}{2}\right]$, γ_λ (weakly) decreases with λ.

The first case obtains when either $\bar{t}$ or p^* is sufficiently high.

Proposition 5 shows that when the leader is inclined to favor adaptation $\left(\gamma^L \geq \frac{1}{2}\right)$, firms which care sufficiently about coordination find it profitable to publish the leader's mission statement. This is because attaching career concerns to the leader makes him more predictable for the follower. She, in turn, chooses not to disregard his information so often, which improves coordination. The follower is less likely to contradict the leader and he is less likely to contradict himself. This makes firms that care about coordination better off.

Note that this does not require strong disagreement between the firm and the leader about the value of adaptation versus coordination. For low levels of λ $\left(\lambda \in \left(\frac{2\gamma^L - 1}{\triangle\tau}, \frac{1}{\triangle\tau}\right], \text{ for example}\right)$, it could be the case that $\gamma \geq \gamma^L$ and still the firm chooses to publish the leader's mission statement if $\int_{p'}^{p^*} \left(\bar{t} - s\right) g(s)\, \mathrm{d}s$ is large enough.

1.5.3.2 $\gamma^L < \frac{1}{2}$

If, on the other hand, $\gamma^L < \frac{1}{2}$, then the firm's payoff is

$$\begin{aligned} V^{**} &= \bar{t}\left[1 - \int_{p^*}^{1} (1-s)\, g(s)\, \mathrm{d}s + \int_{p^*}^{1} (1-s)\, g(s)\, \mathrm{d}s \left[\rho(B,A)\gamma + (1-\rho(B,A))(1-\gamma)\right]\right] \\ &\quad + \left(1-\bar{t}\right)\left[\left(1 - \int_{p^*}^{1} s g(s)\, \mathrm{d}s\right)(1-\gamma) + \int_{p^*}^{1} s g(s)\, \mathrm{d}s \left(1 - \rho^*(B,B)\right)\right] \end{aligned}$$

Note again that if the equilibrium is a pooling one $\left(\lambda \leq \frac{1-2\gamma^L}{(1-r^p(B,A))\triangle\tau}\right)$, where $\rho^*(B,A) = \rho^*(B,B) = 0$ and $p^{**} = p''$, then $V^{**} = V''$. Otherwise there will be a non-zero difference in payoff, and the condition for this difference being positive is

$$\frac{\gamma}{1-\gamma} > \frac{\bar{t}\rho^*(B,A)\int_{p^{**}}^{1}(1-s)\, g(s)\, \mathrm{d}s + \left(1-\bar{t}\right)\rho^*(B,B)\int_{p^{**}}^{1} s g(s)\, \mathrm{d}s}{\int_{p''}^{p^{**}}\left(\bar{t}-s\right) g(s)\, \mathrm{d}s + \bar{t}\rho^*(B,A)\int_{p^{**}}^{1}(1-s)\, g(s)\, \mathrm{d}s - \left(1-\bar{t}\right)\rho^*(B,B)\int_{p^{**}}^{1} s g(s)\, \mathrm{d}s}.$$

Here the firm is trading coordination for adaptation. When the increase in payoff from better adaptation outweighs the cost in coordination, the firm publishes the mission statement.

Proposition 6. Let assumption 1 hold and let $\gamma^L < \frac{1}{2}$. For all $p^{**} > p''$ and all $\lambda > \frac{1-2\gamma^L}{(1-r^p(B,A))\triangle\tau}$, there exists a $\gamma_\lambda < 1$ such that the firm prefers to publish the mission statement if and only if $\gamma \in [\gamma_\lambda, 1]$. γ_λ (weakly) increases in λ.

Proposition 6 shows that when the leader is biased in favor of coordination, firms that care sufficiently about adaptation find it profitable to publish the leader's mission statement. This is because when career concerns do not apply, the leader simply matches the follower. This means the follower has effectively no decision concerns and makes her decision entirely based on the effect it will have on her in-firm reputation. Publishing the mission statement, and therefore making the leader less willing to change his mind, disciplines the follower. If she too zealously disregards the leader's information in order to signal her own ability, she risks the leader being unwilling to match her and losing out on the decision component of her payoff. This makes her less likely to favor her own information over the leader's better information, meaning that the final decision is more likely to be the correct one even if the leader most often chooses to match the follower. Note again that it is not necessary for the firm and the leader to have radically different priorities. When λ is close to $\frac{1}{\triangle\tau}$, it could be the case that $\gamma \leq \gamma^L < \frac{1}{2}$ and the firm would still prefer to publish.

1.6 Non-strategic Conservatism

Instead of publishing the mission statement to attach career concerns to the leader's decision problem, the firm may opt to improve coordination by other, non-strategic, means. By this I mean the firm can choose a leader's characteristics or adopt a policy that prevents the leader from revisiting his initial decision in some cases. One such mechanism is hiring an overconfident leader. This would entail a leader who is not subject to reputation concerns, but who is nevertheless more prone to sticking with his initial impression because he simply believes he is right. Thus, overconfidence in the leader serves the same function as career concerns—both make the leader less likely to change his mind once he's decided on the mission statement, and therefore more predictable for the follower. Another option would be to select a heavy workload for the leader such that some decisions cannot be revisited in time to adjust direction before profits are generated.

In the context of this model, non-strategic conservatism is modeled as follows: a fraction $\alpha \in (0, 1)$ of decisions are never revised given new information. It could be that in some cases the leader is so convinced his initial impression is correct that he does not even bother looking at the new information that comes in period 3, or that instead he simply does not have time to consider new evidence before the decision is irreversible. I assume that the leader's failure to consider new information does not depend on the follower's action. α can be interpreted as the degree of overconfidence/overwork; as α gets close to 1, the leader is unlikely ever to change directions once he has made his initial choice.

1.6.1 Conservative leader

As the following sections show, non-strategic conservatism produces a similar effect on follower anti-herding. Regardless of the leader's attitude toward coordination and adaptation, if he occasionally fails to revise his initial decision, the follower's incentives to strategically disregard the mission statement are reduced. Whether or not the firm wishes to take advantage of this depends on γ.

1.6.1.1 $\gamma^L \geq \frac{1}{2}$

With probability $1-\alpha$ the leader acts as in section 1.4 on page 17. The follower chooses action A if and only if

$$\frac{Pr\left(w=A\mid\eta\right)}{Pr\left(w=B\mid\eta\right)}\geq\frac{1-2\alpha+\mu\left[\pi\left(B;B,\sigma^{\alpha}\right)-\pi\left(A;B,\sigma^{\alpha}\right)\right]}{1+\mu\left[\pi\left(A;A,\sigma^{\alpha}\right)-\pi\left(B;A,\sigma^{\alpha}\right)\right]}$$

with p^{α} defined as the value of p that satisfies the expression with equality. Note that $p^{\alpha}>p'$ and that p^{α} increases in α. Here again, if α is large enough it may be the case that the follower herds rather than anti-herds $\left(p^{\alpha}>\bar{t}\right)$.

The firm's payoff

The firm's payoff is given by

$$\begin{aligned}V^{\alpha} \;=\;& \bar{t}\left[\gamma+(1-\gamma)\left(1-\int_{p^{\alpha}}^{1}(1-s)\,g(s)\,\mathrm{d}s\right)\right]\\&+\left(1-\bar{t}\right)\left[\left(1-\int_{p^{\alpha}}^{1}sg(s)\,\mathrm{d}s\right)\left(\alpha(1-\gamma)+(1-\alpha)\gamma\right)+\int_{p^{\alpha}}^{1}sg(s)\,\mathrm{d}s\,(1-\alpha)\right].\end{aligned}$$

Comparing this value to the firm's payoff when $\alpha=0$ (and $p^{\alpha}=p'$), the condition for the firm to prefer an $\alpha>0$ is

$$\frac{\gamma}{1-\gamma}<\frac{\int_{p'}^{p^{\alpha}}\left(\bar{t}-s\right)g(s)\,\mathrm{d}s+\alpha\left(1-\bar{t}\right)\left(1-2\int_{p^{\alpha}}^{1}sg(s)\,\mathrm{d}s\right)}{\alpha\left(1-\bar{t}\right)}. \tag{1.8}$$

Clearly whether or not any type of firm would choose to employ an $\alpha>0$ to engender leader conservatism depends on the sign of the right hand side.

Proposition 7. For any $p^{\alpha}>p'$ and $\alpha>0$ such that $\int_{p'}^{p^{\alpha}}\left(\bar{t}-s\right)g(s)\,\mathrm{d}s+\alpha\left(1-\bar{t}\right)\left(1-2\int_{p^{\alpha}}^{1}sg(s)\,\mathrm{d}s\right)>0$, there exists a $\gamma_{\alpha}\in(0,1)$ such that the firm prefers conservatism of degree $\alpha>0$ to $\alpha=0$ if and only if $\gamma\in[0,\gamma_{\alpha}]$.

The inequality will be satisfied when $\bar{t}$ or p^{α} is sufficiently high. This result parallels Proposition 5. If the leader prefers adaptation, then firms which care sufficiently about coordination prefer that the leader be conservative.

1.6.1.2 $\gamma^L < \frac{1}{2}$

The follower chooses A if and only if

$$\frac{Pr\left(w = A \mid \eta\right)}{Pr\left(w = B \mid \eta\right)} \geq \frac{-\alpha + \mu\left[\pi\left(B; B, \sigma^{\alpha\alpha}\right) - \pi\left(A; B, \sigma^{\alpha\alpha}\right)\right]}{\alpha + \mu\left[\pi\left(A; A, \sigma^{\alpha\alpha}\right) - \pi\left(B; A, \sigma^{\alpha\alpha}\right)\right]}$$

with $p^{\alpha\alpha}$ defined as the value of p that satisfies this expression with equality, noting that $p^{\alpha\alpha} > p''$ and that $p^{\alpha\alpha}$ increases in α with the possibility of herding $\left(p^{\alpha\alpha} > \bar{t}\right)$.

The firm's payoff

The firm's payoff becomes

$$\begin{aligned} V^{\alpha\alpha} &= \bar{t}\left[1 - \int_{p^{\alpha\alpha}}^{1} (1-s)\, g(s)\, \mathrm{d}s + \int_{p^{\alpha\alpha}}^{1} (1-s)\, g(s)\, \mathrm{d}s \left(\alpha\gamma + (1-\alpha)(1-\gamma)\right)\right] \\ &\quad + \left(1 - \bar{t}\right)\left[\left(1 - \int_{p^{\alpha\alpha}}^{1} s g(s)\, \mathrm{d}s\right)(1-\gamma) + \int_{p^{\alpha\alpha}}^{1} s g(s)\, \mathrm{d}s\, (1-\alpha)\right] \end{aligned}$$

and the condition for $V^{\alpha\alpha} > V''$ is

$$\frac{\gamma}{1-\gamma} > \frac{\bar{t}(1-\alpha)\int_{p^{\alpha\alpha}}^{1}(1-s)\, g(s)\, \mathrm{d}s + \left(1-\bar{t}\right)\alpha \int_{p^{\alpha\alpha}}^{1} s g(s)\, \mathrm{d}s}{\int_{p''}^{p^{\alpha\alpha}}\left(\bar{t}-s\right) g(s)\, \mathrm{d}s + \bar{t}(1-\alpha)\int_{p^{\alpha\alpha}}^{1}(1-s)\, g(s)\, \mathrm{d}s - \left(1-\bar{t}\right)\alpha\int_{p^{\alpha\alpha}}^{1} s g(s)\, \mathrm{d}s}. \tag{1.9}$$

Proposition 8. For all $p^{\alpha\alpha} > p''$ and $\alpha > 0$ such that $\int_{p''}^{p^{\alpha\alpha}}\left(\bar{t}-s\right) g(s)\, \mathrm{d}s + \bar{t}(1-\alpha)\int_{p^{\alpha\alpha}}^{1}(1-s)\, g(s)\, \mathrm{d}s - \left(1-\bar{t}\right)\alpha\int_{p^{\alpha\alpha}}^{1} s g(s)\, \mathrm{d}s \neq 0$, there exists $\gamma_\alpha < 1$ such that the firm prefers to induce conservatism of degree α if and only if $\gamma \in [\gamma_\alpha, 1]$.

This result parallels Proposition 6. If the leader prefers coordination, then firms which care sufficiently about adaptation prefer that the leader be conservative.

The results in this section echo those of section 1.5.3 on page 27. This indicates that strategic and non-strategic conservatism in the leader are substitutes from the firm's perspective.

1.6.2 Non-strategic follower

If it is the case that leaders are apt to overestimate their ability in some cases, then it stands to reason that followers might also incorrectly perceive their ability. And if leaders are occasionally overworked and let due diligence slip through the cracks, then so too might followers.

Unlike the leader's case, overconfidence and overwork produce opposite results for the follower. An overconfident follower is likely to place *too much* weight on her own information because she believes it to be highly accurate. By contrast, an overworked (or underconfident) follower is likely to place *too little* weight on her private information in favor of simply using publicly available information in the form of the mission statement.

1.6.2.1 Overconfident follower

Let the follower's overconfidence be modeled as follows: with some probability $\beta \in (0,1)$, she becomes convinced that her signal is correct. This means that in her indifference condition the probability ratio $\frac{Pr(w=A|\eta)}{Pr(w=B|\eta)}$ is either 0 or unboundedly large, so that she will follow her signal regardless of the mission statement. With probability $1-\beta$ she behaves as in the previous section, with the cutoff value p^{α} (or $p^{\alpha\alpha}$) appropriate to the leader's behavior. Overconfidence in the follower thus enhances the effect of career concerns, in that the follower too often contradicts the mission statement relative to the benchmark. An overconfident follower is 'immoderate' rather than conservative.

Let $\bar{p} \equiv E[p]$. Then, for $\gamma^L \geq \frac{1}{2}$ the firm's payoff is

$$\begin{aligned} V^{\beta} &= \bar{t}\left[\gamma + (1-\gamma)\left[\beta\bar{p} + (1-\beta)\left(1 - \int_{p^{\alpha}}^{1}(1-s)\,g(s)\,\mathrm{d}s\right)\right]\right] \\ &\quad + \left(1-\bar{t}\right)(1-\alpha)\left[\gamma + (1-\gamma)\left(\beta\bar{p} + (1-\beta)\int_{p^{\alpha}}^{1} sg(s)\,\mathrm{d}s\right)\right] \\ &\quad + \left(1-\bar{t}\right)\alpha(1-\gamma)\left(\beta(1-\bar{p}) + (1-\beta)\left(1 - \int_{p^{\alpha}}^{1} sg(s)\,\mathrm{d}s\right)\right). \end{aligned} \tag{1.10}$$

If $\gamma^L < \frac{1}{2}$, then

$$\begin{aligned} V^{\beta\beta} &= \bar{t}\beta\left(\bar{p} + (1-\bar{p})\left(\alpha\gamma + (1-\alpha)(1-\gamma)\right)\right) \\ &\quad + \bar{t}(1-\beta)\left(1 - \int_{p^{\alpha\alpha}}^{1}(1-s)\,g(s)\,\mathrm{d}s + \int_{p^{\alpha\alpha}}^{1}(1-s)\,g(s)\,\mathrm{d}s\left(\alpha\gamma + (1-\alpha)(1-\gamma)\right)\right) \\ &\quad + \left(1-\bar{t}\right)\beta\left(\bar{p}(1-\alpha) + (1-\bar{p})(1-\gamma)\right) \\ &\quad + \left(1-\bar{t}\right)(1-\beta)\left(\left(1 - \int_{p^{\alpha\alpha}}^{1} sg(s)\,\mathrm{d}s\right)(1-\gamma) + \int_{p^{\alpha\alpha}}^{1} sg(s)\,\mathrm{d}s\,(1-\alpha)\right). \end{aligned}$$

1.6.2.2 Overworked/underconfident follower

A follower can also be time-constrained and not be able to fully incorporate all the available information when she takes her decision. When this is the case, she would excessively rely on public information. It is also possible for the follower to be *underconfident*. This is a story in which the follower occasionally becomes convinced her signal is wrong, and that she is better off following the leader and acting based on the mission statement than paying attention to her own signal. Both of these scenarios produce the same result.

In this model a conservative follower has probability $\hat{\beta}$ of following the mission statement for sure, regardless of s_F. With probability $1-\hat{\beta}$, the follower acts as above (p^{α} or $p^{\alpha\alpha}$). For simplicity say that $\hat{\beta}=\beta$, so that the firm's choice is between followers who are equally likely to depart from the benchmark, just in opposite directions. The firm's payoff in the case of underconfidence when $\gamma^{L}\geq\frac{1}{2}$ is

$$\begin{aligned} V^{\hat{B}} &= \bar{t}\left[\gamma+(1-\gamma)\left[\beta+(1-\beta)\left(1-\int_{p^{\alpha}}^{1}(1-s)\,g(s)\,\mathrm{d}s\right)\right]\right] \\ &\quad +\left(1-\bar{t}\right)(1-\alpha)\left[\gamma+(1-\gamma)(1-\beta)\int_{p^{\alpha}}^{1}sg(s)\,\mathrm{d}s\right] \\ &\quad +\left(1-\bar{t}\right)\alpha(1-\gamma)\left(\beta+(1-\beta)\left(1-\int_{p^{\alpha}}^{1}sg(s)\,\mathrm{d}s\right)\right). \end{aligned}$$

If $\gamma^{L}<\frac{1}{2}$, then

$$\begin{aligned} V^{\hat{\beta}\hat{\beta}} &= \bar{t}\left[\beta+(1-\beta)\left(1-\int_{p^{\alpha\alpha}}^{1}(1-s)\,g(s)\,\mathrm{d}s\right)\right] \\ &\quad +\bar{t}(1-\beta)\int_{p^{\alpha\alpha}}^{1}(1-s)\,g(s)\,\mathrm{d}s\,(\alpha\gamma+(1-\alpha)(1-\gamma)) \\ &\quad +\left(1-\bar{t}\right)\left[\beta(1-\gamma)+(1-\beta)\left(1-\int_{p^{\alpha\alpha}}^{1}sg(s)\,\mathrm{d}s\right)(1-\gamma)\right] \\ &\quad +\left(1-\bar{t}\right)(1-\beta)\int_{p^{\alpha\alpha}}^{1}sg(s)\,\mathrm{d}s\,(1-\alpha). \end{aligned}$$

1.6.2.3 The firm's payoff

To see which type of follower is preferred by the firm, I compare the payoffs from employing a conservative versus an immoderate follower. The firm strictly prefers a conservative follower to an immoderate follower if and only if

$$\frac{\bar{t}\left(1-\bar{p}\right)}{\left(1-\bar{t}\right)\bar{p}} > \begin{cases} 1-2\alpha & \gamma^L \geq \frac{1}{2} \\ \frac{\gamma-\alpha}{\alpha(1-\gamma)+\gamma(1-\alpha)} & \gamma^L < \frac{1}{2}. \end{cases} \tag{1.11}$$

In each case the right hand side is less than 1. This leads to the following result.

Proposition 9. For any level of leader overconfidence α and any β for the follower, if $E\left[t\right] > E\left[p\right]$, the firm prefers a conservative follower to an immoderate one.

If the titles 'leader' and 'follower' are taken seriously, it is quite natural to assume that a leader would be on average of better quality than a follower. Leaders have been promoted up because they have demonstrated high ability in their previous positions.

Proposition 9 has different implications depending on the source of non-strategic behavior. If personality traits are the source, then the firm needs an overconfident leader but an underconfident follower. This would affect the firm's hiring and promotion practice in that the firm might want to hire different types of individuals for different positions in the firm hierarchy. However, the firm's payoff is increasing in the expected ability of the leader, and ability is likely easier to observe within the firm than without.[16]

One (unmodeled) implication of this is that the firm might face a trade-off between hiring overconfident followers but getting to select the best ones for promotion to leadership positions, or hiring underconfident followers and having to recruit leaders from outside the firm. This trade-off was not present previously, when the agent characteristic of interest was the level of reputation concern.

If, on the other hand, the firm uses workload to affect the leader's incentives, then increasing the workload (and reducing agents' ability to fully incorporate all available information into their decisions) produces changes in the appropriate direction for each level of the hierarchy. Overworked leaders occasionally fail to reverse bad decisions, and overworked followers sometimes implement the mission statement by rote.

[16]For a slightly different take on the career prospects of overconfident followers, see Goel and Thakor (2008).

1.7 Empirical Predictions

In this section I briefly consider the empirical predictions of this model. I have indicated a way in which it might be possible to distinguish between reputation concern and overconfidence.

The analysis in sections 1.5 and 1.6 suggests a way to differentiate leaders who are strategically conservative from those who are non-strategically conservative. Leaders in the latter category will change their minds too infrequently regardless of the level of information provided to the market about their early decisions. Publishing the mission statement or not plays no role in their decision to revise their initial decision or follow through on the original plan.

In contrast, leaders who are interested in *appearing* to be confident should respond to the dissemination of information about their initial prediction. Lemma 1 showed that, if the mission statement is not published, a rational leader has no interest in taking an inefficient decision in the final period, regardless of how the follower behaves. By Proposition 3, however, he might fail to incorporate new information into his final decision if it reveals unfavorable information to the market because his early predictions were publicized. The prediction is therefore that overconfident or overworked leaders fail to update based on new information whether or not the firm chooses to disclose information about corporate strategy to the market, whereas rational (but career-concerned) leaders will only under-respond to new information when doing so makes them look better to outsiders because they avoid owning up to mistakes.

1.8 Conclusion

When should a firm disclose its strategic direction to outsiders? When should a firm hire an overconfident leader (or heap responsibilities on a rational one)? How might one tell the difference between reputation concerns and rote conservatism in leader behavior? In this paper I have attempted to address these questions, and shown that the answers depend on how important coordination and adaptation are to the firm. Firms will choose to attach career concerns to their adaptation-minded leaders when coordination is sufficiently important, and to their coordination-minded leaders if the converse is true.

If anti-herding is a concern, such that the firm needs to stop followers taking inefficient actions in order to signal high ability, then even firms who strongly prefer one of the two payoff components prefer to publish. In some cases, a firm could achieve the same effect by hiring an overconfident leader, but overconfident followers present a problem. Firms prefer followers to be underconfident, which poses difficulty for organizations that hope to use follower positions as proving grounds for future leaders. As an alternative, a firm could just crowd the in-box of its employees at all levels.

This model is quite simple, and would benefit from extensions. It would be interesting and useful to consider an extension with multiple followers. The results from BBV suggest that coordination amongst the followers themselves would be improved by the two mechanisms discussed here, making it even more attractive for firms to publish the mission statement, but it would be useful to check this in context. Finally, the binary state in which the leader learns the true state in period 3 is somewhat limiting, and it would be interesting to consider a model with a richer action space and some residual uncertainty.

The main message of the paper is that reputation concerns are a useful lever for firms who wish to provide incentives to coordinate, as are behavioral traits like overconfidence and firm-wide policies like increased workload. I have provided some empirical predictions, and hopefully with further research this line of inquiry will produce useful insight into firm behavior.

Appendix

Preliminaries

Follower's indifference condition

Section 1.4.2 on page 20: The cutoff strategy σ' satisfies

$$\begin{aligned} & Pr\left(w=A \mid \eta\right)\left(1+\mu\pi\left(A;A,\sigma'\right)\right)+Pr\left(w=B \mid \eta\right)\mu\pi\left(A;B,\sigma'\right) \\ = \; & Pr\left(w=A \mid \eta\right)\mu\pi\left(B;A,\sigma'\right)+Pr\left(w=B \mid \eta\right)\left(1+\mu\pi\left(B;B,\sigma'\right)\right) \end{aligned}$$

if $\gamma^L \geq \frac{1}{2}$, and if $\gamma^L < \frac{1}{2}$ it satisfies

$$\begin{aligned} & 1+Pr\left(w=A \mid \eta\right)\mu\pi\left(A;A,\sigma'\right)+Pr\left(w=B \mid \eta\right)\mu\pi\left(A;B,\sigma'\right) \\ = \; & 1+Pr\left(w=A \mid \eta\right)\mu\pi\left(B;A,\sigma'\right)+Pr\left(w=B \mid \eta\right)\mu\pi\left(B;B,\sigma'\right). \end{aligned}$$

Section 1.5.2 on page 26: The cutoff strategy σ^* satisfies

$$\begin{aligned} & Pr\left(w=A \mid \eta\right)\left(1+\mu\pi\left(A;A,\sigma^*\right)\right)+Pr\left(w=B \mid \eta\right)\left(\rho\left(A,B\right)+\mu\pi\left(A;B,\sigma^*\right)\right) \\ = \; & Pr\left(w=A \mid \eta\right)\mu\pi\left(B;A,\sigma^*\right)+Pr\left(w=B \mid \eta\right)\left(1-\rho\left(B,B\right)+\mu\pi\left(B;B,\sigma^*\right)\right) \end{aligned}$$

if $\gamma^L \geq \frac{1}{2}$, and if $\gamma^L < \frac{1}{2}$ it satisfies

$$\begin{aligned} & Pr\left(w=A \mid \eta\right)\left(1+\mu\pi\left(A;A,\sigma^{**}\right)\right)+Pr\left(w=B \mid \eta\right)\left(1+\mu\pi\left(A;B,\sigma^{**}\right)\right) \\ = \; & Pr\left(w=A \mid \eta\right)\left(1-\rho\left(B,A\right)+\mu\pi\left(B;A,\sigma^{**}\right)\right) \\ & +Pr\left(w=B \mid \eta\right)\left(1-\rho\left(B,B\right)+\mu\pi\left(B;B,\sigma^{**}\right)\right). \end{aligned}$$

Section 1.6.1.1 on page 32: The cutoff strategy σ^α satisfies

$$\begin{aligned} & Pr\left(w=A \mid \eta\right)\left(1+\mu\pi\left(A;A,\sigma^\alpha\right)\right)+Pr\left(w=B \mid \eta\right)\left(\alpha+\mu\pi\left(A;B,\sigma^\alpha\right)\right) \\ = \; & Pr\left(w=A \mid \eta\right)\mu\pi\left(B;A,\sigma^\alpha\right)+Pr\left(w=B \mid \eta\right)\left(1-\alpha+\mu\pi\left(B;B,\sigma^\alpha\right)\right). \end{aligned}$$

Section 1.6.1.2 on page 33: The cutoff strategy $\sigma^{\alpha\alpha}$ satisfies

$$\begin{aligned} & Pr\left(w=A \mid \eta\right)\left(1+\mu\pi\left(A;A,\sigma^{\alpha\alpha}\right)\right)+Pr\left(w=B \mid \eta\right)\left(1+\mu\pi\left(A;B,\sigma^{\alpha\alpha}\right)\right) \\ = \; & Pr\left(w=A \mid \eta\right)\left(1-\alpha+\mu\pi\left(B;A,\sigma^{\alpha\alpha}\right)\right)+Pr\left(w=B \mid \eta\right)\left(1-\alpha+\mu\pi\left(B;B,\sigma^{\alpha\alpha}\right)\right). \end{aligned}$$

Formulae for $\mathbf{r}(a_F, a_2)$

From Bayes' rule, given $s_F^* = B$:

$$r(A,A) = \frac{\bar{t}\rho(A,A)\left[1-\int_{p^*}^1 (1-s)\,g(s)\,\mathrm{d}s\right]}{\bar{t}\rho(A,A)\left[1-\int_{p^*}^1 (1-s)\,g(s)\,\mathrm{d}s\right] + \left(1-\bar{t}\right)\rho(A,B)\left[1-\int_{p^*}^1 sg(s)\,\mathrm{d}s\right]} \tag{1.12}$$

$$r(A,B) = \frac{\bar{t}\left(1-\rho(A,A)\right)\left[1-\int_{p^*}^1 (1-s)\,g(s)\,\mathrm{d}s\right]}{\bar{t}\left(1-\rho(A,A)\right)\left[1-\int_{p^*}^1 (1-s)\,g(s)\,\mathrm{d}s\right] + \left(1-\bar{t}\right)\left(1-\rho(A,B)\right)\left[1-\int_{p^*}^1 sg(s)\,\mathrm{d}s\right]} \tag{1.13}$$

$$r(B,A) = \frac{\bar{t}\rho(B,A)\int_{p^*}^1 (1-s)\,g(s)\,\mathrm{d}s}{\bar{t}\rho(B,A)\int_{p^*}^1 (1-s)\,g(s)\,\mathrm{d}s + \left(1-\bar{t}\right)\rho(B,B)\int_{p^*}^1 sg(s)\,\mathrm{d}s} \tag{1.14}$$

$$r(B,B) = \frac{\bar{t}\left(1-\rho(B,A)\right)\int_{p^*}^1 (1-s)\,g(s)\,\mathrm{d}s}{\bar{t}\left(1-\rho(B,A)\right)\int_{p^*}^1 (1-s)\,g(s)\,\mathrm{d}s + \left(1-\bar{t}\right)\left(1-\rho(B,B)\right)\int_{p^*}^1 sg(s)\,\mathrm{d}s} \tag{1.15}$$

Similar expressions can be found for when $s_F^* = A$. Notice that $r(a_F, A)$ decreases in p^* for $a_F \in \{A, B\}$.

A note on equilibrium selection

Recall the equilibrium in the leader's subgame laid out in Proposition 3. For $\lambda \in \left[\frac{1-2\gamma^L}{\triangle\tau}, \frac{1-2\gamma^L}{(1-r^p(B,A))\triangle\tau}\right]$ and $\gamma^L < \frac{1}{2}$, there is also an equilibrium in which a leader with (B, A) randomizes,[17] where he does so with probability

$$\rho(B,A) = 1 - \frac{1-\bar{t}}{\bar{t}}\frac{\int_{p^*}^1 sg(s)\,\mathrm{d}s}{\int_{p^*}^1 (1-s)\,g(s)\,\mathrm{d}s}\frac{\lambda\triangle\tau - \left(1-2\gamma^L\right)}{1-2\gamma^L}.$$

It is immediately apparent that $\rho(B, A)$ decreases with λ, which is somewhat perverse, given that stronger career concerns generally encourage sticking with the mission statement rather than reversing it.

Any time there is an equilibrium in which a leader with (B, A) randomizes, there is also an equilibrium in which he chooses B for sure. The latter is the equilibrium preferred by the leader and is the one I analyze for the results in Proposition 6. If instead I consider the semi-separating equilibrium, then the range of λ over which some types of firm choose to publish increases.

In Proposition 6, $\lambda \in \left[\frac{1-2\gamma^L}{\triangle\tau}, \frac{1-2\gamma^L}{(1-r^p(B,A))\triangle\tau}\right]$ means the firm is indifferent between publishing and not for all values of γ, since the pooling equilibrium arises. If the

[17] Recall that if (B, A) randomizes, then (B, B) always plays B.

semi-separating equilibrium arises, then for all λ in this range, there exists a $\gamma_\lambda < 1$ such that the firm prefers to publish the mission statement if and only if $\gamma \in [\gamma_\lambda, 1]$. I have therefore given a lower bound on how often firms would be expected to publish.

Proofs

Proof of Proposition 1: Consider first the case of $\gamma^L \geq \frac{1}{2}$. Given $a_1 = A$, the follower should take action A if and only if $\frac{Pr(w=A|\eta)}{Pr(w=B|\eta)} \geq 1$. For any p, the left hand side of this expression is $\frac{\bar{t}p}{(1-\bar{t})(1-p)}$ if $s_F = A$, which is always greater than 1. If $s_F = B$, the left hand side is $\frac{\bar{t}(1-p)}{(1-\bar{t})p}$, which strictly decreases in p and equals 1 at exactly $p = \bar{t}$. If, on the other hand, $\gamma^L < \frac{1}{2}$, then since the follower is indifferent between actions A and B given the leader's strategy, an admissible equilibrium strategy is for her to choose $a_F = A$ if and only if $Pr\left(w = A \mid \eta\right) \geq Pr\left(w = B \mid \eta\right)$. ■

Proof of Proposition 2. This follows from Proposition 1 of Levy (2004) and backwards induction.■

Proof of Lemma 1.

The leader's choice between A and B requires comparing (assuming $a_F = A$)

$$\gamma^L Pr\left(w = A \mid w\right) + \left(1 - \gamma^L\right) + \lambda\tau_{-a_1}\left(A; A, \rho\right) \gtreqless \gamma^L Pr\left(w = B \mid w\right) + \lambda\tau_{-a_1}\left(B; A, \rho\right) \tag{1.16}$$

where $\tau_{-a_1}\left(a_2; a_F, \rho\right)$ is the evaluator's belief about the leader's type given that a_1 is unobserved, and given the observed action pair $(a_F, a_2)$and conjectured strategy ρ for the leader.

From (1.16), any equilibrium in which $\gamma^L \geq \frac{1}{2}$ but $a_2 \neq w$ must involve some beliefs for E given a_F that involve $\tau_{-a_1}\left(a; a_F, \rho\right) < \tau_{-a_1}\left(a'; a_F, \rho\right)$, otherwise when $w = a$, L would deviate. This means that, given a_F, E believes that a leader who chooses action a is more likely to have been incorrect than a leader who chooses action a'. But if the true state is indeed a, then any beliefs for E that make it worthwhile for type $s_1 \neq w$ to choose a also make it worthwhile for type $s_1 = w$ to choose a. Divinity thus eliminates beliefs that place higher probability on type $s_1 \neq w$ playing a, and hence all equilibria in which $a_2 \neq w$ because there is some action pair that E conjectures is more likely to come from an incorrect leader. Similarly, any equilibrium in which $\gamma^L < \frac{1}{2}$ but $a_2 \neq a_F$ must involve some beliefs for E that satisfy $\tau_{-a_1}\left(a; a, \rho\right) < \tau_{-a_1}\left(a'; a, \rho\right)$, otherwise L would certainly match the follower regardless of w. But regardless of the state, any beliefs for E that make it worthwhile for type $s_1 \neq w$ to choose $a_2 = a_F$ also make it worthwhile for type $s_1 = w$ to choose a_F. Divinity eliminates such beliefs, and therefore all equilibria in which $a_2 \neq a_F$ because E conjectures only incorrect leaders match the follower. $\square$

Proof of Lemma 2.

Suppose there is a pooling equilibrium where $\rho(A, A) = \rho(A, B) = 0$. That is, both types of leader reverse their initial decisions if the follower agrees with them. If this is an equilibrium, it must be that E has off-equilibrium-path beliefs that a deviation is more likely to have come from a leader who was initially incorrect, or $r(A, A) < r(A, B)$. But for all beliefs of E that make type (A, B) deviate, type (A, A) also wants to deviate since $1 \geq 1 - 2\gamma^L$. Divinity then requires $r(A, A) \geq r(A, B)$. This restriction on beliefs refines away all pooling equilibria in which both types reverse their initial decisions. Next consider fully separating equilibria. If $\rho(A, A) = 0$ (i.e. type (A, A) always plays B), then type (A, B) would strictly prefer to play B as well, so no separation can occur in that direction. In a semi-separating equilibrium, if type (A, A) plays B with positive probability, he must be indifferent between playing A and B, which requires $r(A, A) < r(A, B)$. This means that type (A, B) strictly prefers B. Bayesian updating then requires $r(A, A) > r(A, B)$, which is a contradiction.$\square$

Proof of Proposition 3.

Lemma 1 establishes that $\rho^*(A, A) = 1$. By 1.13 and Bayes' rule, $r(A, B) = 0$. Furthermore, only one of (B, A) and (B, B) can randomize in any equilibrium. To see this, note that if one of them randomizes, he must be indifferent between his action choices. If this is the case, then the other strictly prefers one of the options.

1. $a_F = A$
   The only history I need to consider is (A, B), since Lemma 1 covers (A, A). There are 3 possible types of equilibrium behavior for (A, B): pooling, separating, and semi-separating.

   a) A necessary and sufficient condition for a pooling equilibrium ($\rho(A, B) = 1$) is $\lambda \geq \frac{2\gamma^L - 1}{r^p(A,A)\triangle\tau}$. To see this, note that beliefs off the equilibrium path are pinned down by D1 from Cho and Kreps (1987) as $r(A, B) = 0$, since a leader with (A, B) has the most to gain from a deviation from this pooling equilibrium. The surviving equilibrium is unique. To see necessity, notice that (A, B) plays A if and only if the inequality is satisfied. Sufficiency comes from the fact that if (A, B) has no incentive to deviate, then (A, A) has even less incentive to deviate.

b) A necessary and sufficient condition for a separating equilibrium ($\rho(A, B) = 0$) is $\lambda \leq \frac{2\gamma^L - 1}{\triangle\tau}$. Beliefs from Bayes' rule are $r(A, A) = 1$ and $r(A, B) = 0$. To see necessity, note that (A, B) plays B for sure if and only if the inequality is satisfied. To see sufficiency, note that if (A, B) has no incentive to deviate, neither does (A, A).

c) A necessary and sufficient condition for a semi-separating equilibrium ($\rho(A, B) \in (0, 1)$) is $\lambda \in \left(\frac{2\gamma^L - 1}{\triangle\tau}, \frac{2\gamma^L - 1}{r^p(A,A)\triangle\tau}\right)$. If $(A, B)$is willing to randomize, it must be the case that

$$\lambda r(A, A) \triangle\tau = 2\gamma^L - 1.$$

Given (1.12) and p^*, this pins down

$$\rho(A, B) = \frac{\bar{t}}{1 - \bar{t}} \frac{1 - \int_{p^*}^1 (1 - s) g(s) \,\mathrm{d}s}{1 - \int_{p^*}^1 sg(s) \,\mathrm{d}s} \left[\frac{\lambda\triangle\tau - \left(2\gamma^L - 1\right)}{2\gamma^L - 1}\right].$$

This value is unique. A necessary and sufficient condition for this $\rho(A, B)$ to be between 0 and 1 is the condition given above. Sufficiency for an equilibrium follows from the fact that if (A, B) does not want to deviate, then neither does (A, A).
Note that if $\gamma^L < \frac{1}{2}$, then $\rho^*(A, B) = 1$ for all λ.

2. $a_F = B$
There are five possible combinations of strategies given histories (B, A) and (B, B), depending on the value of λ. Some of these may overlap.

a) $\rho(B, A) = \rho(B, B) = 0$
Beliefs on the equilibrium path are $r(B, B) = r^p(B, A)$. D1 from Cho and Kreps pins down $r(B, A) = 1$ since (B, A) has the most to gain from a deviation. A necessary and sufficient condition for an equilibrium in which there is pooling on B is $\lambda \leq \frac{1 - 2\gamma^L}{(1 - r^p(B,A))\triangle\tau}$. To see necessity, note that (B, A) only plays B for sure if the inequality is satisfied. If (B, A) has no incentive to deviate, then neither does (B, B), which provides sufficiency.

b) $\rho(B, A) \in (0, 1)$, $\rho(B, B) = 0$

Bayes' rule requires $r(B,A)=1$. If (B,A) is willing to randomize, it is necessary that

$$\lambda \triangle \tau = \lambda r(B,B)\triangle \tau + 1 - 2\gamma^L.$$

Given (1.15) and p^*, this pins down

$$\rho(B,A) = 1 - \frac{1-\bar{t}}{\bar{t}} \frac{\int_{p^*}^1 sg(s)\,\mathrm{d}s}{\int_{p^*}^1 (1-s)\,g(s)\,\mathrm{d}s} \frac{\lambda \triangle \tau - \left(1-2\gamma^L\right)}{1-2\gamma^L}.$$

This value is unique. A necessary and sufficient condition for this $\rho(B,A)$ to be between 0 and 1 is $\lambda \in \left(\frac{1-2\gamma^L}{\triangle \tau}, \frac{1-2\gamma^L}{(1-r^p(B,A))\triangle \tau}\right)$. Sufficiency for an equilibrium comes from the fact that if (B,A) is willing to randomize and does not want to deviate, then (B,B) strictly prefers B and has no incentive to deviate.

c) $\rho(B,A)=1$, $\rho(B,B)=0$
Beliefs are $r(B,A)=1$ and $r(B,B)=0$. A necessary and sufficient condition for a fully separating equilibrium is $\lambda \in \left(\frac{1-2\gamma^L}{\triangle \tau}, \frac{1}{\triangle \tau}\right)$. To see necessity, see that (B,A) playing A for sure given these beliefs requires

$$\lambda \triangle \tau > 1 - 2\gamma^L$$

and that (B,B) playing B for sure requires

$$\lambda \triangle \tau < 1.$$

Sufficiency follows from the fact that if λ is in this range, then neither (B,A) nor (B,B) has a profitable deviation.

d) $\rho(B,A)=1$, $\rho(B,B)\in(0,1)$
Bayes' rule determines $r(B,B)=0$. If (B,B) is willing to randomize, it must be the case that

$$\lambda r(B,A)\triangle \tau = 1.$$

Given (1.14) and p^*, this pins down

$$\rho(B,B) = \frac{\bar{t}}{1-\bar{t}} \frac{\int_{p^*}^1 (1-s)\,g(s)\,\mathrm{d}s}{\int_{p^*}^1 sg(s)\,\mathrm{d}s} (\lambda \triangle \tau - 1).$$

This value is unique. A necessary and sufficient condition for this $\rho(B,B)$ to be between 0 and 1 is $\lambda \in \left(\frac{1}{\triangle \tau}, \frac{1}{r^p(B,A)\triangle \tau}\right)$. Sufficiency for an equilibrium comes from the fact that if (B,B) is willing to randomize and does not want to deviate, then (B, A) strictly prefers A and has no incentive to deviate.

e) $\rho(B, A) = \rho(B, B) = 1$
Beliefs on the equilibrium path are $r(B, A) = r^p(B, A)$. D1 from Cho and Kreps pins down $r(B, B) = 0$ since (B, B) has the most to gain from a deviation. (B, B) plays A for sure if and only if $\lambda \geq \frac{1}{r^p(B,A)\triangle\tau}$. This is sufficient for an equilibrium since if (B, B) has no incentive to deviate, then neither does a leader with (B, A).

The rest of the proposition follows from considering all possible values of λ and pairing the equilibrium strategies of the leader following the possible histories given that value of λ (for all values of γ^L). For $\gamma^L < \frac{1}{2}$, multiple equilibria are possible for some values of λ due to the behavior of (B, A). When this is the case, I select the equilibrium least likely to induce the firm to publish. When $\gamma^L \geq \frac{1}{2}$, $\rho^*(B, A) = 1$ for all λ, so since the behavior of (A, B) and (B, B) is unique (given λ), the equilibrium is unique.■

Proof of Proposition 4.

Backwards inducting from period 3, the leader's behavior is as described in Proposition 3. To see that an equilibrium exists in the follower's subgame, recall the equilibrium condition

$$\frac{\bar{t}(1-p)}{\left(1-\bar{t}\right)p} = \begin{cases} \frac{1-\rho(A,B)-\rho(B,B)+\mu[\pi(B;B,\sigma)-\pi(A;B,\sigma)]}{1+\mu[\pi(A;A,\sigma)-\pi(B;A,\sigma)]} & \gamma^L \geq \frac{1}{2} \\ \frac{-\rho(B,B)+\mu[\pi(B;B,\sigma)-\pi(A;B,\sigma)]}{\rho(B,A)+\mu[\pi(A;A,\sigma)-\pi(B;A,\sigma)]} & \gamma^L < \frac{1}{2} \end{cases}$$

Consider first $\gamma^L \geq \frac{1}{2}$. Section 1.5.1 showed that $\rho^*(A, B)$ and $\rho^*(B, B)$ decrease in p^*, so together with the results from Levy (2004), the right hand side of the equilibrium condition increases in p^*. The left hand side decreases with p^* from something greater than 1 to 0. Therefore, if a p^* exists that equates the two sides, it must be when $s_F^* = B$.

To see how such a p^* differs from p', consider the case of $p^* = p'$. Since $\rho^*(A, B) \geq 0$ and $\rho^*(B, B) \geq 0$, There is only one instance in which indifference is maintained at p'. This is when $\rho^*(A, B) = \rho^*(B, B) = 0$, which happens when λ is sufficiently low (see Proposition 3). If $\rho^*(A, B)$ and/or $\rho^*(B, B)$ are greater than 0, then indifference is broken at the cutoff point p'. The right hand side is now lower, which means that a follower of type (B, p') would strictly prefer action A. Since the left hand side decreases with p, this implies a higher cutoff value to restore indifference.

Hence, $p^* > p'$. By the same reasoning, $p^{**} > p''$ when $\gamma^L < \frac{1}{2}$.

As in Proposition 2, p^* and p^{**} decrease in μ since the right hand side of the equilibrium condition increases in μ, implying lower cutoff values. To see that p^* and p^{**} increase in λ, consider a value $\tilde{p}$ that equalizes the two sides of the equilibrium condition. Recall that for a given p^*, $\rho\left(a_F, s_2\right)$ increases in λ. So regardless of γ^L, an increase in λ reduces the right hand side, breaking indifference. This requires a higher cutoff $p > \tilde{p}$ to restore indifference at the higher value of λ. In the event of no indifference (for example if λ is quite large relative to μ), the follower would always follow the leader and $p^* = 1$. ■

Proof of Proposition 5.

The proof proceeds via a number of lemmas.

If $\lambda \in \left(\frac{2\gamma^L-1}{\triangle\tau}, \frac{1}{\triangle\tau}\right]$, then $\rho^*\left(B,B\right) = 0$, and the firm's condition is

$$\frac{\gamma}{1-\gamma} < \frac{\int_{p'}^{p^*}\left(\bar{t}-s\right)g\left(s\right)\mathrm{d}s + \left(1-\bar{t}\right)\rho^*\left(A,B\right)\left(1-\int_{p^*}^{1} sg\left(s\right)\mathrm{d}s\right)}{\left(1-\bar{t}\right)\rho^*\left(A,B\right)\left(1-\int_{p^*}^{1} sg\left(s\right)\mathrm{d}s\right)}$$

Substituting in the expression for $\rho\left(A,B\right)$ from (1.5) and solving, I get that this expression is equivalent to

$$\frac{\gamma}{1-\gamma} < \frac{\int_{p'}^{p^*}\left(\bar{t}-s\right)g\left(s\right)\mathrm{d}s + \bar{t}\left(1-\int_{p^*}^{1}\left(1-s\right)g\left(s\right)\mathrm{d}s\right)\left(\frac{\lambda\triangle\tau-\left(2\gamma^L-1\right)}{2\gamma^L-1}\right)}{\bar{t}\left(1-\int_{p^*}^{1}\left(1-s\right)g\left(s\right)\mathrm{d}s\right)\left(\frac{\lambda\triangle\tau-\left(2\gamma^L-1\right)}{2\gamma^L-1}\right)}.$$

Lemma 5a. Let assumption 1 hold. For all $p^* > p'$, and all $\lambda \in \left(\frac{2\gamma^L-1}{\triangle\tau}, \frac{1}{\triangle\tau}\right]$, there exists $\gamma_\lambda \geq \frac{1}{2}$ such that $\gamma \in [0, \gamma_\lambda]$ implies the firm (weakly) prefers to publish the mission statement. γ_λ decreases with λ.

Proof of Lemma 5a. The right hand side of the inequality is greater than or equal to 1 for all $p^* > p'$, and all $\lambda \in \left(\frac{2\gamma^L-1}{\triangle\tau}, \frac{1}{\triangle\tau}\right]$ provided $\gamma^L \geq \frac{1}{2}$ and assumption 1 holds.

Define γ_λ as the value of $\gamma \geq \frac{1}{2}$ such that

$$\frac{\gamma_\lambda}{1-\gamma_\lambda} = \frac{\int_{p'}^{p^*}\left(\bar{t}-s\right)g\left(s\right)\mathrm{d}s + \bar{t}\left(1-\int_{p^*}^{1}\left(1-s\right)g\left(s\right)\mathrm{d}s\right)\left(\frac{\lambda\triangle\tau-\left(2\gamma^L-1\right)}{2\gamma^L-1}\right)}{\bar{t}\left(1-\int_{p^*}^{1}\left(1-s\right)g\left(s\right)\mathrm{d}s\right)\left(\frac{\lambda\triangle\tau-\left(2\gamma^L-1\right)}{2\gamma^L-1}\right)}$$

so that a firm of γ_λ is indifferent between publishing and not publishing. Then, all firms with $0 \leq \gamma < \gamma_\lambda$ strictly prefer to publish the mission statement. To see that γ_λ (weakly) decreases in λ, note that the right hand side of the firm's condition decreases with λ, corresponding to a lower γ that renders the firm indifferent between publishing and not.$\square$

If $\frac{1}{\triangle\tau} < \lambda < \frac{2\gamma^L-1}{r^p(A,A)\triangle\tau}$ (which requires $\gamma^L \geq \frac{1+r^p(A,A)}{2}$), then both $\rho^*(A,B) \in (0,1)$ and $\rho^*(B,B) \in (0,1)$. The firm's condition in this case is $\frac{\gamma}{1-\gamma}$ less than

$$\frac{\int_{p'}^{p^*}\left(\bar{t}-s\right)g\left(s\right)\mathrm{d}s+\bar{t}\left[\left(1-\int_{p^*}^{1}\left(1-s\right)g\left(s\right)\mathrm{d}s\right)\left(\frac{\lambda\triangle\tau+1-2\gamma^L}{2\gamma^L-1}\right)-\int_{p^*}^{1}\left(1-s\right)g\left(s\right)\mathrm{d}s\left(\lambda\triangle\tau-1\right)\right]}{\bar{t}\left[\left(1-\int_{p^*}^{1}\left(1-s\right)g\left(s\right)\mathrm{d}s\right)\left(\frac{\lambda\triangle\tau-(2\gamma^L-1)}{2\gamma^L-1}\right)+\int_{p^*}^{1}\left(1-s\right)g\left(s\right)\mathrm{d}s\left(\lambda\triangle\tau-1\right)\right]}.$$

Lemma 5b. Let assumption 1 hold. For all $p^* > p'$, and all $\lambda \in \left(\frac{1}{\triangle\tau}, \frac{2\gamma^L-1}{r^p(A,A)\triangle\tau}\right)$, there exists $\gamma_\lambda > 0$ such that $\gamma \in [0,\gamma_\lambda]$ implies the firm (weakly) prefers to publish the mission statement.

Proof of Lemma 5b. The right hand side of the firm's condition is always positive if assumption 1 holds, since $\left(1-\int_{p^*}^{1}(1-s)\,g(s)\,\mathrm{d}s\right) > \int_{p^*}^{1}(1-s)\,g(s)\,\mathrm{d}s$ and $\frac{\lambda\triangle\tau-\left(2\gamma^L-1\right)}{2\gamma^L-1} > \lambda\triangle\tau-1$ for $\gamma^L \geq \frac{1}{2}$. Then the reasoning from Lemma 5a applies.$\square$

This is the only case for which I have not proved that γ_λ decreases in λ.

If $\lambda \in \left(\max\left\{\frac{1}{\triangle\tau}, \frac{2\gamma^L-1}{r^p(A,A)\triangle\tau}\right\}, \frac{1}{r^p(B,A)\triangle\tau}\right)$, then $\rho^*(A,B) = 1$ and $\rho^*(B,B) \in (0,1)$. The condition for the firm publishing the mission statement is (1.7) which simplifies to the following after substitution using (1.6)

$$\frac{\gamma}{1-\gamma} < \frac{\int_{p'}^{p^*}\left(\bar{t}-s\right)g\left(s\right)\mathrm{d}s+\left(1-\bar{t}\right)\left(1-\int_{p^*}^{1}sg\left(s\right)\mathrm{d}s\right)-\bar{t}\int_{p^*}^{1}\left(1-s\right)g\left(s\right)\mathrm{d}s\left(\lambda\triangle\tau-1\right)}{\left(1-\bar{t}\right)\left(1-\int_{p^*}^{1}sg\left(s\right)\mathrm{d}s\right)+\bar{t}\int_{p^*}^{1}\left(1-s\right)g\left(s\right)\mathrm{d}s\left(\lambda\triangle\tau-1\right)}.$$

Lemma 5c. Let assumption 1 hold. One of the following is true:

1. For all $p^* > p'$, and all $\lambda \in \left(\max\left\{\frac{1}{\triangle\tau}, \frac{2\gamma^L-1}{r^p(A,A)\triangle\tau}\right\}, \frac{1}{r^p(B,A)\triangle\tau}\right)$, there exists $\gamma_\lambda > 0$ such that $\gamma \in [0,\gamma_\lambda]$ implies the firm (weakly) prefers to publish the mission statement. γ_λ decreases with λ.

2. There exists a $\bar{\lambda} \in \left(\max\left\{\frac{1}{\triangle\tau}, \frac{2\gamma^L-1}{r^p(A,A)\triangle\tau}\right\}, \frac{1}{r^p(B,A)\triangle\tau}\right)$ such that for all $p^* > p'$, and all $\lambda \in \left(\max\left\{\frac{1}{\triangle\tau}, \frac{2\gamma^L-1}{r^p(A,A)\triangle\tau}\right\}, \bar{\lambda}\right)$, there exists $\gamma_\lambda > 0$ such that $\gamma \in [0,\gamma_\lambda]$ implies the firm (weakly) prefers to publish the mission statement. γ_λ decreases with λ. If $\lambda \geq \bar{\lambda}$, no type of firm prefers to publish.

Proof of Lemma 5c. The right hand side of the firm's condition decreases with λ. If the right hand side is positive, there exists a $\gamma_\lambda > 0$ that equalizes the two sides. Then, for all $\gamma < \gamma_\lambda$, the firm strictly prefers to publish. If $\max\left\{\frac{1}{\triangle\tau}, \frac{2\gamma^L-1}{r^p(A,A)\triangle\tau}\right\} = \frac{1}{\triangle\tau}$, the right hand side is maximally $\frac{\int_{p'}^{p^*}(\bar{t}-s)g(s)\mathrm{d}s+(1-\bar{t})\left(1-\int_{p^*}^1 sg(s)\mathrm{d}s\right)}{(1-\bar{t})\left(1-\int_{p^*}^1 sg(s)\mathrm{d}s\right)}$, which is positive for all p^* given assumption 1. If $\max\left\{\frac{1}{\triangle\tau}, \frac{2\gamma^L-1}{r^p(A,A)\triangle\tau}\right\} = \frac{2\gamma^L-1}{r^p(A,A)\triangle\tau}$, then the right hand side is maximally $\frac{\int_{p'}^{p^*}(\bar{t}-s)g(s)\mathrm{d}s+(1-\bar{t})\left(1-\int_{p^*}^1 sg(s)\mathrm{d}s\right)-\bar{t}\int_{p^*}^1(1-s)g(s)\mathrm{d}s\left(\frac{2\gamma^L-1}{r^p(A,A)}-1\right)}{(1-\bar{t})\left(1-\int_{p^*}^1 sg(s)\mathrm{d}s\right)+\bar{t}\int_{p^*}^1(1-s)g(s)\mathrm{d}s\left(\frac{2\gamma^L-1}{r^p(A,A)}-1\right)}$, which is minimal when $\gamma^L = 1$. In this case, the condition becomes

$$\frac{\int_{p'}^{p^*}\left(\bar{t}-s\right)g\left(s\right)\mathrm{d}s+\left(1-\bar{t}\right)\left(1-\int_{p^*}^1 sg\left(s\right)\mathrm{d}s\right)-\bar{t}\int_{p^*}^1\left(1-s\right)g\left(s\right)\mathrm{d}s\frac{1-\bar{t}}{\bar{t}}\frac{\left(1-\int_{p^*}^1 sg(s)ds\right)}{\left(1-\int_{p^*}^1(1-s)g(s)ds\right)}}{\left(1-\bar{t}\right)\left(1-\int_{p^*}^1 sg\left(s\right)\mathrm{d}s\right)+\bar{t}\int_{p^*}^1\left(1-s\right)g\left(s\right)\mathrm{d}s\frac{1-\bar{t}}{\bar{t}}\frac{\left(1-\int_{p^*}^1 sg(s)ds\right)}{\left(1-\int_{p^*}^1(1-s)g(s)ds\right)}}$$

$$=\frac{\int_{p'}^{p^*}\left(\bar{t}-s\right)g\left(s\right)\mathrm{d}s+\left(1-\bar{t}\right)\left(1-\int_{p^*}^1 sg\left(s\right)\mathrm{d}s\right)\frac{1-2\int_{p^*}^1(1-s)g(s)ds}{1-\int_{p^*}^1(1-s)g(s)ds}}{\left(1-\bar{t}\right)\left(1-\int_{p^*}^1 sg\left(s\right)\mathrm{d}s\right)\frac{1}{1-\int_{p^*}^1(1-s)g(s)ds}}$$

which is positive for all p^* since $1 > 2\int_{p^*}^1\left(1-s\right)g\left(s\right)\mathrm{d}s$ and given assumption 1. Then for all $\gamma^L \in \left[\frac{1+r^p(A,A)}{2}, 1\right]$, the right hand side is also positive at its maximum. Recall that γ^L in this range is required for $\max\left\{\frac{1}{\triangle\tau}, \frac{2\gamma^L-1}{r^p(A,A)\triangle\tau}\right\} = \frac{2\gamma^L-1}{r^p(A,A)\triangle\tau}$.

As $\lambda \to \frac{1}{r^p(B,A)\triangle\tau}$, the right hand side of the firm's condition goes to $\frac{\int_{p'}^{p^*}(\bar{t}-s)g(s)\mathrm{d}s+(1-\bar{t})\left(1-2\int_{p^*}^1 sg(s)\mathrm{d}s\right)}{1-\bar{t}}$. If this is positive, then over the whole range of λ in question, low γ types of firm find it profitable to publish the mission statement. If this is negative, then by continuity there is some $\bar{\lambda} \in \left(\max\left\{\frac{1}{\triangle\tau}, \frac{2\gamma^L-1}{r^p(A,A)\triangle\tau}\right\}, \frac{1}{r^p(B,A)\triangle\tau}\right)$ such that $\int_{p'}^{p^*}\left(\bar{t}-s\right)g\left(s\right)\mathrm{d}s+\left(1-\bar{t}\right)\left(1-\int_{p^*}^1 sg\left(s\right)\mathrm{d}s\right)-\bar{t}\int_{p^*}^1\left(1-s\right)g\left(s\right)\mathrm{d}s\left(\bar{\lambda}\triangle\tau-1\right) = 0$. Then for all $\lambda < \bar{\lambda}$, there there exists a $\gamma_\lambda > 0$ that equalizes the two sides, and for all $\gamma < \gamma_\lambda$ the firm strictly prefers to publish. For all $\lambda \geq \bar{\lambda}$, no type of firm publishes the mission statement.

To see that when γ_λ exists it decreases with λ, recall that the right hand side of the firm's condition decreases with λ, implying a lower value of γ_λ to equalize the two sides.$\square$

If $\lambda \geq \frac{1}{r^p(B,A)\triangle\tau}$, then $\rho^*\left(A,B\right) = \rho^*\left(B,B\right) = 1$. The firm's condition becomes

$$\frac{\gamma}{1-\gamma} < \frac{\int_{p'}^{p^*}\left(\bar{t}-s\right)g\left(s\right)\mathrm{d}s+\left(1-\bar{t}\right)\left(1-2\int_{p^*}^1 sg\left(s\right)\mathrm{d}s\right)}{1-\bar{t}}.$$

Lemma 5d. Let $\lambda \geq \frac{1}{r^p(B,A)\triangle\tau}$. For all $p^* > p'$, if $\int_{p'}^{p^*}\left(\bar{t}-s\right)g\left(s\right)\mathrm{d}s+\left(1-\bar{t}\right)\left(1-2\int_{p^*}^{1}sg\left(s\right)\mathrm{d}s\right) > 0$, then there exists $\bar{\gamma} > 0$ such that $\gamma \in [0,\bar{\gamma}]$ implies the firm (weakly) prefers to publish the mission statement. If $\int_{p'}^{p^*}\left(\bar{t}-s\right)g\left(s\right)\mathrm{d}s+\left(1-\bar{t}\right)\left(1-2\int_{p^*}^{1}sg\left(s\right)\mathrm{d}s\right) \leq 0$, then no firm ever chooses to publish.

Proof of Lemma 5d. This follows from the proof of Lemma 5c. □

To see that if $\gamma^L \in \left[\frac{1}{2},\frac{1+r^p(A,A)}{2}\right]$, γ_λ decreases with λ, note that the only time both types (A,B) and (B,B) randomize, making it uncertain whether the right hand side decreases in λ, is when $\max\left\{\frac{1}{\triangle\tau},\frac{2\gamma^L-1}{r^p(A,A)\triangle\tau}\right\} = \frac{2\gamma^L-1}{r^p(A,A)\triangle\tau}$, which requires $\gamma^L > \frac{1+r^p(A,A)}{2}$. Otherwise, the right hand side of (1.7) weakly decreases in λ, corresponding to a decreasing γ_λ that equalizes the two sides. This concludes the proof of Proposition 5.■

Proof of Proposition 6.

Since $p'' < p'$ and $p^{**} < p^*$, if assumption 1 holds, then $\int_{p''}^{p^{**}}\left(\bar{t}-s\right)g\left(s\right)\mathrm{d}s$ is also non-negative.

The proof proceeds in two lemmas.

Lemma 6a. Let assumption 1 hold. For all $p^{**} > p''$ and all $\lambda \in \left(\frac{1-2\gamma^L}{(1-r^p(B,A))\triangle\tau},\frac{1}{r^p(B,A)\triangle\tau}\right)$, there exists a $\gamma_\lambda < 1$ such that $\gamma \in [\gamma_\lambda, 1]$ implies the firm (weakly) prefers to publish the mission statement. γ_λ increases in λ.

Proof of Lemma 6a.

If $\lambda \in \left(\frac{1-2\gamma^L}{(1-r^p(B,A))\triangle\tau},\frac{1}{r^p(B,A)\triangle\tau}\right)$, then by Proposition 3 $\rho^*(B,A) = 1$ and $\rho^*(B,B) \in [0,1)$.

Substituting in for $\rho^*(B,B)$ from (1.6), the condition becomes

$$\frac{\gamma}{1-\gamma} > \begin{cases} \frac{\bar{t}\int_{p^{**}}^{1}(1-s)g(s)\mathrm{d}s}{\int_{p''}^{p^{**}}(\bar{t}-s)g(s)\mathrm{d}s+\bar{t}\int_{p^{**}}^{1}(1-s)g(s)\mathrm{d}s} & \rho^*(B,B)=0 \\ \frac{\bar{t}\int_{p^{**}}^{1}(1-s)g(s)\mathrm{d}s\lambda\triangle\tau}{\int_{p''}^{p^{**}}(\bar{t}-s)g(s)\mathrm{d}s+\bar{t}\int_{p^{**}}^{1}(1-s)g(s)\mathrm{d}s(2-\lambda\triangle\tau)} & \rho^*(B,B)\in(0,1) \end{cases} \tag{1.17}$$

If λ is low enough that $\rho^*(B,B) = 0$, then the first case in (1.17) obtains. This happens if $\frac{1}{2} > \gamma^L \geq \frac{r^p(B,A)}{2}$ and $\lambda \in \left(\frac{1-2\gamma^L}{(1-r^p(B,A))\triangle\tau},\frac{1}{\triangle\tau}\right]$. By assumption 1, the right hand side is positive and less than 1. Define $\underline{\gamma}$ as the value of γ that equalizes the two sides of the firm's condition. $\underline{\gamma} < \frac{1}{2}$ since the right hand side is less than 1.

Then, all firms with $\gamma > \underline{\gamma}$ strictly prefer to publish the mission statement.

For intermediate values of λ $\left(\lambda \in \left(\max\left\{\frac{1-2\gamma^L}{(1-r^p(B,A))\triangle\tau}, \frac{1}{\triangle\tau}\right\}, \frac{1}{r^p(B,A)\triangle\tau}\right)\right)$ such that $\rho(B,B) \in (0,1)$, the right hand side of (1.17) increases with λ. As $\lambda \to \frac{1}{r^p(B,A)\triangle\tau}$, the right hand side of the inequality goes to $\frac{\bar{t}\int_{p^{**}}^{1}(1-s)g(s)\mathrm{d}s+(1-\bar{t})\int_{p^{**}}^{1}sg(s)\mathrm{d}s}{2\int_{p''}^{p^{**}}(\bar{t}-s)g(s)\mathrm{d}s}$, which is positive by assumption 1 and finite. For each value of λ in this range, let γ_λ be the value of γ that makes the firm indifferent between publishing and not. Then, $\gamma > \gamma_\lambda$ implies the firm strictly prefers to publish the mission statement. γ_λ increases in λ since, again, the right hand side of the firm's condition increases with λ, implying a higher value of γ that equalizes the two sides of the inequality.$\square$

Lemma 6b. Let assumption 1 hold. For all $p^{**} > p''$ and all $\lambda \geq \frac{1}{r^p(B,A)\triangle\tau}$, there exists a $\overline{\gamma} < 1$ such that $\gamma \in [\overline{\gamma}, 1]$ implies the firm (weakly) prefers to publish the mission statement.

Proof of Lemma 6b.

This follows from the proof of Lemma 6a and the fact that when $\rho(B,B) = 1$ the right hand side of (1.17) is $\frac{\bar{t}\int_{p^{**}}^{1}(1-s)g(s)\mathrm{d}s+(1-\bar{t})\int_{p^{**}}^{1}sg(s)\mathrm{d}s}{2\int_{p''}^{p^{**}}(\bar{t}-s)g(s)\mathrm{d}s}$, which is positive and finite. $\square$

To see that γ_λ weakly increases with λ note that when λ is low, $\gamma_\lambda = \underline{\gamma}$. As λ increases, γ_λ increases as shown in Lemma 6a. Then, as λ tops $\frac{1}{r^p(B,A)\triangle\tau}$, $\gamma_\lambda = \overline{\gamma}$. This concludes the proof of Proposition 6.$\blacksquare$

Proof of Proposition 7.

If $\int_{p'}^{p^\alpha}\left(\bar{t}-s\right)g(s)\,\mathrm{d}s + \alpha\left(1-\bar{t}\right)\left(1-2\int_{p^\alpha}^{1}sg(s)\,\mathrm{d}s\right) > 0$, then define γ_α as the value of $\gamma \in (0,1)$ that satisfies

$$\frac{\gamma_\alpha}{1-\gamma_\alpha} = \frac{\int_{p'}^{p^\alpha}\left(\bar{t}-s\right)g(s)\,\mathrm{d}s + \alpha\left(1-\bar{t}\right)\left(1-2\int_{p^\alpha}^{1}sg(s)\,\mathrm{d}s\right)}{\alpha\left(1-\bar{t}\right)}.$$

A firm with preferences γ_α is indifferent between employing policies that produce $\alpha > 0$ and not. This implies that all firms with $\gamma < \gamma_\alpha$ strictly prefer the conservatism. If the reverse inequality holds, then (1.8) is not satisfied for any $\gamma \geq 0$.$\blacksquare$

Proof of Proposition 8.

If $\int_{p''}^{p^{\alpha\alpha}}\left(\bar{t}-s\right)g(s)\,\mathrm{d}s+\bar{t}(1-\alpha)\int_{p^{\alpha\alpha}}^{1}(1-s)\,g(s)\,\mathrm{d}s-\left(1-\bar{t}\right)\alpha\int_{p^{\alpha\alpha}}^{1}sg(s)\,\mathrm{d}s > 0$, then let γ_α be the value of γ that equalizes the two sides of (1.9). A firm with preferences γ_a is indifferent between inducing conservatism of degree α and not. This implies that all firms with $\gamma > \gamma_\alpha$ strictly prefer to induce. If $\int_{p''}^{p^{\alpha\alpha}} \left(\bar{t} - s\right) g\left(s\right) \mathrm{d}s + \bar{t}\left(1-\alpha\right) \int_{p^{\alpha\alpha}}^{1} \left(1-s\right) g\left(s\right) \mathrm{d}s - \left(1-\bar{t}\right) \alpha \int_{p^{\alpha\alpha}}^{1} sg\left(s\right) \mathrm{d}s < 0$, then the right hand side of (1.9) is negative, and all types of firm ($\gamma \in [0,1]$) prefer to induce conservatism. Only if $\int_{p''}^{p^{\alpha\alpha}} \left(\bar{t} - s\right) g\left(s\right) \mathrm{d}s + \bar{t}\left(1-\alpha\right) \int_{p^{\alpha\alpha}}^{1} \left(1-s\right) g\left(s\right) \mathrm{d}s - \left(1-\bar{t}\right) \alpha \int_{p^{\alpha\alpha}}^{1} sg\left(s\right) \mathrm{d}s = 0$ does the right hand side grow unboundedly large, in which case no firm would choose to induce α-conservatism.■

Proof of Proposition 9.

If $E\left[t\right] > E\left[p\right]$, then the left hand side of (1.11) is greater than 1. For any values of β, α, and γ the right hand side of each of these equations is less than 1, satisfying the inequalities.■

1.9 Chapter 1 References

1. Banks, Jeffrey and Joel Sobel. 1987. Equilibrium Selection in Signaling Games. *Econometrica* 55 (May): 647-661.

2. Blanes i Vidal, Jordi and Marc Möller. 2007. When Should Leaders Share Information with Their Subordinates? *Journal of Economics & Management Strategy* 16 (Summer): 251-283.

3. Bolton, Patrick, Markus Brunnermeier and Laura Veldkamp. 2011. Leadership, Coordination and Corporate Culture. Working Paper (August).

4. Boot, Arnoud. 1992. Why Hang on to Losers? Divestitures and Takeovers. *Journal of Finance* 47(September): 1401-1423.

5. Camerer, Colin and Dan Lovallo. 1999. Overconfidence and excess entry: An experimental approach. *The American Economic Review* 89 (March): 306-318.

6. Cho, In-Koo and David Kreps. 1987. Signaling Games and Stable Equilibria. *The Quarterly Journal of Economics* 102 (May): 179-221.

7. Englmaier, Florian. 2004. A Strategic Rationale for Having Overconfident Managers. Working Paper (December).

8. Fama, Eugene. 1980. Agency Problems and the Theory of the Firm. *Journal of Political Economy* 88 (April): 288-307.

9. Ferreira, Daniel and Marcelo Rezende. 2007. Corporate Strategy and Information Disclosure. *RAND Journal of Economics* 38 (Spring): 164-184.

10. Goel, Anand and Anjan Thakor. 2008. Overconfidence, CEO Selection and Corporate Governance. *Journal of Finance* 63(December): 2737-2784.

11. Hart, Oliver and Bengt Holmström. 2002. A Theory of Firm Scope. MIT Department of Economics Working Paper No. 02-42.

12. Holmström, Bengt. 1999. Managerial Incentive Problems: A Dynamic Perspective. *The Review of Economic Studies* 66 (January): 169-182. Special Issue: Contracts.

13. Kanodia, Chandra, Robert Bushman and John Dickhaut. 1989. Escalation

Errors and the Sunk Cost Effect: An Explanation Based on Reputation and Information Asymmetries. *Journal of Accounting Research* 27 (Spring): 59-77.

14. Komai, Mana, Mark Stegeman and Benjamin Hermalin. 2007. Leadership and Information. *American Economic Review* 97 (June): 944-947.

15. Lazear, Edward and Sherwin Rosen. 1981. Rank-order Tournaments as Optimum Labor Contracts. *Journal of Political Economy* 89 (October): 841-864.

16. Leaver, Clare. 2009. Bureaucratic minimal squawk behavior: Theory and evidence from regulatory agencies. *American Economic Review* 99 (June): 572-607.

17. Levy, Gilat. 2004. Anti-herding and Strategic Consultation. *European Economic Review* 48 (June): 503-525.

18. Malmendier, Ulrike and Geoffrey Tate. 2005. CEO overconfidence and corporate investment. *The Journal of Finance* 60 (December): 2661-2700.

19. _______. 2008. Who makes acquisitions? CEO overconfidence and the market's reaction. *Journal of Financial Economics* 89 (July): 20-43.

20. Ottaviani, Marco and Peter Sørensen. 2000. Herd Behavior: A Comment. *American Economic Review* 90 (June): 695-704.

21. Prendergast, Canice and Lars Stole. 1996. Impetuous Youngsters and Jaded Old-timers: Acquiring a Reputation for Learning. *Journal of Political Economy* 104 (December): 1105-1134.

22. Rotemberg, Julio and Garth Saloner. 2000. Visionaries, Managers and Strategic Direction. *RAND Journal of Economics* 31 (Winter): 693-716.

23. Scharfstein, David and Jeremy Stein. 1990. Herd Behavior and Investment. *American Economic Review* 80 (June): 465-479.

24. Van den Steen, Eric. 2005. Organizational Beliefs and Managerial Vision. *Journal of Law, Economics, and Organization* 21 (1): 256-283.

25. Zwiebel, Jeffrey. 1995. Corporate Conservatism and Relative Compensation. *Journal of Political Economy* 103(February): 1-25.

Chapter 2

Re-election of Corrupt Politicians

2.1 Introduction

In September 2002, William Braker, a member of the Hudson County Board of Freeholders in New Jersey, was publicly implicated in an extortion scandal. He resigned his elected office, only to rescind his resignation later in compliance with "the wishes of his supporters...who had urged him to remain in office".[1] Two months later he was re-elected to the Board, where he served until his conviction for taking bribes in 2004.

This paper presents a model of corruption and re-election, and attempts to provide an explanation for why rational voters might vote for a candidate they observe is corrupt, contributing to the persistence of such behavior in elected officials. I use a two-period political agency model with re-election. There are two parts: a baseline game with only rational voting and an extension with campaign activity. In both cases, an incumbent politician (who can be good or bad) observes the outcome of a public project and decides how to report it to the voters (i.e. whether to be truthful about the returns or to lie and free up funds for her personal use). In the extension, an incumbent who lies has the option to spend the funds on campaign activity in the hopes of influencing election results. Voters then choose to re-elect the incumbent or to oust her and elect a challenger, who may have an in-built electoral advantage over the incumbent which is independent of his type. The campaign activity is formulated in terms of campaign spending, but the overall mechanism is meant generally to encompass any way in which a sitting politician can reduce public welfare and by

[1] Smothers (2002).

doing so increase her own consumption and/or her chances of re-election.

I appeal to two explanations for voter tolerance of corruption among elected office-holders:

1. Voters observe that all politicians are potentially corrupt and vote for the best option available.

2. Voters observe that politics is a ‘dirty game’ and are willing to forgive a certain amount of corruption in a politician who will serve them best in the long term.

With frequent emphasis on corruption in government (as opposed to bureaucracy), it is unsurprising that a large section of the literature concerns politician and voter behavior. Starting from the ‘Leviathan hypothesis’ of Brennan and Buchanan (1980), many authors have sought to describe the various ways in which political systems constrain the expropriating power of the government. Myerson (1993) analyzes several electoral systems for their efficacy in keeping corrupt parties out of power. Besley and Case (1995) and Ferraz and Finan (2008) show that the prospect of re-election can provide discipline, reducing the amount of public wealth extracted by government officials. Indeed, one of the most common arguments in favor of democracy is that it gives citizens the ability to revisit their choice of elected official every so often to remove unsatisfactory ones. Given this political setup, and given that voters dislike corruption, it remains something of a paradox that so many corrupt politicians thrive, facing election after election victoriously (and oftentimes with quite comfortable margins). It is this seeming inconsistency that is addressed in this paper.

The puzzling re-election of corrupt officials received a significant amount of attention from political scientists in the 1970s, resulting in three basic explanations.[2] Firstly, that voter ignorance about corrupt acts leads them to vote for corrupt politicians. Since most politicians obviously would wish to hide any corrupt dealings, perhaps this can account for some otherwise inexplicable voter behavior. However, this explanation does not address cases of egregious corruption which were well-known and yet seemingly disregarded at the polls.[3] Secondly, that voters will elect corrupt politicians in exchange for favors. This explanation is unsatisfactory simply

[2] Rundquist et al. (1977).

[3] A good example is that of Rickey Peete, former Memphis, TN City Council member, who was elected to the Council in the late 1980s, was convicted of taking bribes, and served a two-year prison sentence. On release, he was elected again to the Council in 1995 and re-elected twice more. He was again convicted of taking bribes in exchange for votes in 2007 and is currently serving his sentence (Dries 2007).

because it is unlikely that voters are being materially rewarded in sufficient numbers to explain electoral victories. Finally, that voters implicitly trade-off corruption with other traits, such as position on certain issues or party affiliation.

The issue was taken up in the economics literature in the 1990s, starting in earnest with the Myerson (1993) paper. Much of this work centers on coordination failure and preference heterogeneity as the reasons for inefficiency in democracies.[4] A notable departure from the voter-side emphasis is Caselli and Morelli (2004), who take a supply-side approach, arguing that re-election of bad (corrupt or incompetent) politicians can be explained by a lack of good politicians willing to hold office. In this case voters simply make the best of bad choices. A simplified variant on this explanation will be addressed in this paper, which suggests that rational voters who observe corruption may opt to vote for the person they see as the less corrupt of two politicians.

One possible explanation which has not (to my knowledge) received much attention in the literature is that of politics as a dirty game. Intuitively, if a certain amount of corruption is necessary for re-election (because it garners a politician either much-needed campaign money or the support of vote-getting groups), and if voters realize this, they may be willing to overlook a certain amount of corruption if they believe the politician in question will ultimately serve their interests. Politicians, counting on this forgiveness, may then decide to act corruptly to ensure re-election, creating a vicious cycle. I also address this possibility.

I find that even good politicians will choose to act corruptly in order to secure re-election, and that rational voters will support them in their bids to stay in office. I also find that, in many cases, both good politicians and voters would be made better off by eliminating the mechanism by which corruption increases re-election chances. Only when the campaign activity provides more information to rational voters than benefit to bad incumbents is it beneficial to leave it in place.

2.2 Related Literature

The two most closely related works are Coate and Morris (1995) and Grossman and Helpman (2001). I follow Coate and Morris in many aspects of the baseline model setup. They present two types of politicians—good and bad—who are charged with the decision of implementing or not implementing a public project. The incumbent

[4] Besley and Coate (1998), for example.

has private information about the public benefit of such a project, and so is in a position to make the socially optimal choice. However, bad politicians also have ties to a special interest, and may implement the project more often than is efficient as an indirect transfer to that interest. I adapt their project to my setting in the following ways. Whereas they give the incumbent the power to decide whether or not the project goes forward, in this model the project is automatically implemented. The only thing the politician must do is observe the outcome and decide how to report it to the citizenry. It is here, rather than at the implementation stage, that the incumbent can act corruptly by extracting funds from the project. Furthermore, where Coate and Morris's bad politician acts inefficiently via an interest group,[5] the bad politician in this model acts inefficiently in an unspecified manner (somehow expropriating money from a successful project).

The re-election structure is also adapted from Coate and Morris. I use the tactic of drawing challengers in the election from a known distribution, which allows an incumbent to forward-estimate her probability of re-election after taking certain actions, given the manner in which voters update their beliefs.[6]

The most important difference between the baseline model and the extension with campaign spending relies on a feature of Grossman and Helpman's (2001) model of electoral competition with campaign spending.[7] They break the electorate up into two blocs: a strategic fraction, which"understand[s] the political environment and the implications of [its] votes",[8] and an impressionable fraction, which is susceptible to campaign advertising. The vote-getting power of campaign spending is left in reduced form.[9] In Grossman and Helpman's model, the two fractions actually correspond to two distinct groups of voters, both of whom act according to a probabilistic voting model, only differing in how they make their decisions. In this model, there is only one voter. Therefore the two different kinds of voter correspond to two potential types he could assume, and the composition of the electorate in Grossman and Helpman corresponds here to the prior probability he will be realized as one or the other. This is primarily for ease of computation, but it does have some real-world analogues. It could, for example, be interpreted as uncertainty on the politician's part as to what voters will base their decisions on in the election—legislative record,

[5]Coate and Morris remain technically agnostic as to how exactly helping the interest group increases the politician's utility. However, it could happen in a number of ways. One way which is mentioned in their paper is that the bad politician is susceptible to bribes from the interest group (p. 1216).

[6]For a detailed discussion of similar models, see Besley (2006).

[7]Chapter 10.

[8]Page 321.

[9]Following Baron (1994).

ethnicity, personal charm, etc.[10]

As an important aside, I have couched the action of this extension in terms of campaign spending, which is most often associated with campaign advertising. The results are therefore related to those of Prat (2002) and Coate (2004), who find that aggregate voter welfare is improved under limits on campaign financing, even when voters are fully rational. Coate considers directly informative campaign advertising, in which a candidate truthfully conveys information about himself, but which may be paid for by interest groups as a *quid pro quo* for more favorable policies. Prat takes a similar policy-for-contributions scenario,[11] but gives the information about candidates' quality to the interest group, so that voters correctly perceive candidates who receive larger campaign contributions as higher quality. In both of their papers, reducing or outright banning campaign advertising can improve voter welfare, even though in both cases it provides information.

My model, like Prat's (2002), involves campaign advertising which is indirectly informative to rational voters in equilibrium. However, I do not mean to firmly embed this version of the model in that context. The incumbent's choice is meant to stand as a metaphor for any way a politician can act that decreases public benefits but gives her the possibility of increasing her own utility and/or her chances of being re-elected. This is partly to highlight the supposition that politicians can be corrupt at many points during their incumbency, not just at re-election time when they may 'sell' influence to interest groups, and partly to leave in sufficient flexibility for later extensions. To give an example, it could be that the incumbent allows a firm to over-invoice on a public contract in exchange for kick-backs and/or campaign contributions, but it could also be the case that she puts forward a socially sub-optimal transportation bill which earns her the support of transport workers' unions in the next election cycle.

The fact that the challenger automatically has funds to spend is therefore important. The origin of the funds is left unspecified, but they could represent either a large campaign 'war-chest' (say from wealthy constituents) or dedicated support from a portion of the population (perhaps for reasons of ethnicity). It is this disadvantage that the incumbent must decide to rectify or not.

[10] By way of illustration, many feel the 2004 U.S. Presidential election was decided on "moral values", as approximately 4 million evangelical Christians turned out to vote. This was unexpected by most political observers, and probably also by the challenger, Sen. John Kerry (see Egan (2004)).

[11] For an early example of the trade-off between campaign contributions and policy choices with rational voters, see Austen-Smith (1987).

2.3 Benchmark

2.3.1 Model

2.3.1.1 The policy

In each period a public project is automatically implemented. The social benefit to this project, B, can be either high or low—$B \in \{H, L\}$, $H > L > 0$. The exogenous probability that the social benefits are high is denoted β. The incumbent observes the value of B and then decides what to report to voters. She selects an announcement $A \in \{H, L\}$. If the return is low, the incumbent must report L to the voters; she cannot pretend the outcome was better than it was. If $B = H$, however, the incumbent has a choice. She can choose to report H, in which case the full benefits of the project accrue to the voters, or she can choose to report L, in which case the voters receive the lower value of benefits. In this way, the voters' utility depends only on the incumbent's decision and the exogenous probability of a high return β.

Reporting L when in fact $B = H$ frees up an amount b for the incumbent's use. This act could represent a number of potentially corrupt activities, such as over-invoicing for goods and services or diverting public funds for personal gain.[12] The important feature is that the incumbent can make a decision which lowers voter utility and gives her access to funds. In this part of the model, it is assumed that the only use for diverted funds is personal. That is, an incumbent who extracts b from the project can only use it to increase her personal consumption.

2.3.1.2 The citizenry

There is a single representative citizen/voter who receives utility from the public project. The voter receives $u_c = u(H)$ when the incumbent reports H, and $u_c = u(L)$ when the incumbent reports L. $u(H) > u(L)$. At the end of the first period, he decides which candidate, incumbent or challenger, to vote for in the election.

[12] For an example see Reinikka and Svensson (2004). An alternative way of seeing the setup is that public wealth is exogenously determined (say by tax revenue), but only the incumbent knows how much is collected. If the public wealth is passed on to voters in the form of benefits, it may be possible for the incumbent to expropriate some of the public wealth and pass on fewer benefits to the citizenry.

2.3.1.3 The politicians

There are two types of politician: good (g), and bad (b). Both types have discount rate δ, and they receive no utility if they are not in office.[13] A good politician cares about the public benefit her announcement decision nets the citizenry. Her utility per period is $v_g(A)$, $A \in \{H, L\}$, where $v_g(H) > v_g(L)$.

A bad politician, on the other hand, cares about the amount of money she can expropriate for her personal use. Her utility is $v_b(b)$ when she lies about the project's return and uses all the freed-up funds herself, and $v_b(0)$ (which can be thought of as ego-rent) either if she does not lie about the project's return or if the return really is low ($B = L$), in which case it is not possible to expropriate anything. $v_b(\cdot)$ is increasing in its argument, for example $v_b(b) > v_b(0)$.

Note that $v_g(\cdot)$ and $v_b(\cdot)$ are defined on different spaces—v_g on the amount of public good provided to the voter, v_b on expropriated funds from the project.

2.3.1.4 The information structure

The citizen is hampered in his decision-making by two dimensions of uncertainty. First, he is uncertain about the actual return to the public project, as he is unable to observe B directly. This implies a certain lack of transparency in the political system, since I assume that the voter's only information about the project's outcome comes from the incumbent politician. It is this uncertainty that enables a bad politician to expropriate public funds for his own use, and 'hide' behind the possibility (as far as the voter is concerned) of a low return to the project.

Secondly, the citizen is uncertain about the incumbent's type. He has no direct information on whether the incumbent is good or bad. I assume, however, that the voter does have some, albeit imperfect, information about the incumbent prior to the action of the game. Specifically, the voter observes a signal $\lambda_i \in (0, 1)$ that corresponds to the prior probability of the incumbent's being 'good' at the beginning of the first period. One could think of λ_i as the incumbent's reputation upon assuming office. The challenger who runs against the incumbent in the election at the end of the first period also has a reputation, denoted λ_c, which is realized after the action of the first period, right before the votes are cast. I assume that λ_c

[13]This assumption follows Maskin and Tirole (2004, p. 1035): "...it is not enough for the official that great things be done; she wants to be the one who does them."

is drawn from a cumulative distribution function $G(\lambda)$, which is smooth and strictly increasing, satisfying the property $G(0) = 0$ and $G(1) = 1$.[14]

2.3.1.5 The game

This game is played out over two periods between the citizen and the two politicians (incumbent and challenger). At the start of the game, Nature selects the incumbent's type (b or g) and the outcome of the public project (H or L). The voter receives an imperfect signal, λ_i, of the incumbent's type. The incumbent observes the outcome B, and if $B = H$ decides on her announcement $A \in \{H, L\}$ and receives utility accordingly. The citizen observes the incumbent's announcement and receives utility according to A. The incumbent's record after the first period is A.

At the end of the first period, the challenger is realized. Nature determines the challenger's type (b or g), and the voter receives an imperfect signal of that type λ_c, representing the probability he assigns to the challenger's being good. The signal is drawn from the cumulative distribution function $G(\lambda)$. The voter uses λ_i, λ_c, and the incumbent's first-period record to decide which candidate to elect. The winner of the election takes office in the second period, observes the outcome of the new public project, and makes her reporting decision. Payoffs are realized, and the game ends here.

The incumbent has two relevant decisions to make that comprise her strategy. First, her strategy must specify for each type of incumbent what to report in the first period in the event $B = H$. Second, the strategy must specify for each type her reporting decision in the second period following re-election.

The challenger's strategy has only one component—a rule specifying his reporting decision in the event he is elected and $B = H$. This will depend only on his type, since the game ends after he is called on to act.

The voter's strategy is a rule specifying the probability of re-electing the incumbent given the incumbent's record A, her initial reputation λ_i, and the challenger's initial reputation λ_c. The citizen's strategy will depend on his beliefs about the incumbent's type. These beliefs are determined by the incumbent's reputation and her record at the end of the first period.

The solution concept is perfect Bayesian equilibrium (PBE). A PBE of this game

[14] This structure of voter information is the same as in Coate and Morris (1995).

will consist of a strategy for the incumbent, a strategy for the challenger, beliefs for the citizen, and a strategy for the citizen. The citizen's beliefs must be consistent. That is, for every action the incumbent takes with positive probability, the citizen derives his beliefs on the incumbent's type using Bayes' rule. Furthermore, all strategies must be optimal given the strategies (and possibly beliefs) of the other players.

2.3.2 Equilibrium

In this section I solve for the equilibrium of the game laid out above. Equilibrium is solved for by backward induction.

2.3.2.1 Second period behavior of politicians

An incumbent who is re-elected to office in the second period will act depending on her type. If she is good, she will report $A = H$ when the return to the project is high since $v_g(H) > v_g(L)$. When it is low she can only report truthfully. Her second-period expected utility is $\Delta v_g \equiv \beta v_g(H) + (1-\beta) v_g(L)$. A bad incumbent, on the other hand, will expropriate as much as possible from the project, reporting $A = L$ no matter what. His second-period expected utility will therefore be $\Delta v_b \equiv \beta v_b(b) + (1-\beta) v_b(0)$. If the challenger is elected at the end of period one, he will follow the same strategy as the incumbent——tell the truth if good, lie if bad.

2.3.2.2 Citizen's voting behavior

The citizen always prefers to have a good politician in office in the second period (compare $u_c = \beta u(H) + (1-\beta) u(L)$ with a good politician to $u_c = u(L)$ with a bad one). Therefore, he will vote for the candidate he believes is most likely to be good given the information available to him. Let $\lambda(A)$ denote the voter's best estimate (from Bayes' rule) of the probability the incumbent is good when A is her record. Given a challenger with reputation λ_c, the citizen will choose to re-elect the incumbent if and only if $\lambda(A) > \lambda_c$. Recall that the challenger's initial reputation is drawn from the cumulative distribution function $G(\lambda)$. Therefore, the probability of an incumbent's being re-elected is given by $G(\lambda(A))$.

2.3.2.3 First period behavior of incumbent and citizen's beliefs

The only time the incumbent has to make a decision about her report is when $B = H$. It is therefore necessary to specify strategies for good and bad incumbents only in this event. Consider the following payoffs:

$$\begin{aligned} V_g(H) &= v_g(H) + \delta G(\lambda(H)) \Delta v_g && (2.1) \\ V_g(L) &= v_g(L) + \delta G(\lambda(L)) \Delta v_g && (2.2) \\ V_b(H) &= v_b(0) + \delta G(\lambda(H)) \Delta v_b && (2.3) \\ V_b(L) &= v_b(b) + \delta G(\lambda(L)) \Delta v_b && (2.4) \end{aligned}$$

Equations (2.1) and (2.2) give the expected total payoff of a good incumbent who reports H and L, respectively. Equations (2.3) and (2.4) give the expected total payoff of a bad incumbent who reports H and L, respectively.

Clearly the strategy for the incumbent that maximizes her expected payoff depends on the citizen's beliefs. On the equilibrium path, the citizen's beliefs must be determined using Bayes' rule. Off the equilibrium path, after first-period actions that occur with zero probability in equilibrium, the citizen's beliefs are not tied down. As a result, there exist equilibria which depend on unreasonable out-of-equilibrium beliefs. For example, there is a set of equilibria where both types lie about the return to the project, and their actions are supported by beliefs that any incumbent who tells the truth must be very unlikely to be a good type. I address the issue by focusing on equilibria that satisfy a kind of monotonicity in beliefs.[15] I will say that the voter's beliefs satisfy this monotonicity property if $\lambda(H) \geq \lambda(L)$. What this means is that a voter cannot attribute a lower probability of being good to an incumbent who reports H than to one who reports L. This follows Coate and Morris (1995), and so will use their terminology, denoting this an equilibrium with monotonic beliefs (EMB).

Knowing that the citizen updates his beliefs in this manner, the incumbent chooses a first-period strategy that maximizes her expected payoff. I will call any untruthful reporting 'corruption' (i.e. a politician is corrupt if she reports $A = L$ when the project return is high). It should be noted here that in the interests of clarity I will focus on fully separating EMBs only. There is a set of semi-separating equilibria (depending on parameter values) which do not add to overall understanding of the

[15] This will become a focus on equilibria satisfying lexicographic monotonicity of beliefs in the extension in 2.4.

model's implications, and so will not be considered here.

2.3.3 Results

The first result states that in equilibrium there is always some corruption.[16]

Lemma 1. There is no equilibrium in which both types of incumbent behave honestly.

Lemma 1 establishes that universal honesty is never an equilibrium. The bad incumbent always has an incentive to deviate from the pooled strategies given the voter's beliefs. It remains to be seen that it is possible for at least one type of incumbent to behave honestly in equilibrium. Let $\{A_g, A_b\}$ be a pair of strategies, where the first object is the good type's strategy and the second object the bad type's.

Proposition 1. $\{H, L\}$ is an EMB of the benchmark model for all λ_i that satisfy

$$G\left(\frac{\lambda_i (1-\beta)}{\lambda_i (1-\beta) + (1-\lambda_i)}\right) \geq 1 - \frac{v_b(b) - v_b(0)}{\delta \Delta v_b}.$$

Note that, unlike in traditional signaling games,[17] there is no discipline effect on first-period behavior of the bad type. Even though she is looking forward to re-election, which requires voters to attach a high probability to her being a good type, she is unrestrained in her actions in the first period. This is due to the uncertainty the voter has over the outcome of the public project. If the incumbent's initial reputation is high enough, a bad incumbent can 'hide' behind the possibility (from the voter's perspective) that she is a good incumbent who had the bad luck of being in office when the project return was low.

An important feature of this equilibrium is that it is an equilibrium with honesty, in which good types do not lie/steal from the public (though bad types do).

[16] Omitted proofs are in the appendix.
[17] A classic example is Maskin and Tirole (2004).

2.4 Campaign Spending

2.4.1 Model

2.4.1.1 The Policy

In this extension, there are no changes in the mechanics of the public project. The incumbent still chooses how to report the outcome of a high-return project. The only change is that now, if the incumbent reports $A = L$ when in fact $B = H$, freeing up the amount b, she has the opportunity to spend that money on either personal consumption or on campaign-related activities or a combination of the two.

2.4.1.2 The Citizenry

Now, the citizen will be one of two types. With probability σ, the voter will be strategic. That is, he will be exactly like the voter in 2.3.1.2, observing the incumbent's actions (including her announcement and her level of campaign spending) and updating his beliefs via Bayes' rule. A strategic voter cares only about the types of the politicians. With probability $1 - \sigma$, he will be impressionable. His vote is susceptible to campaign activity in that he will vote for whichever candidate spends the most money in the campaign. The voter's type is realized just before the election is held, so that before that time it is unknown to all players.[18]

2.4.1.3 The Politicians

The politicians remain largely unchanged. Good types receive utility from the benefits their decisions render the citizenry, bad types from expropriated funds. Now, the challenger is realized at the end of period one with both an initial reputation λ_c and an amount of money available for campaign spending c_c. c_c is assumed to be drawn randomly from a distribution $H(c)$, which is smooth and increasing, with $H(0) = 0$. This endowment of campaign funds does not contribute to either politician's utility, except indirectly via its effect on re-election probability.

I make the following assumption on the good incumbent's utility, which states

[18] This is not crucial—what matters is that the voter cannot credibly signal to the incumbent what type he is, so that the incumbent makes her decision with some uncertainty as to which type of voter she will face.

that the benefits of being in office in the second period outweigh the present costs of lying to the voter.

Assumption 1. $v_g(H) - v_g(L) < \delta \Delta v_g$

2.4.1.4 The Game

Timing remains the same as in 2.3.1.5. The incumbent's strategy now specifies an announcement A and an amount of campaign spending $c_i \in [0, b]$. Her record now consists of (A, c_i). A strategic voter forms the belief $\lambda(A, c_i)$, and votes for the incumbent with probability $G(\lambda(A, c_i))$. Impressionable voters vote for the incumbent if $c_i > c_c$, which occurs with probability $H(c_i)$. This makes the total probability of re-election $\sigma G(\lambda(A, c_i)) + (1 - \sigma) H(c_i)$.

2.4.2 Equilibrium

Equilibrium in this part is also solved for by backward induction. Second period actions are the same. A strategic voter forms beliefs based on the incumbent's initial reputation and first period record and votes accordingly. An impressionable voter observes campaign spending and votes accordingly. It remains only to determine the incumbent's first period strategy.

Again in the interests of eliminating unnatural voter beliefs (and subsequently the unreasonable equilibria supported by them), I will focus on a slightly different class of equilibria than in the benchmark. Here, I will say that the citizen's beliefs satisfy 'lexicographic monotonicity' if the following properties are satisfied: first, it must be that $\lambda(H, \cdot) \geq \lambda(L, \cdot)$, which is essentially the same idea as in the benchmark—namely that an incumbent who reports high must be thought as least as good as an incumbent who reports low. Secondly, given a low report, it must be that $\lambda(L, x) \geq \lambda(L, y)$ if $x > y > 0$. This says that, if money is being spent in the campaign, a voter cannot believe an incumbent less likely to be good if she spends more money in the campaign (and hence less on her own personal consumption). This eliminates equilibria that require the good incumbent to throw money away if the voter believes that spending all of b in the campaign is the mark of a bad type. Equilibria satisfying this property will be called equilibria with lexicographic monotonic beliefs (ELMB).

It is immediately apparent that a good incumbent will only take the actions $(H, 0)$

or (L, b) in equilibrium, meaning that she will either report a high return truthfully, or if she opts to lie she will spend all the funds on her campaign. A bad incumbent, on the other hand, chooses campaign spending c_i to maximize

$$v_b\left(b - c_i\right) + \delta\left(\sigma \cdot 0 + (1 - \sigma) H\left(c_i\right)\right) \Delta v_b$$

(taking into account that $\lambda\left(L, c_i\right) = 0$ for all c_i that separate her from a good type), which yields c^* satisfying

$$h\left(c^*\right) = \frac{v_b'\left(b - c^*\right)}{\delta\left(1 - \sigma\right) \Delta v_b}. \tag{2.5}$$

With the following assumptions on the bad type's preferences, I can characterize the equilibria of this game.

Assumption 2a. $v_b\left(b\right) - v_b\left(b - c^*\right) \leq \delta\left(1 - \sigma\right) H\left(c^*\right) \Delta v_b$

Assumption 2b. $v_b\left(b - c^*\right) - v_b\left(0\right) \geq \delta\left[\sigma + (1 - \sigma)\left(H\left(b\right) - H\left(c^*\right)\right)\right] \Delta v_b$

Taken together, Assumptions 2a and 2b mean that a bad incumbent finds it worthwhile to forgo some immediate consumption to increase her re-election chances, but not worthwhile to forgo all immediate consumption. These are restrictions on the parameters that yield the equilibrium results of interest. One further assumption ensures that all probabilities are between 0 and 1.

Assumption 3. $v_b\left(b\right) - v_b\left(b - c^*\right) \geq \delta\left[(1 - \sigma) H\left(c^*\right) - \sigma\right] \Delta v_b$

Intuitively Assumption 3 says that the difference in re-election probability between spending c^* to win the impressionable voters and convincing the strategic voters cannot be too large.

2.4.3 Results

Let $\left\{\left(A_g, c_g\right), \left(A_b, c_b\right)\right\}$ be a pair of strategies, where the first object is the good type's strategy and the second object the bad type's.

Proposition 2. There are four fully-separating ELMBs of this game. They fall into two classes:

1. *Equilibria with honesty:* $\left\{(H, 0), (L, 0)\right\}$ and $\left\{(H, 0), (L, c^*)\right\}$, both supported by out-of-equilibrium beliefs $\lambda\left(L, c\right) = 0$ for all $c > 0$.

2. *Equilibria without honesty:* $\{(L,b),(L,0)\}$ and $\{(L,b),(L,c^*)\}$, both supported by out-of-equilibrium beliefs $\lambda(L,c)=0$ for $c\in(0,b)$.

In equilibria with honesty, a good incumbent will always report a high return truthfully, and a bad type will always lie, either taking all the funds for personal use or spending some on campaign activities. In equilibria without honesty, both types take money from the project. The good type uses it all in her campaign; the bad type either takes all of it for herself, or takes some and uses some in the campaign.

Result 1. In both classes of equilibrium (that is, regardless of what a good type does), the bad incumbent's strategy depends only on her initial reputation. Specifically, the bad incumbent's strategy is

1. $(L,0)$ for all $\lambda_i \geq \lambda_{(L,0)}$ where $\lambda_{(L,0)} \in (0,1)$ is defined by

$$G\left(\frac{\lambda_{(L,0)}(1-\beta)}{\lambda_{(L,0)}(1-\beta)+\left(1-\lambda_{(L,0)}\right)}\right)=\frac{1-\sigma}{\sigma}H(c^*)-\frac{v_b(b)-v_b(b-c^*)}{\sigma\delta\Delta v_b} \tag{2.6}$$

2. (L,c^*) for all $\lambda_i \leq \lambda_{(L,c^*)}$ where $\lambda_{(L,c^*)} \in (0,1)$ is defined by

$$G\left(\lambda_{(L,c^*)}\right)=\frac{1-\sigma}{\sigma}H(c^*)-\frac{v_b(b)-v_b(b-c^*)}{\sigma\delta\Delta v_b}. \tag{2.7}$$

The first case occurs when the incumbent's initial reputation is high enough that a bad type will rely on the strategic voter having a high enough posterior belief following a low report to compensate for not spending any money in the campaign, similar to the benchmark model. The second case occurs when her initial reputation is low enough that it becomes worth it for the bad incumbent to forgo unrestrained consumption in the first period to increase her chance of re-election by winning over the impressionable voter.[19]

Result 2. Which class of equilibrium arises depends on the relationship between the good type's preferences and the voter composition. In particular, it depends on the direction of inequality of the following expression:

$$\frac{v_g(H)-v_g(L)}{\sigma\delta\Delta v_g} \gtreqless \frac{1-\sigma}{\sigma}H(b). \tag{2.8}$$

[19] Note that $\lambda_{(L,c^*)} < \lambda_{(L,0)}$, which means there are intermediate values of λ_i for which no separating ELMB exists.

Informally, if the left hand side is greater than the right hand side, then it is not worthwhile for a good incumbent to steal from the public project in the first period to increase her chances of re-election.[20] If the right hand side is greater, meaning she is sufficiently more likely to encounter an impressionable voter than a strategic one, the opposite is true, and even a good incumbent will steal money from the public project in order to ensure she is re-elected.

2.5 Discussion

The results laid out above seem to suggest a possible explanation for the re-election of visibly corrupt public officials. If a strategic voter cares about getting a good politician into office, then he will vote for a politician who steals from him if he observes that she acted in the interests of being re-elected to do good in the future. Knowing this, a good politician may be willing to trade off her present desire to pass on high returns to the public against the competing desire to be in office in the future. While the strategic voter might wish that the good politician did not steal from him in the first period, it would be imprudent of him to eject a politician he knows to be good in favor of a challenger who is of unknown type.

Looking again at equation (2.8), notice that the left hand side is the normalized benefit of truth-telling for the good incumbent. This is the relationship between the utility lost by stealing from the public project in the first period and the discounted utility of being in office in the second period. Let this ratio be called her 'preferences'. The right hand side is then the increased probability of re-election achieved by spending b. For simplicity throughout the rest of the discussion, let us take $H(b) = 1$. That is, an incumbent who spends b in the campaign is sure to win the impressionable voter.

As shown in Figure 2.1 on page 74, for any given voter composition (any σ), there are some good politicians who value re-election highly enough to make it worthwhile for them to act corruptly in the first period. In this case, the strategic voter sees the incumbent as the proverbial lesser of two evils—not ideal, but better than the alternative (a challenger who may be a bad type). This is loosely in line with Caselli and Morelli (2004) in that voters might prefer a politician with preferences that induce her to act honestly, but none may be available. If that is the case, they will vote for the dishonest one who at least seeks to serve their interests in the long

[20] Note that if Assumption 1 is violated, then this is always true, and a good type will never be dishonest in the first period.

term.

Alternatively, the same figure illustrates that, for any given set of preferences,[21] there is a voter composition that will induce a good incumbent to act corruptly in the first period. This is the case of politics as a dirty game, where there is a good chance that the election is decided not on merit, but on other factors (money or influence), and any incumbent wishing to be re-elected must play by the rules. The voter may prefer that the incumbent doesn't steal from the project, and the incumbent herself may be reluctant to do so ($v_g(H) > v_g(L)$), but as long as she values re-election at least slightly, she can be induced to act dishonestly. The strategic voter will still support her.

Consider now the bad incumbent, who always lies. The first period now involves a more complicated trade-off between consumption and re-election. For her, the presence of the impressionable voter provides a perverse kind of discipline, in as much as she now sometimes forgoes maximal consumption in the first period. However, this discipline does not benefit society, since it goes to winning over the impressionable voter with campaign spending.

2.5.1 A Note on Welfare

I turn now to the question of policy. If it were possible to do away with the means of influencing impressionable voters, would society be better off? In order to make this comparison, I need an assumption on the behavior of impressionable voters when campaign spending is disallowed. Say that they flip a coin and vote with probability $\frac{1}{2}$ for the incumbent and with probability $\frac{1}{2}$ for the challenger.

Consider first the baseline model. The expression for the citizen's expected utility before the first period action (given that λ_i is such that we have $\{H, L\}$ as an equilibrium)[22] is somewhat cumbersome and becomes much easier to work with if I make some simplifying assumptions. Take voter utility $u(H) = 1$, $u(L) = 0$, and G as the standard uniform distribution ($\lambda_c \sim U[0,1]$). The general expression and the simplified version are provided in full in the appendix.

Making the same simplifications, the citizen's expected utility in the four equilib-

[21] Provided they satisfy Assumption 2.

[22] In the presence of impressionable voters who randomize when indifferent, the condition for $\{H, L\}$ as an equilibrium is $G\left(\frac{\lambda_i(1-\beta)}{\lambda_i(1-\beta)+(1-\lambda_i)}\right) \geq 1 - \frac{v_b(b)-v_b(0)}{\delta\sigma\Delta v_b}$, so that any time $\{H, L\}$ is an equilibrium of the game without impressionable voters, it is also an equilibrium of the game with impressionable voters.

ria of the extended model—$\{(H,0),(L,0)\}$, $\{(L,b),(L,0)\}$, $\{(H,0),(L,c^*)\}$, $\{(L,b),(L,c^*)\}$—is also given in the appendix.

It is straightforward to see that $E\left[U^{H,L}\right] \geq E\left[U^{(H,0),(L,0)}\right]$ if and only if $\lambda_i \geq \frac{1}{3-\beta}$. So if the incumbent has a good enough reputation, the voter is better off if campaign spending is banned. Otherwise, if the incumbent has a bad reputation, welfare is improved by letting the incumbent lose to the (expected average quality) incumbent on spending.[23]

What if it is impossible to ban the campaign activity? Since a good incumbent has a better chance of getting re-elected if she spends b in the campaign than she does if she simply passes that amount on to the voter, it might be feasible that voter welfare is higher when the good incumbent steals from the public project, provided she uses it to get re-elected. However, that turns out not to be the case. Comparing $\{(H,0),(L,0)\}$ with $\{(L,b),(L,0)\}$ and $\{(H,0),(L,c^*)\}$ with $\{(L,b),(L,c^*)\}$, the voter always prefers a good politician to act honestly with regards to the returns to the public project.[24]

Comparing $\{(H,0),(L,0)\}$ to $\{(H,0),(L,c^*)\}$, it is necessary to say something about $H(c^*)$. $E\left[U^{(H,0),(L,c^*)}\right]$ is minimal when $H(c^*) = 1$. So I will consider that scenario, and if it is the case that voters are better off when the bad incumbent reveals himself when he is guaranteed to win the impressionable vote, it is also the case if he might be beaten by the challenger. Some algebra yields $E\left[U^{(H,0),(L,c^*)}\right] \geq E\left[U^{(H,0),(L,0)}\right] \leftrightarrow \sigma \geq \frac{\lambda_i(1-\beta)+(1-\lambda_i)}{\lambda_i(1-\beta)+(1-\lambda_i)+\lambda_i^2(1-\beta)}$.

If the voter is sufficiently likely to be strategic (and hence able to expel an incumbent who is revealed to be bad), then voters would like for bad incumbents to reveal themselves. However, recall from (2.7) that higher values of σ mean it is less likely that a bad incumbent is willing to reveal himself, instead preferring to masquerade as an unlucky good incumbent.

To give an example, for $\beta = \frac{1}{2}$, if $\lambda = \frac{2}{5}$, $E\left[U^{(H,0),(L,0)}\right] \geq E\left[U^{H,L}\right]$ so voters prefer campaign activity not be banned. Then, if $\sigma \geq 0.9$, voters would prefer the bad type to reveal himself through spending a positive amount less than b on the campaign. However, if $\sigma = 0.9$ then (under the simplifying assumptions above) $\lambda_{(L,c^*)} < \frac{1}{9}$, so

[23] If in the absence of any campaign spending the impressionable voter votes for the incumbent (a sort of incumbency advantage), then welfare is higher under no campaign spending if $\lambda \geq \frac{1}{2} = E[\lambda_c]$.

[24] The condition for voter welfare to be higher when the good politician acts corruptly is $\delta > \frac{2}{\beta(1-\sigma)}$, so voters have to value future payoffs extremely highly in order to be willing to trade present payoff for future honesty.

the bad incumbent would not find any campaign spending worthwhile—she would rather lose the impressionable voter in order to consume more in the present and hide behind the possibility of project failure in the reputation comparison with the challenger. Note that these results are highly dependent on the simplifying assumptions made above, and it may be the case that campaign spending is more attractive under alternative parameter and distributional assumptions.

To summarize, a sufficiently high reputation for the incumbent means that campaign spending is undesirable. In almost all cases, voter welfare would be higher if it were possible to ban the campaign activity and let the strategic voter determine the outcome of the election based on his assessment of the candidates' types. This is in line with much of the campaign finance literature. It is only if the incumbent has a bad reputation that voter welfare is improved by letting the challenger defeat the bad incumbent on spending. Voters never want good politicians to ransack public funds to finance campaigns, but they may want bad incumbents to reveal themselves through their spending patterns, in which case leaving open the option of campaign spending may be optimal.

2.5.2 Further Inquiry

There are a number of avenues of further inquiry indicated by the results of this paper. If voters strongly prefer a candidate of their political or ethnic affiliation, or if they feel very strongly about a particular policy issue, might they be more eager to admit some corruption in order to have their preferred candidate in office? The policy space is very limited here, with management of the public project the only responsibility of the incumbent. As such, this paper does not address the possibility that voters trade-off between their policy preferences (or perhaps personal characteristics such as ethnicity) and corruption. Adding parties, multi-dimensional policy/corruption space, or ethnicity could therefore add some realism to the model.

In addition to adding the aforementioned characteristics to flesh out the political space, there are a number of unsatisfactory elements to this model, which further extension may address. Firstly, the challenger is little more than a straw man in this set-up. He takes no action, does nothing specific to earn his reputation, and cannot account for his campaign funds. This last criticism, while very much intended in the case of this model to introduce an in-built advantage over the incumbent, is the most troubling. Why one candidate should have access to funds or electoral support and not another should be made explicit in a model that tries to explain election results.

2.6 Conclusion

I have proposed an explanation for the re-election of corrupt politicians, namely that if good politicians can count on rational voters' recognizing them as such and supporting them in elections, they may in some circumstances be induced to behave corruptly in pursuit of re-election. This result is counter-intuitive, especially considering the theoretical and empirical emphasis on the disciplining effect of re-election. I do not attempt to account for cases in which re-election motives provide discipline, but rather for the many puzzling cases in which corrupt incumbents are re-elected, sometimes by large margins.

The results of this paper suggest two mechanisms whereby this may happen. Firstly, if voters perceive that all politicians are potentially corrupt, they have no choice but to elect the corrupt politician whom they believe will best serve their interests in the long run. Secondly, if some corruption is necessary for re-election, and if politicians like staying in office, then they can be induced to behave dishonestly, correctly perceiving that rational voters will support them anyway.

These results also suggest some instances in which it may be better for voters to leave the mechanism for corrupt campaign activity in place. If the voter composition is such that an incumbent is very likely to face a strategic voter on election day, then it may be preferable to risk a bad incumbent's re-election if it allows strategic voters to eject the bad types more frequently. Furthermore, if voters place a higher weight on future consumption than present consumption, they may prefer that a good incumbent uses whatever means are at her disposal to get re-elected, even if it means stealing from them.

Figures

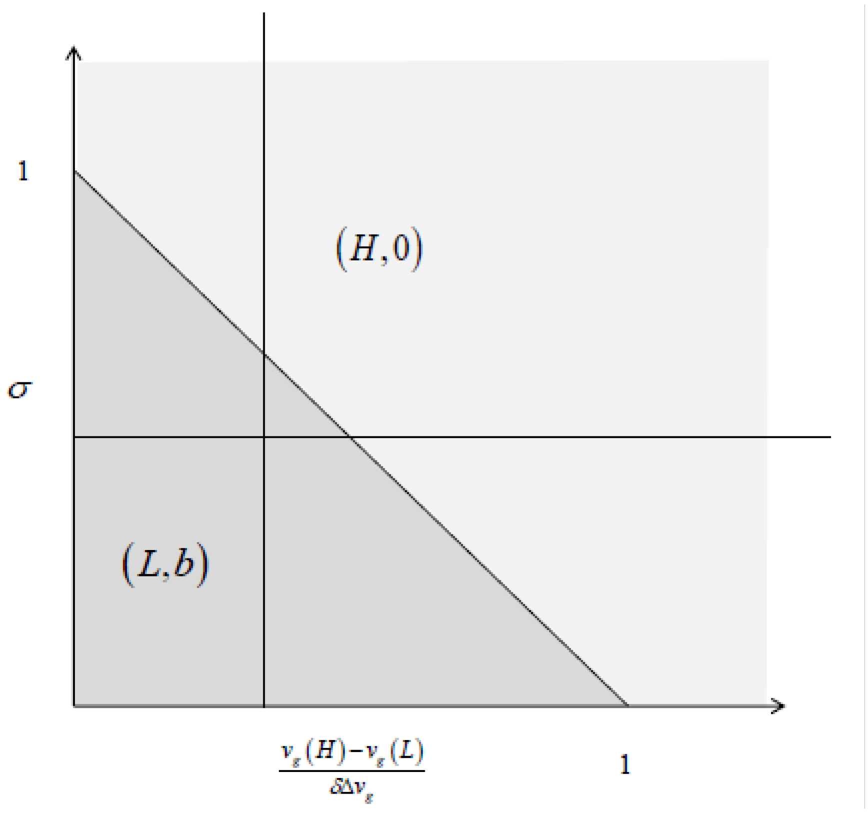

Figure 2.1: Good incumbent's trade-off

The figure shows the good incumbent's trade-off between 'preferences' $\frac{v_g(H)-v_g(L)}{\delta\Delta v_g}$ and re-election. For a given set of preferences, there is a small enough σ to make a good incumbent play (L, b). For a given voter composition (σ), there are incumbent preferences that lead to the choice of (L, b) in the first period.

Appendix

Proofs

Proof of Lemma 1. If both types of incumbent behave honestly when the project return is high, that is, $A = H$ for both types, then the citizen cannot distinguish them from their records. His posterior must equal his prior

$$\lambda(H) = \lambda(L) = \lambda_i.$$

A bad incumbent expects payoff $v_b(0) + \delta G(\lambda_i)\Delta v_b$ if she reports truthfully upon observing a high return. If she were to lie, that is report $A = L$, she would expect to receive $v_b(b) + \delta G(\lambda_i)\Delta v_b$, which is greater. The bad type, therefore, has a profitable deviation from her postulated strategy.□

Proof of Proposition 1. Given the specified equilibrium strategies, the citizen's beliefs calculated from Bayes' rule are

$$\begin{aligned} \lambda(H) &= 1 \\ \lambda(L) &= \frac{\lambda_i(1-\beta)}{\lambda_i(1-\beta)+(1-\lambda_i)}. \end{aligned}$$

A good incumbent expects payoff $v_g(H) + \delta\Delta v_g$ given these beliefs. If she were to deviate and lie, she would expect $v_g(L) + \delta G\left(\frac{\lambda_i(1-\beta)}{\lambda_i(1-\beta)+(1-\lambda_i)}\right)\Delta v_g$, which is lower. She will not deviate. A bad incumbent expects payoff $v_b(b)+\delta G\left(\frac{\lambda_i(1-\beta)}{\lambda_i(1-\beta)+(1-\lambda_i)}\right)\Delta v_b$ from her equilibrium strategy. If she were to deviate and truthfully report the high return, she would get $v_b(0) + \delta\Delta v_b$, which is less than her equilibrium payoff for all λ_i that satisfy the condition specified.[25] ■

Proof of Proposition 2.

It is obvious that a good incumbent will only play the pure strategies $(H, 0)$ or (L, b) in ELMB. Therefore these are the only strategies that need to be considered for the good type. A bad type will never pool with a good type in equilibrium, the proof of which is the same as that of Proposition 1. Since we are looking at fully separating ELMBs we therefore only need to consider the potential equilibria $\{(H,0),(L,0)\}$; $\{(H,0),(L,c^*)\}$; $\{(L,b),(L,0)\}$; and $\{(L,b),(L,c^*)\}$, since a bad type will either take all the money for personal use, or if she spends some, she will optimize her spending according to equation (2.5).

[25] A sufficient condition for the existence of such an equilibrium is $v_b(b) - v_b(0) \leq \delta\Delta v_b$.

The payoffs set down in equations (2.1) to (2.4) now become

$$\begin{aligned}
V_g(H,0) &= v_g(H)+\delta\sigma G(\lambda(H,0))\Delta v_g \\
V_g(L,b) &= v_g(L)+\delta[\sigma G(\lambda(L,b))+(1-\sigma)H(b)]\Delta v_g \\
V_b(L,c) &= v_b(b-c)+\delta[\sigma G(\lambda(L,c))+(1-\sigma)H(c)]\Delta v_b \\
V_b(L,0) &= v_b(b)+\delta\sigma G(\lambda(L,0))\Delta v_b
\end{aligned}$$

1. $\{(H,0),(L,0)\}$
   This strategy pair results in the beliefs $\lambda(H,0)=1$, $\lambda(L,0)=\frac{\lambda_i(1-\beta)}{\lambda_i(1-\beta)+(1-\lambda_i)}$; $\lambda(L,c)\in[0,1]$ for $c\neq 0$.
   First consider a good type. Given these beliefs, the payoffs above, and Assumption 2b, a good type will not deviate to (L,b) for any beliefs $\lambda(L,b)\in[0,1]$ provided $v_g(H)-v_g(L)\geq\delta(1-\sigma)H(b)\Delta v_g$, which follows from Assumption 2.
   Now consider a bad type. This equilibrium is supported by out of equilibrium beliefs $\lambda(L,c)=0$ for $c\neq 0$. If a bad incumbent spends money in the campaign, she will spend c^* as defined in (2.5). The condition for a bad type's not wanting to deviate to $(H,0)$ or (L,c^*) is $G\left(\frac{\lambda_i(1-\beta)}{\lambda_i(1-\beta)+(1-\lambda_i)}\right)\geq\frac{1-\sigma}{\sigma}H(c^*)-\frac{v_b(b)-v_b(b-c^*)}{\sigma\delta\Delta v_b}$. The second condition holds for all $\lambda_i\geq\lambda_{(L,0)}$, where $\lambda_{(L,0)}$ is defined as in (2.6), and is between 0 and 1 by Assumptions 2a and 3. If these conditions are satisfied, then $\{(H,0),(L,0)\}$is an ELMB.

2. $\{(H,0),(L,c^*)\}$
   This strategy pair results in the beliefs $\lambda(H,0)=1$, $\lambda(L,c^*)=0$, $\lambda(L,0)=\lambda_i$, $\lambda(L,c)\in[0,1]$ for $c\notin\{0,c^*\}$.
   Given these beliefs, the payoffs above, and Assumption 2b, the condition for the good type's not wanting to deviate to (L,b) is the same as in part 1 above. Regarding a bad type, this equilibrium is supported by out of equilibrium beliefs $\lambda(L,c)=0$ for $c\notin\{0,c^*\}$. A bad type will not deviate to any other spending amount c by the definition of c^*as optimal. The condition for a bad type's never wanting to deviate to $(H,0)$ or $(L,0)$ is $G(\lambda_i)\leq\frac{1-\sigma}{\sigma}H(c^*)-\frac{v_b(b)-v_b(b-c^*)}{\sigma\delta\Delta v_b}$. The second condition holds for all $\lambda_i\leq\lambda_{(L,c^*)}$, where $\lambda_{(L,c^*)}$is defined as in (2.7) and is between 0 and 1 by Assumptions 2a and 3. If these conditions are satisfied, then $\{(H,0),(L,c^*)\}$ is an ELMB.

3. $\{(L,b),(L,0)\}$
   This strategy pair results in the beliefs $\lambda(L,b)=1$, $\lambda(L,0)=\frac{\lambda_i(1-\beta)}{\lambda_i(1-\beta)+(1-\lambda_i)}$, $\lambda(H,0)=1$, and $\lambda(L,c)\in[0,1]$ for $c\notin\{0,b\}$. Note that $\lambda(H,0)$ is pinned down by lexicographic monotonicity. Since $\lambda(L,c)$ is not pinned down for

$c \notin \{0, b\}$, let $\lambda(L, c) = 0$ for those parameter values.

Given these beliefs, the payoffs above, and Assumption 2b, the condition for a good type's not wanting to deviate to $(H, 0)$ is $v_g(H) - v_g(L) \leq \delta(1-\sigma) H(b) \Delta v_g$. If a bad type spends money in the campaign, she will spend c^* as defined in equation (2.5). The condition for a bad type's not wanting to deviate to $(H, 0)$, (L, b), or (L, c^*) is the same as in part 1. The second condition holds for all $\lambda_i \geq \lambda_{(L,0)}$. If these conditions are satisfied, then $\{(L, b), (L, 0)\}$ is an ELMB.

4. $\{(L, b), (L, c^*)\}$

   This strategy pair results in beliefs $\lambda(L, b) = 1$, $\lambda(L, c^*) = 0$, $\lambda(L, 0) = \lambda_i$, $\lambda(H, 0) = 1$, and $\lambda(L, c) \in [0, 1]$ for $c \notin \{0, c^*, b\}$. Again, note that $\lambda(H, 0)$ is pinned down by lexicographic monotonicity. Since $\lambda(l, c)$ is not pinned down for $c \notin \{0, c^*, b\}$, let $\lambda(L, c) = 0$ for those values.

   Given these beliefs, the payoffs above, and Assumption 2b, the condition for a good type's not wanting to deviate to $(H, 0)$ is the same as in part 3. A bad type will not deviate to any other spending amount $c \in [0, b)$ by the definition of a maximum. The condition for a bad type's not wanting to deviate to $(H, 0)$, (L, b), or $(L, 0)$ is the same as in part 2. The second condition holds for all $\lambda_i \leq \lambda_{(L,c^*)}$. If these conditions are satisfied, then $\{(L, b), (L, c^*)\}$ is an ELMB.■

Proof of Result 1.

1. Consider parts 1 and 3 from the proof of the claim above. If $\lambda_i \geq \lambda_{(L,0)}$, then regardless of the strategy of a good incumbent, a deviation from $(L, 0)$ is not worthwhile.

2. Consider parts 2 and 4 from the proof of the claim above. If $\lambda_i \leq \lambda_{(L,c^*)}$, then regardless of the strategy of a good incumbent, a deviation from (L, c^*) is not worthwhile.□

Proof of Result 2.

From the proof of Claim 1, it is clear that the good type's strategy, independent of what a bad type plays, only depends on the direction of inequality of equation (2.8).□

A note on welfare

Baseline model:

$$
\begin{aligned}
E\left[U^{H,L}\right] \;=\;& \lambda_i\beta\left[u\left(H\right)+\delta\sigma\left(\beta u\left(H\right)+\left(1-\beta\right)u\left(L\right)\right)\right] \\
&+\lambda_i\beta\frac{1}{2}\delta\left(1-\sigma\right)\left(\beta u\left(H\right)+\left(1-\beta\right)u\left(L\right)\right) \\
&+\lambda_i\beta\frac{1}{2}\delta\left(1-\sigma\right)\left(\beta\bar{\lambda}u\left(H\right)+\left(1-\beta\bar{\lambda}\right)u\left(L\right)\right) \\
&+\lambda_i\left(1-\beta\right)\left[u\left(L\right)+\delta\sigma\left(G\left(\lambda\left(L\right)\right)\left(\beta u\left(H\right)+\left(1-\beta\right)u\left(L\right)\right)\right)\right] \\
&+\lambda_i\left(1-\beta\right)\delta\sigma\left(1-G\left(\lambda\left(L\right)\right)\right)\left(\beta\tilde{\lambda}u\left(H\right)+\left(1-\beta\tilde{\lambda}\right)u\left(L\right)\right) \\
&+\lambda_i\left(1-\beta\right)\delta\left(1-\sigma\right)\frac{1}{2}\left(\beta u\left(H\right)+\left(1-\beta\right)u\left(L\right)\right) \\
&+\lambda_i\left(1-\beta\right)\delta\left(1-\sigma\right)\frac{1}{2}\left(\beta\bar{\lambda}u\left(H\right)+\left(1-\beta\bar{\lambda}\right)u\left(L\right)\right) \\
&+\left(1-\lambda_i\right)\left[u\left(L\right)+\delta\sigma G\left(\lambda\left(L\right)\right)\left(\beta u\left(H\right)+\left(1-\beta\right)u\left(L\right)\right)\right] \\
&+\left(1-\lambda_i\right)\delta\sigma\left(1-G\left(\lambda\left(L\right)\right)\right)\left(\beta\tilde{\lambda}u\left(H\right)+\left(1-\beta\tilde{\lambda}\right)u\left(L\right)\right) \\
&+\left(1-\lambda_i\right)\delta\left(1-\sigma\right)\left[\frac{1}{2}u\left(L\right)+\frac{1}{2}\left(\beta\bar{\lambda}u\left(H\right)+\left(1-\beta\bar{\lambda}\right)u\left(L\right)\right)\right]
\end{aligned}
$$

where $\bar{\lambda}=E\left[\lambda_c\right]$ and $\tilde{\lambda}\equiv E\left[\lambda_c \mid \lambda_c>\lambda\left(L\right)\right]$.

Under the simplifying assumptions $u(H)=1$, $u(L)=0$, and G as the standard uniform distribution ($\lambda_c\sim U\left[0,1\right]$), the citizen's expected utility becomes

$$
\begin{aligned}
E\left[U^{H,L}\right] \;=\;& \lambda_i\beta\left[1+\delta\sigma\beta+\delta\left(1-\sigma\right)\beta\frac{3}{4}\right] \\
&+\lambda_i\left(1-\beta\right)\left[\delta\sigma\beta\lambda\left(L\right)+\delta\sigma\beta\left(1-\lambda\left(L\right)\right)\frac{1+\lambda\left(L\right)}{2}+\delta\left(1-\sigma\right)\beta\frac{3}{4}\right] \\
&+\left(1-\lambda_i\right)\left[\delta\sigma\beta\left(1-\lambda\left(L\right)\right)\frac{1+\lambda\left(L\right)}{2}+\delta\left(1-\sigma\right)\beta\frac{1}{4}\right]
\end{aligned}
$$

where $\lambda\left(L\right)=\frac{\lambda_i(1-\beta)}{\lambda_i(1-\beta)+(1-\lambda_i)}$.

Extended model:

1. $\{(H,0),(L,0)\}$

$$\begin{aligned}
E\left[U^{(H,0),(L,0)}\right] &= \lambda_i\beta\left[1+\delta\sigma\beta+\delta(1-\sigma)\beta\frac{1}{2}\right] \\
&\quad +\lambda_i(1-\beta)\left[\delta\sigma\beta\lambda(L,0)+\delta\sigma\beta(1-\lambda(L,0))\frac{1+\lambda(L,0)}{2}\right] \\
&\quad +\lambda_i(1-\beta)\delta(1-\sigma)\beta\frac{1}{2} \\
&\quad +(1-\lambda_i)\left[\delta\sigma\beta(1-\lambda(L,0))\frac{1+\lambda(L,0)}{2}+\delta(1-\sigma)\beta\frac{1}{2}\right]
\end{aligned}$$

2. $\{(L,b),(L,0)\}$

$$\begin{aligned}
E\left[U^{(L,b),(L,0)}\right] &= \lambda_i\beta[\delta\beta] \\
&\quad +\lambda_i(1-\beta)\left[\delta\sigma\beta\lambda(L,0)+\delta\sigma\beta(1-\lambda(L,0))\frac{1+\lambda(L,0)}{2}\right] \\
&\quad +\lambda_i(1-\beta)\frac{1}{2}\delta(1-\sigma)\beta \\
&\quad +(1-\lambda_i)\left[\delta\sigma\beta(1-\lambda(L,0))\frac{1+\lambda(L,0)}{2}+\delta(1-\sigma)\beta\frac{1}{2}\right]
\end{aligned}$$

3. $\{(H,0),(L,c^*)\}$

$$\begin{aligned}
E\left[U^{(H,0),(L,c^*)}\right] &= \lambda_i\beta\left[1+\delta\sigma\beta+\delta(1-\sigma)\beta\frac{1}{2}\right] \\
&\quad +\lambda_i(1-\beta)\left[\delta\sigma\beta\lambda_i+\delta\sigma\beta(1-\lambda_i)\frac{1+\lambda_i}{2}+\delta(1-\sigma)\beta\frac{1}{2}\right] \\
&\quad +(1-\lambda_i)\beta\left[\delta\sigma\beta\frac{1}{2}+\delta(1-\sigma)\beta(1-H(c^*))\frac{1}{2}\right] \\
&\quad +(1-\lambda_i)(1-\beta)\left[\delta\sigma\beta(1-\lambda_i)\frac{1+\lambda_i}{2}+\delta(1-\sigma)\beta\frac{1}{2}\right]
\end{aligned}$$

4. $\{(L,b),(L,c^*)\}$

$$\begin{aligned}
E\left[U^{(L,b),(L,c^*)}\right] &= \lambda_i\beta[\delta\beta] \\
&\quad +\lambda_i(1-\beta)\left[\delta\sigma\beta\lambda_i+\delta\sigma\beta(1-\lambda_i)\frac{1+\lambda_i}{2}+\delta(1-\sigma)\beta\frac{1}{2}\right] \\
&\quad +(1-\lambda_i)\beta\left[\delta\sigma\beta\frac{1}{2}+\delta(1-\sigma)\beta(1-H(c^*))\frac{1}{2}\right] \\
&\quad +(1-\lambda_i)(1-\beta)\left[\delta\sigma\beta(1-\lambda_i)\frac{1+\lambda_i}{2}+\delta(1-\sigma)\beta\frac{1}{2}\right]
\end{aligned}$$

2.7 Chapter 2 References

1. Austen-Smith, David. 1987. Interest groups, campaign contributions, and probabilistic voting. *Public Choice* 54 (2): 123-139.

2. Baron, David. 1994. Electoral competition with informed and uninformed voters. *The American Political Science Review* 88 (March): 33-47.

3. Besley, Timothy. 2006. *Principled Agents? The Political Economy of Good Government.* New York, NY: Oxford University Press, Inc.

4. _______ and Stephen Coate. 1998. Sources of inefficiency in a representative democracy: A dynamic analysis. *American Economic Review* 88 (March): 139-156.

5. Brennan, Geoffrey and James Buchanan. 2000. *The Power to Tax: Analytical Foundation of a Fiscal Constitution.* Library of Economics and Liberty.

6. Caselli, Francesco and Massimo Morelli. 2004. Bad politicians. *The Journal of Public Economics* 88: 759-782.

7. Coate, Stephen. 2004. Pareto-improving campaign finance policy. *The American Economic Review* 94 (June): 628-655.

8. Coate, Stephen and Stephen Morris. 1995. On the form of transfers to special interests. *The Journal of Political Economy* 103 (December): 1210-1235.

9. Dries, Bill. 2007. Peete sentenced to 51 months in prison. *Memphis Daily News.* 15 November.

10. Egan, Tim. 2004. The evangelical vote. *State of the Union.* 9 November.

11. Ferraz, Claudio and Frederico Finan. 2008. Exposing corrupt politicians: The effect of Brazil's publicly released audits on electoral outcomes. *The Quarterly Journal of Economics* 123 (2): 703-745.

12. Grossman, Gene and Elhanan Helpman. 2001. *Special Interest Politics.* Cambridge, MA: MIT Press.

13. Maskin, Eric and Jean Tirole. 2004. The politician and the judge: Accountability in government. *The American Economic Review* 94 (September): 1034-

1054.

14. Myerson, Roger. 1993. Effectiveness of electoral systems for reducing government corruption: A game-theoretic analysis. *Games and Economic Behavior* 5 (1): 118-132.

15. Prat, Andrea. 2002. Campaign advertising and voter welfare. *Review of Economic Studies* 69 (4): 997-1017.

16. Reinikka, Ritva and Jakob Svensson. 2004. Local capture: Evidence from a central government transfer program in Uganda. *The Quarterly Journal of Economics* 119 (May): 679-705.

17. Rundquist, Bary, Gerald Strom, and John Peters. 1977. Corrupt politicians and their electoral support: Some experimental observations. *The American Political Science Review* 71 (September): 954-963.

18. Smothers, Ronald. 2002. Hudson county freeholder charged with extorting bribes. *The New York Times.* 12 December.

Chapter 3

Overconfidence in Group Decision-Making

3.1 Introduction

Models of information aggregation generally assume agreement among committee members about the distribution of information within the group. All strategic behavior takes place within the context of a known distribution (or known expected distribution) of relevant information. Overconfidence can undermine this agreement. If I think I am incredibly knowledgeable about the subject to be decided, and act accordingly, but my partner believes me to be wholly uninformed (and also acts accordingly), our voting behavior will depart from standard predictions. We may also fail to aggregate our information efficiently. This paper is concerned with strategic responses to overconfidence, and with institutional remedies to it.

I model a three-person committee which must decide on a policy. The members agree on which policy is best given the state of the world, but have differing ability to assess the state. In particular, one of the committee members is completely uninformed (receives a pure noise signal) but believes that his signal is as informative as that of his best-informed colleague. This causes him to vote informatively (following his signal) when he should ignore his signal and vote to make better-informed agents pivotal.

The other two members of the committee know about the first agent's biased belief, and must respond to it in their voting strategy. I show that under majority rule the effect of overconfidence in an uninformed committee member is to cause the middle-informed member to ignore her signal. In trying to make the most informed agent pivotal, she discards her information. Given these voting strategies, overconfidence results in trading better information for worse, and causes a loss of social welfare. I then show that alternative voting rules can mitigate the loss associated with overconfidence. When the uninformed member suffers from biased beliefs about his ability, at least one of the unanimity rules does strictly better than majority rule to aggregate information. Indeed, the voting rule which requires unanimity to enact the *a priori* most likely policy becomes highest-ranked, which represents a reversal of preferences from the non-biased baseline.

3.2 Related Literature

This paper takes for granted that agents in committees vote strategically, taking into account the information contained in other voters' strategies. It therefore follows the seminal work of Austen-Smith and Banks (1996). Following their contribution, there is a body of work on optimal voting rules given strategic behavior by individuals in group decision-making.[1] Even with identical preferences,[2] voting rules can be important for achieving socially optimal outcomes. Feddersen and Pesendorfer (1998) show that unanimity requirements on juries result in a higher probability of error than simple majority rule, since a single juror's private information may not be convincing in the event she is pivotal.

When there is asymmetric information on a committee, strategic voting must take this heterogeneity into account as well. Feddersen and Pesendorfer (1996) demonstrate that less well-informed voters may strictly prefer to abstain to avoid adding noise to the decision process.[3] I do not allow for abstention, but something of its flavor remains in the uninformed player's decision to vote in such a way that guarantees that better-informed agents decide the outcome.

In this paper I analyze strategic voting behavior in a committee when there is heterogeneity in information quality and the least-informed agent is overconfident about his ability. Overconfidence was first discussed in the psychological literature,

[1]See, for example, Feddersen and Pesendorfer (1997, 1998, 1999), McLennan (1998) and Coughlan (2000). There is a large literature on optimal voting rules with endogenous information acquisition which is not relevant to this paper. See, for example, Persico (2004), Gerardi and Yariv (2008), or Gershkov and Szentes (2009). Levy (2007) explores optimal voting rules in an adverse selection setting.

[2]Visser and Swank (2007) find the optimal voting rule when committee members have divergent preferences.

[3]Battaglini et al. 2010 provide experimental evidence in support of Feddersen and Pesendorfer's model.

but there is also a substantial economic literature on overconfidence, particularly in individual decision-making. Odean (2004) reviews much of the literature on overconfidence among traders, and tests empirically whether a certain type of investor trades excessively due to overconfidence. Malmendier and Tate (2005, 2008) find evidence of overconfidence in CEO investment and acquisition decisions. Camerer and Lovallo (1999) experimentally test the relationship between personal overconfidence and business entry mistakes. In this literature, overconfidence is usually modeled as a boundedly rational assessment of one's likelihood of success (either by overestimating one's ability or underestimating the challenges posed by one's environment).

More specifically, I consider a particular type of overconfidence similar to the Dunning-Kruger effect (after the psychologists who experimentally identified it). The Dunning-Kruger effect disproportionately affects low-ability individuals, and makes them overestimate their ability both in absolute terms and relative to the rest of the population (Kruger and Dunning 1999). Their interpretation is that the same skills that produce good performance are necessary to evaluate good performance, so there is a meta-cognitive gap that renders low-skilled individuals simultaneously incompetent and unaware of their incompetence. Further studies have found such an effect in various circumstances: college debate teams and hunters tested on their knowledge of firearms (Ehrlinger et al. 2003),[4] medical lab technicians tested on knowledge of relevant terminology and problem solving (Haun et al. 2000), and medical residents evaluating their interview skills (Hodges et al. 2001). When I assume that the uninformed committee member has an incorrect belief about his ability, I have in mind something like this effect.

It should be noted that although overconfidence is most often thought of in terms of bounded rationality, there are many researchers who argue that there is nothing irrational about it. Benoit and Dubra (2009), taking issue with prevailing explanations of the 'better than average effect' whereby a (possibly large) majority of people consider their skills better than average, argue that it is fully compatible with rationality for a majority of the population to regard themselves as 'above average'. This seemingly impossible state of affairs arises because of rational updating based on random realizations (e.g. drivers who happen not to have an accident will reasonably surmise they are likely to be good drivers, even though only a small minority of all drivers—skilled or not—actually have an accident in a given period of time). Krajč and Ortmann (2008) make a similar argument that the Dunning-Kruger effect is due

[4] The effect was evident in this paper even when subjects were monetarily incentivized to assess their performance accurately.

to inference based on biased information rather than a meta-cognitive impairment among low-ability individuals.[5] Both of these arguments contend that overconfidence is more properly described as an inference or signal-extraction problem than as a failure of rationality. Notably, they do not quibble with the existence of the phenomenon, merely its source. As I am interested in the effects of overconfidence on group decision-making rather than its psychological or micro-foundational roots, I remain agnostic on that point.

3.3 Model

In this section I introduce the basic elements of the model, along with an extension to account for overestimation of ability in uninformed agents.

3.3.1 Benchmark model

There are two states of the world, 0 and 1, where $W = \{0, 1\}$ denotes the set of states with generic element w. A 3-member committee $\{A, B, C\}$ must make a decision about which policy to implement, where the policy space is $D = \{0, 1\}$ with generic element d. All committee members have identical preferences over the decision:

$$u(d, w) = \begin{cases} 1 & d = w \\ 0 & d \neq w. \end{cases}$$

At the beginning of the game, nature chooses a state of the world $w \in W$. All agents know that the probability of state 1 being realized is q, where $q > \frac{1}{2}$. Each committee member I then receives a signal $s_I \in \{0, 1\}$, which contains private information about the true state of the world. The signals are conditionally independent, and members differ in the accuracy of their signals. $\Pr(s_I = w \mid w) = i$. Agent A receives an uninformative signal $\left(a = \frac{1}{2}\right)$. Agents B and C, however, receive informative signals $\frac{1}{2} < b < q < c$. Note that B is informed $\left(b > \frac{1}{2}\right)$, but her signal is less accurate than the prior belief on the state of the world, whereas C is more accurate. To start with, I assume that all of the above information is common knowledge.

After observing his signal, each agent lodges a vote $v_I \in \{0, 1\}$. Agents are not allowed to abstain, but they are not required to vote in a way that reflects their

[5] See also Krueger and Mueller (2002) and Burson et al. (2006).

private information. The outcome d is determined by majority rule (MR), and I focus on the case where agents vote as if they are pivotal.

3.3.2 Model with overconfidence

Next I consider the case in which agent A suffers from overconfidence. Whilst the true accuracy of his signal is $a = \frac{1}{2}$, he believes his signal is accurate with probability $\hat{a} = c$. That is, though his signal is in reality uninformative, he believes it is as informative as the signal of the best-informed member of the committee. To keep things relatively simple, I will allow for common knowledge of disagreement about the quality of A's information. A will consider himself well-informed, but will accurately perceive the quality of the other members' information. B and C will correctly perceive all signal qualities, and that A overestimates his signal accuracy. A knows that B and C think he is uninformed, and that they will be acting accordingly; he simply disagrees with them.

This disagreement deserves some justification, since it departs from standard rationality in a pretty significant way. Firstly, common knowledge of A's overconfidence is the most conservative assumption I could make here, since it involves the smallest departure from full recognition of reality. If A believed instead that B and C were also bullish about his ability, the effects I analyze would be reinforced. Secondly, note that disagreement is compatible with Bayesian rationality since the agents do not share a common prior about the accuracy of A's signal. I therefore do not fall foul of Aumann (1976), which shows that with common priors players cannot 'agree to disagree'.

By way of example, it could be that B and C are more experienced, and therefore able to accurately pinpoint their own abilities and that of others, whereas A, being less experienced, overestimates his knowledge. This can happen on medical teams with both recent medical school graduates and experienced doctors. The recent graduates, lacking experience, may not be able to judge which of competing diagnoses is more likely, whereas more senior team members can discount unlikely possibilities and reach a more reasonable conclusion.[6] The parallels with my modeling assumptions are as follows: senior doctors correctly estimate the likelihood that they come to the correct diagnosis, and they also know how likely new graduates are to be correct. New graduates, on the other hand, accurately perceive that senior doctors are often correct, but think that their classroom training, which is of

[6]Thanks to Dr. Jake Young for this example.

more recent vintage than that of senior doctors, gives them better real-world diagnostic ability than it does. They think that senior doctors incorrectly underweight textbook knowledge relative to experience.[7]

3.3.3 Strategies and beliefs

Each agent chooses an action $v_I \in \{0,1\}$. Define the set of actions $v = \{v_A, v_B, v_C\}$. All agents seek to maximize expected utility

$$U = \sum_{w\in\{0,1\}} \Pr(w)\Pr(d = w \mid w, v).$$

Given that agents' incentives are aligned, they vote for whichever state of the world is more likely given their beliefs stemming from the prior q and their private signals, and conditional on being pivotal. Agents assess this likelihood using Bayes' Rule.

It will be convenient to refer to a transformation of the various probabilities above when going through the results. Define $l_q \equiv \ln \frac{q}{1-q}$, and similarly, for $i \in \{a,b,c\}$ define $l_i = \ln \frac{i}{1-i}$. Using these transformations, I can represent beliefs additively. For example, if A always votes 1 regardless of his signal, and C votes sincerely (that is, $v_C = s_C$), and B is pivotal, then it must be that $s_C = 0$. Say $s_B = 0$. By Bayes' Rule,

$$\Pr(w = 0 \mid q, s_B = 0, s_C = 0) = \frac{(1-q)\,bc}{(1-q)\,bc + q(1-b)(1-c)}$$

so that the posterior likelihood ratio is

$$L \equiv \ln\left[\frac{\Pr(w=0 \mid q, s_B = 0, s_C = 0)}{\Pr(w=1 \mid q, s_B = 0, s_C = 0)}\right] = \ln\left[\frac{(1-q)\,bc}{q(1-b)(1-c)}\right] = -l_q + l_b + l_c > 0$$

meaning that $\Pr(w = 0 \mid q, s_B = 0, s_C = 0) > \frac{1}{2}$ and B should vote 0.

3.4 Majority Rule

I limit attention to equilibria in pure strategies. Details can be found in the appendix.

[7]Other papers that make similar assumptions are Gervais and Goldstein (2007), Blanes i Vidal and Möller (2007), and Morris (1996).

3.4.1 Benchmark model

In the basic model, there are two relevant parameter regions, where the parameter of interest is q (or rather, l_q). Throughout the analysis, I assume that the difference in ability between B and C is 'large enough':[8]

Assumption 1: $l_b < \frac{1}{2} l_c$

3.4.1.1 $l_q \in [l_b, l_c - l_b)$

Here, there are two equivalent equilibria.[9] The equilibria, given in the form (v_A, v_B, v_C) are:

1. $(1, 0, s_C)$ with $U = c$

2. $(0, 1, s_C)$ with $U = c$

Player A correctly perceives himself as being uninformed, and therefore ignores his signal. Player B's two sources of information, her signal and the prior, are both inaccurate enough that she would prefer to ignore her signal and vote in such a way that makes C, the best-informed agent, pivotal.

3.4.1.2 $l_q \in [l_c - l_b, l_c]$

Here, there is a single pure strategy equilibrium.

1. $(1, s_B, s_C)$ with $U = (1 - q)\, bc + q\, (1 - (1 - b)\, (1 - c))$

Again, player A knows that his signal is uninformative, so when voting as if pivotal he casts a vote for the more likely state according to the prior. Now, however, player B, though knowing her information is not as accurate as C's, can vote informatively. This is because the prior is strong enough that when B is pivotal (i.e. C has received $s_C = 0$), if B has received $s_B = 1$, that information plus the prior is sufficient to outweigh $C's$ signal.

[8] All of the main results go through with the weaker assumption that $\frac{c-b}{(c+b)(1-c)} > \frac{2b-1}{2-b}$. Making the simpler assumption allows me to focus on a particular case for completeness of exposition. It essentially requires some minimal heterogeneity in ability on the committee. For example, if $b = .6$, this assumption requires $c \geq .692$. The weaker assumption requires $c \geq .661$.

[9] There is also a third, mixed-strategy equilibrium which I will not discuss here.

3.4.2 Overconfidence model

Consider the range of values of q (l_q). The action is in whether B thinks she ought to play her signal once A believes his signal is informative ($\hat{a} = c$).

3.4.2.1 $l_q \in [l_b, l_c]$

Now, there is a single equilibrium.

1. $(s_A, 1, s_C)$ with $U = \frac{1}{2}(c + q)$

A's overconfidence leads him to vote according to his signal, which the other players recognize as uninformative. In essence, A is injecting noise into the voting process. B's best response is to ignore her signal and vote 1. She cannot distinguish between the two events $(s_A = 1, s_C = 0)$ and $(s_A = 0, s_C = 1)$ both of which leave her pivotal. She therefore compares her signal to the prior and, as her signal is less accurate than the prior, votes 1 regardless of her private information.

The consequences of this are noteworthy. When $l_q < l_c - l_b$, B was already ignoring her signal, so the effect of A's overconfidence is simply to add noise to the outcome. C is no longer always pivotal, which results in a loss because he is the best-informed. The expected loss is $\frac{1}{2}(c - q)$.

When $l_q \in [l_c - l_b, l_c]$, the effect is to trade B's information for A's. Whereas B would have voted informatively, contributing the information contained in her signal when A correctly perceived his ability, she ceases to vote informatively when he is overconfident. She strategically counteracts A's uninformed voting by lodging a vote for the *a priori* more likely state of the world. The expected loss is $\left(b - \frac{1}{2}\right)(c(1 - q) + q(1 - c))$.

3.5 Other voting rules

The loss generated by A's overconfidence can therefore be substantial. In this section, I explore whether the choice of voting rule can mitigate this loss. Up to now, I have assumed that decision is by majority rule, but other rules can also be applied. Specifically, there are two unanimity rules that might be implemented in this committee, and it is possible that the loss associated with these other rules might be less than that associated with majority rule.

Let $x \in \{1, 2, 3\}$ be the number of votes required to enact policy 1. Then $x = 1$ corresponds to a unanimity requirement for 0 (U0), $x = 2$ is again majority rule (MR), and $x = 3$ is a unanimity requirement for 1(U1). I covered majority rule in section 3.4. In this section I will cover equilibrium voting behavior under the other two voting rules.

3.5.1 0-Unanimity (U0, $x = 1$)

A similar pattern of results to those under majority rule arises here. Consider the same partition of the parameter space.

3.5.1.1 $l_q \in [l_b, l_c - l_b)$

1. The basic model involves equilibrium strategies $(0, 0, s_C)$ with $U = c$. When there is no bias, A and B ignore their private information and vote in such a way to ensure that the best-informed member, C, is pivotal.

2. The overconfidence model has strategies $(s_A, 0, s_C)$, with expected utility $U = \frac{1}{2}(c + q)$. The expected loss is again $\frac{1}{2}(c - q)$, the same as under majority rule.

3.5.1.2 $l_q \in [l_c - l_b, l_c]$

1. In the basic model, the equilibrium is $(0, s_B, s_C)$ with expected utility $U = (1 - q) bc + q (1 - (1 - b)(1 - c))$.

2. Under the model with overconfidence, the equilibrium is (s_A, s_B, s_C), with expected utility $U = (1 - q) \frac{1}{2} bc + q \left(1 - \frac{1}{2}(1 - b)(1 - c)\right)$ and expected loss $\frac{1}{2}[(1 - q) bc - q (1 - b)(1 - c)]$.

Here, unlike the case of $x = 2$ (majority rule), when A votes informatively, B votes informatively as well. This is due to the greater uncertainty when the voting rule is majoritarian as opposed to unanimous. Under majority rule, B cannot tell the difference between the pivotal events $(s_A = 1, s_C = 0)$ and $(s_A = 0, s_C = 1)$, so she is left to compare her own information with the (more accurate) prior, and disregards her signal. Under U0, however, B is pivotal only if C has received a 0-signal. In this case, when the prior information is accurate enough that agreement between s_B and the prior outweigh s_C, B can vote informatively. The loss due to A's bias here is smaller than under majority rule, so that U0 is preferable to MR if overconfidence is expected.

3.5.2 1-Unanimity (U1, $x = 3$)

For all $l_q \in [l_b, l_c]$, there is only one equilibrium regardless of the presence of overconfidence. The unique equilibrium under U1 is $(1, 1, s_C)$, with expected utility $U = c$. Here, A's bias is insufficient to induce him to follow his private signal. A and B always vote such that C is pivotal.

3.5.3 Optimal voting rules

In this section, I identify the optimal voting rules for different values of the prior q. Implicitly, this requires that the committee (or the social planner in charge of setting the committee's rules) can observe the prior before setting the voting rule.[10] In the next section I will consider the alternative case, in which a 'constitution' must be written which cements a voting rule before the prior is realized.

As seen in the last two sections, for every value of q and every voting rule, with and without bias, either there is a unique pure strategy equilibrium or the multiple pure strategy equilibria produce the same expected utility. It is therefore possible to rank the voting rules. These rankings are contained in following propositions.

Proposition 1: When A is unbiased (correctly perceives that $a = \frac{1}{2}$), the following ranking obtains:

1. $l_q \in [l_b, l_c - l_b)$ implies that U0$\sim$MR$\sim$U1

2. $l_q \in [l_c - l_b, l_c]$ implies that U0$\sim$MR$\succ$U1

Proposition 2: When A is overconfident ($\hat{a} = c$), there exists $q^* \in (l_c - l_b, l_c)$ such that the following ranking obtains:

[10] As B and C would agree on preferences over voting rule, pairwise majority rule would select the optimal voting rule. Obviously, other voting rules would produce different results, but such meta-voting procedures are beyond the scope of this paper.

1. $l_q \in [l_b, l_c - l_b)$ implies that U1$\succ$U0$\sim$MR

2. $l_q \in [l_c - l_b, l_{q^*})$ implies that U1$\succ$U0$\succ$MR

3. $l_q \in [l_{q^*}, l_c]$ implies that U0$\succ$U1$\succ$MR

The presence of overconfidence in the uninformed member of the committee results in some noteworthy reversals of preferences regarding voting rules. Specifically, majority rule is always weakly top-ranked when there is no bias, but once there is bias it becomes weakly bottom-ranked.[11] Also, when $l_q \in [l_c - l_b, l_{q^*}]$ 1-unanimity goes from strictly bottom-ranked to strictly top-ranked. For low and intermediate levels of the prior, it is optimal to choose a rule that is not vulnerable to A's bias, as A votes uninformatively under U1 to make C pivotal. This can mean losing the information contained in B's private signal, which is why U1 is weakly dominated by the other voting rules in the absence of bias. However, when policy 1 becomes more likely, it is better to choose the voting rule most biased in favor of 1, even though it means the voting outcome will be affected by uninformed voting by A. Then, both informed agents vote informatively.

To summarize, when the voting rule is decided with full information about the strength of the prior, the optimal voting rule depends on q. For low and intermediate levels of q, U1 is optimal. This voting rule is biased against 1, which is *a priori* more likely, but it prevents A from voting according to his uninformative signal, making C's information decisive. For high enough q, however, it becomes worthwhile to tolerate A's voting informatively in order to use both B's and C's information, so U0 (which is biased in favor of 1) becomes the optimal voting rule.

3.5.4 Optimal constitution

Occasionally voting rules must be set before any information is available, including any public information in the form of the prior. The problem then is to write a constitution, which specifies a universal voting rule that does not depend on any degree of confidence about the right decision. If the voting rule must be set before the prior is realized, the presence of overconfidence results in a reversal of preferences.

Assume that $q \sim U[b, c]$. When this is the only information available to the

[11] This is due to the uncertainty over which state of the world B is in when she is pivotal. Communication between her and C prior to voting might rehabilitate MR as the optimal voting rule. Since B and C have aligned incentives, they can share information, and they would prefer a voting rule that makes the pair of them decisive.

constitution writer, the optimal *ex ante* voting rule is summarized in the following propositions:

Proposition 3: When A is unbiased, the optimal constitution specifies either U0 or MR. The ranking is U0$\sim$MR$\succ$U1.

Because 1 is most likely $\left(q > \frac{1}{2}\right)$, in the absence of bias the constitution writer would never choose the voting rule which is biased against the most likely outcome. Given the strategies of the players, expected payoffs are equal between U0 and MR.

Proposition 4: When A is overconfident, the optimal constitution specifies U1. The ranking is U1$\succ$U0$\succ$MR.

The presence of bias makes 1-unanimity more attractive than the other voting rules, because under U1 A doesn't use his private information. The information of B is not used either, but this is preferable to the case in which A votes his (uninformed) signal whether or not B votes hers.

3.6 Conclusion

What I have shown is that, on a sufficiently heterogeneous committee, overconfidence in an uninformed member can cause a better-informed member to strategically discard her information, resulting in a social loss. This loss is most pronounced under majority rule, so that when overconfidence is an issue, unanimity rule is preferable. Indeed, when the voting rule must be set before any information is revealed, 1-unanimity becomes the most preferred constitutional voting rule. It is least-preferred when there is no overconfidence.

The dramatic reversals of preferences are due to the fact that agents cannot share their information with each other prior to voting. If B and C could discuss their signals, they could jointly determine the outcome under majority rule, which would resurrect majority rule as preferred. Therefore, the results presented here are most applicable to committees whose members know each others' identities but do not know what votes have been cast or what information other agents are working with at the time they make their voting decisions.

Appendix

Equilibrium

For each player I, each state of the world w, and each voting rule x, it is possible to calculate

$$\Pr\left(w = 1 \mid piv_x, s_I, v\right)$$

where $piv_x = \sum_{s_{-I} \in \{0,1\}^2} \Pr\left(s_{-I}\right) \Pr\left(piv \mid s_{-I}, x, v\right)$ and v is the conjectured equilibrium strategy. Then, each player votes for policy 1 if and only if

$$\Pr\left(w = 1 \mid piv_x, s_I, v\right) > \frac{1}{2}.$$

Alternatively, let $L_I\left(x, s_I, v\right)$ be the as-if-pivotal likelihood ratio for voting rule x and conjectured strategies v given signal s_I.

$$L_I\left(x, s_I, v\right) = \ln \frac{\Pr\left(w = 1 \mid piv_x, s_I, v\right)}{\Pr\left(w = 0 \mid piv_x, s_I, v\right)}$$

Player I votes for policy 1 if and only if

$$L_I\left(x, s_I, v\right) > 0.$$

Majority rule

First consider the benchmark case, where A knows that $a = \frac{1}{2}$.

1. $l_q \in [l_b, l_c - l_b)$:

   a) First I show that there is no equilibrium in which all three follow their signals (s_A, s_B, s_C). Suppose $v_A = s_A$ and $v_C = s_C$. In particular, I will show that B does not want to follow a signal $s_B = 0$.

$$\begin{aligned} L_B\left(2, 0, \left(s_A, \cdot, s_C\right)\right) &= \ln \frac{\Pr\left(w = 1 \mid piv_2, 0, v\right)}{\Pr\left(w = 0 \mid piv_2, 0, v\right)} \\ &= \ln \frac{q\left(\frac{1}{2}c + \frac{1}{2}\left(1 - c\right)\right)\left(1 - b\right)}{\left(1 - q\right)\left(\frac{1}{2}c + \frac{1}{2}\left(1 - c\right)\right)b} \\ &= l_q - l_b \\ &> 0 \end{aligned}$$

So upon receiving $s_B = 0$, B votes 1.

b) Next I show that there is no equilibrium in which two players follow their signals. I show the results for putative equilibria in which the non-responsive player plays 1; results are similar for those in which they play 0.

i. $(s_A, s_B, 1)$ C has a profitable deviation.

$$\begin{aligned} L_C\left(2, 0, (s_A, s_B, \cdot)\right) &= \ln \frac{\Pr\left(w = 1 \mid piv_2, 0, v\right)}{\Pr\left(w = 0 \mid piv_2, 0, v\right)} \\ &= \ln \frac{q\left(\frac{1}{2}b + \frac{1}{2}(1-b)\right)(1-c)}{(1-q)\left(\frac{1}{2}b + \frac{1}{2}(1-b)\right)c} \\ &= l_q - l_c \\ &< 0 \end{aligned}$$

So upon receiving $s_C = 0$, C votes 0.

ii. $(s_A, 1, s_C)$ A has a profitable deviation.

$$\begin{aligned} L_A\left(2, 1, (\cdot, 1, s_C)\right) &= \ln \frac{\Pr\left(w = 1 \mid piv_2, 1, v\right)}{\Pr\left(w = 0 \mid piv_2, 1, v\right)} \\ &= \ln \frac{q(1-c)\frac{1}{2}}{(1-q)c\frac{1}{2}} \\ &= l_q - l_c \\ &< 0 \end{aligned}$$

So upon receiving $s_A = 1$, A votes 0.

iii. $(1, s_B, s_C)$ B has a profitable deviation.

$$\begin{aligned} L_B\left(2, 1, (1, \cdot, s_C)\right) &= \ln \frac{\Pr\left(w = 1 \mid piv_2, 1, v\right)}{\Pr\left(w = 0 \mid piv_2, 1, v\right)} \\ &= \ln \frac{q(1-c)b}{(1-q)c(1-b)} \\ &= l_q - l_c + l_b \\ &< 0 \end{aligned}$$

So upon receiving $s_B = 1$, B votes 0.

c) Finally I show that in the only equilibrium, only player C follows his signal. The results are similar if the non-responsive players' strategies are reversed.

i. $(s_A, 1, 0)$ C has a profitable deviation.

$$\begin{aligned} L_C(2, 1, (s_A, 1, \cdot)) &= \ln \frac{q\frac{1}{2}c}{(1-q)\frac{1}{2}(1-c)} \\ &= l_q + l_c \\ &> 0 \end{aligned}$$

So upon receiving $s_C = 1$, C votes 1.

ii. $(1, s_B, 0)$ C has a profitable deviation.

$$\begin{aligned} L_C(2, 1, (1, s_B, \cdot)) &= \ln \frac{q(1-b)c}{(1-q)b(1-c)} \\ &= l_q - l_b + l_c \\ &> 0 \end{aligned}$$

So upon receiving $s_C = 1$, C votes 1.

iii. $(1, 0, s_C)$ No player has a deviation.

$$\begin{aligned} L_A(2, 0, (\cdot, 0, s_C)) &= \ln \frac{qc\frac{1}{2}}{(1-q)(1-c)\frac{1}{2}} \\ &= l_q + l_c \\ &> 0 \end{aligned}$$

So A has no deviation.

$$\begin{aligned} L_B(2, 1, (1, \cdot, s_C)) &= \ln \frac{q(1-c)b}{(1-q)c(1-b)} \\ &= l_q - l_c + l_b \\ &< 0 \end{aligned}$$

So B has no deviation.

$$\begin{aligned} L_C(2, 1, (1, 0, \cdot)) &= \ln \frac{qc}{(1-q)(1-c)} \\ &= l_q + l_c \\ &> 0 \end{aligned}$$

$$\begin{aligned} L_C(2,0,(1,0,\cdot)) &= \ln \frac{q(1-c)}{(1-q)c} \\ &= l_q - l_c \\ &< 0 \end{aligned}$$

So C has no deviation. There is also an equivalent equilibrium $(0,1,s_C)$.

2. $l_q \in [l_c - l_b, l_c]$
I will show that the equilibrium given in the text is indeed an equilibrium. Showing its uniqueness proceeds as above, but is omitted for brevity.

$$\begin{aligned} L_A(2,0,(\cdot,s_B,s_C)) &= \ln \frac{q(b(1-c)+(1-b)c)\frac{1}{2}}{(1-q)(b(1-c)+(1-b)c)\frac{1}{2}} \\ &= l_q \\ &> 0 \end{aligned}$$

So A has no deviation.

$$\begin{aligned} L_B(2,1,(1,\cdot,s_C)) &= \ln \frac{q(1-c)b}{(1-q)c(1-b)} \\ &= l_q - l_c + l_b \\ &> 0 \end{aligned}$$

$$\begin{aligned} L_B(2,0,(1,\cdot,s_C)) &= \ln \frac{q(1-c)(1-b)}{(1-q)cb} \\ &= l_q - l_c - l_b \\ &< 0 \end{aligned}$$

So B has no deviation.

$$\begin{aligned} L_C(2,1,(1,s_B,\cdot)) &= \ln \frac{q(1-b)c}{(1-q)b(1-c)} \\ &= l_q - l_b + l_c \\ &> 0 \end{aligned}$$

$$\begin{aligned} L_C(2,0,(1,s_B,\cdot)) &= \ln \frac{q(1-b)(1-c)}{(1-q)bc} \\ &= l_q - l_b - l_c \\ &< 0 \end{aligned}$$

So C has no deviation.

Now suppose that A is overconfident, so that $\hat{a} = c$. Again, I will show that the equilibrium provided in the text is indeed an equilibrium. Uniqueness is demonstrated as in part 1 above.

1. $l_q \in [l_b, l_c]$

$$\begin{aligned} L_A(2,1,(\cdot,1,s_C)) &= \ln \frac{q(1-c)\hat{a}}{(1-q)c(1-\hat{a})} \\ &= l_q - l_c + l_{\hat{a}} \\ &> 0 \end{aligned}$$

$$\begin{aligned} L_A(2,0,(\cdot,1,s_C)) &= \ln \frac{q(1-c)(1-\hat{a})}{(1-q)c\hat{a}} \\ &= l_q - l_c - l_{\hat{a}} \\ &< 0 \end{aligned}$$

So A has no deviation.

$$\begin{aligned} L_B(2,0,(s_A,\cdot,s_C)) &= \ln \frac{q\left(\frac{1}{2}c + \frac{1}{2}(1-c)\right)(1-b)}{(1-q)\left(\frac{1}{2}c + \frac{1}{2}(1-c)\right)b} \\ &= l_q - l_b \\ &> 0 \end{aligned}$$

So B has no deviation.

$$\begin{aligned} L_A(2,1,(s_A,1,\cdot)) &= \ln \frac{q\frac{1}{2}c}{(1-q)\frac{1}{2}(1-c)} \\ &= l_q + l_c \\ &> 0 \end{aligned}$$

$$\begin{aligned} L_A(2,0,(s_A,1,\cdot)) &= \ln \frac{q\frac{1}{2}(1-c)}{(1-q)\frac{1}{2}c} \\ &= l_q - l_c \\ &< 0 \end{aligned}$$

So C has no deviation.

0-**Unanimity and** 1-**Unanimity**

The same type of analysis yields the results in the text.

In summary, the strategies and expected payoffs are given in the following tables.

Table 3.1: Benchmark model strategies and payoffs

$l_{q\in}\,[l_b, l_c - l_b)$

	U0	**MR**	**U1**
strategy	$(0, 0, s_C)$	$(1, 0, s_C)/(0, 1, s_C)$	$(1, 1, s_C)$
exp. utility	c	c	c

$l_{q\in}\,[l_c - l_b, l_c]$

	U0	**MR**	**U1**
strategy	$(0, s_B, s_C)$	$(1, s_B, s_C)$	$(1, 1, s_C)$
exp. utility	$(1-q)\,bc + q\,(1-(1-b)\,(1-c))$	see U0	c

Table 3.2: Overconfidence model strategies and payoffs

$l_{q\in}\,[l_b, l_c - l_b)$

	U0	**MR**	**U1**
strategy	$(s_A, 0, s_C)$	$(s_A, 1, s_C)$	$(1, 1, s_C)$
exp. utility	$\frac{1}{2}\,(c+q)$	$\frac{1}{2}\,(c+q)$	c

$l_{q\in}\,[l_c - l_b, l_c]$

	U0	**MR**	**U1**
strategy	(s_A, s_B, s_C)	$(s_A, 1, s_C)$	$(1, 1, s_C)$
exp. utility	$(1-q)\,\frac{1}{2}bc + q\left(1 - \frac{1}{2}\,(1-b)\,(1-c)\right)$	$\frac{1}{2}\,(c+q)$	c

Proofs

Proof of Proposition 1: Follows from comparison of the payoffs in Table 3.1.■

Proof of Proposition 2: Follows from comparison of the payoffs in Table 3.2. $(1-q)\,\frac{1}{2}bc+q\left(1-\frac{1}{2}\,(1-b)\,(1-c)\right) > \frac{1}{2}\,(c+q)$ for all $q > \frac{c(1-b)}{c(1-b)+b(1-c)}$, so U0$\succ$MR. Clearly, $c > \frac{1}{2}\,(c+q)$, so U1$\succ$MR.

It remains to show when U1$\succ$U0 for $l_q \in [l_c - l_b, l_c]$.

$$\begin{aligned} (1-q)\,\frac{1}{2}bc + q\left(1 - \frac{1}{2}\,(1-b)\,(1-c)\right) &\leq c \\ q + qb\,(1-c) &\leq c + (1-q)\,(1-b)\,c \end{aligned}$$

The left hand side increases in q and the right hand side decreases in q. When $q = \frac{c(1-b)}{c(1-b)+b(1-c)}$, the inequality simplifies to

$$1 - b \leq b$$

which is true. When $q = c$, the inequality simplifies to

$$b \leq 1 - b$$

which is false. Therefore by continuity there exists a $q^* \in \left[\frac{c(1-b)}{c(1-b)+b(1-c)}, c\right]$ such that U1$\sim$U0. For all $q < q^*$, U1$\succ$U0; for all $q > q^*$, U0$\succ$U1. ■

Proof of Proposition 3: Consider the payoffs in Table 3.1. Clearly, expected utility is equal for all possible voting rules at low levels of q. For $\{q \mid l_q \geq l_c - l_b\}$, $(1-q)\,bc + q\,(1-(1-b)(1-c)) > c$. This implies that U0 and MR are preferred to U1 for high levels of q. Therefore, a constitution writer is indifferent between U0 and MR, and prefers either of these to U1.■

Proof of Proposition 4: Consider the payoffs in Table 3.2. Clearly, $c > \frac{1}{2}(c+q)$, so U1$\succ$MR for all q. Also, $(1-q)\,\frac{1}{2}bc + q\left(1 - \frac{1}{2}(1-b)(1-c)\right) > \frac{1}{2}(c+q)$ for all $q > \frac{c(1-b)}{c(1-b)+b(1-c)}$, so U0$\succ$MR. It remains to show that U1$\succ$U0.

Let $t = \frac{c(1-b)}{c(1-b)+b(1-c)}$ (i.e. $q \mid l_q = l_c - l_b$). The *ex ante* expected utility from U1 is

$$\int_b^c c\frac{1}{c-b}\,\mathrm{d}q$$

The *ex ante* expected utility from U0 is

$$\int_b^t \frac{1}{2}(c+q)\,\frac{1}{c-b}\,\mathrm{d}q + \int_t^c \left[\frac{1}{2}(1-q)\,bc + q\left(1-\frac{1}{2}(1-b)(1-c)\right)\right]\frac{1}{c-b}\,\mathrm{d}q.$$

U1$\succ$U0 requires $U(U1) > U(U0)$, which gives (after integrating and some manipulation)

$$\begin{aligned}
(c-t)\,2c\,(2-b) + 2c\,(t-b) &> (c-t)\,(c+t)\,(1+b+c-2bc) \\
&\quad + (t+b)\,(t-b) \\
(c-t)^2\,(1+b+c-2bc) + 2c\,(t-b) &> (c-t)\,2c\,(1-c)\,(2b-1) \\
&\quad + (t+b)\,(t-b) \\
(c-t)^2\,(1+b+c-2bc) + 2\,(c-t)\,(t-b) + (t-b)^2 &> (c-t)\,2c\,(1-c)\,(2b-1)\,.
\end{aligned}$$

Because $t = \frac{c(1-b)}{c(1-b)+b(1-c)}$,

$$\begin{aligned}(c-t)^2(1+b+c-2bc)+2(c-t)(t-b)+(t-b)^2 &> (c-t)^2\,2\,(c(1-b)+b(1-c)) \\ 2(c-t)(t-b)+(t-b)^2+b(c-t)^2 &> (c-t)^2(1-c)(2b-1)\,.\end{aligned}$$

Since $b<1$, $b>2b-1$, which gives the result.

So U1$\succ$U0 if the constitution must be written before q is realized.$\blacksquare$

3.7 Chapter 3 References

1. Aumann, Robert. 1976. Agreeing to disagree. *The Annals of Statistics* 4 (6): 1236-1239.

2. Austen-Smith, David and Jeffrey Banks. 1996. Information aggregation, rationality, and the Condorcet jury theorem. *The American Political Science Review* 90 (March): 34-45.

3. Battaglini, Marco, Rebecca Morton and Thomas Palfrey. 2010. The swing voter's curse in the laboratory. *Review of Economic Studies* 77 (January): 61-89.

4. Benoit, Jean-Pierre and Juan Dubra. 2009. Overconfidence? Working paper (January). Available at SSRN: http://ssrn.com/abstract=1088746.

5. Blanes i Vidal, Jordi and Marc Möller. 2007. When Should Leaders Share Information with Their Subordinates? *Journal of Economics & Management Strategy* 16 (Summer): 251-283.

6. Burson, Katherine, Richard Larrick and Joshua Klayman. 2006. Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology* 90 (1): 60-77.

7. Camerer, Colin and Dan Lovallo. 1999. Overconfidence and excess entry: An experimental approach. *The American Economic Review* 89 (March): 306-318.

8. Coughlan, Peter. 2000. In defense of unanimous jury verdicts: Mistrials, communication, and strategic voting. *The American Political Science Review* 94 (June): 375-393.

9. Ehrlinger, Joyce, Kerri Johnson, Matthew Banner, David Dunning and Justin Kruger. 2008. Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes* 105: 98-121.

10. Feddersen, Timothy and Wolfgang Pesendorfer. 1996. The swing voter's curse. *The American Economic Review* 86 (June): 408-424.

11. _______. 1997. Voting behavior and information aggregation in elections with private information. *Econometrica* 65 (September): 1029-1058.

12. _______. 1998. Convicting the innocent: The inferiority of unanimous jury verdicts. *The American Political Science Review* 92 (March): 23-35.

13. _______. 1999. Elections, information aggregation, and strategic voting. *Proceedings of the National Academy of Sciences* 96(September): 10571-10574.

14. Gerardi, Dino and Leeat Yariv. 2008. Information acquisition in committees. *Games and Economic Behavior* 62: 436-459.

15. Gershkov, Alex and Balázs Szentes. 2009. Optimal voting schemes wtih costly information acquisition. *Journal of Economic Theory* 144: 36-68.

16. Gervais, Simon and Itay Goldstein. 2007. The positive effects of biased self-perception in firms. *Review of Finance* 11(August): 453-496.

17. Haun, Daniel, Andrea Zeringue, Argie Leach and Angela Foley. 2000. Assessing the competence of specimen-processing personnel. *Laboratory Medicine* 31: 633-637.

18. Hodges, Brian, Glenn Regehr and Dawn Martin. 2001. Difficulties in recognizing one's own incompetence: Novice physicians who are unskilled and unaware of it. *Academic Medicine* 76: S87-S89.

19. Krajč, Marian and Andreas Ortmann. 2008. Are the unskilled really that unaware? An alternative explanation. *Journal of Economic Psychology* 29 (5): 724-738.

20. Krueger, Joachim and Ross Mueller. 2002. Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology* 82: 180-188.

21. Kruger, Justin and David Dunning. 1999. Unskilled and unaware of it: How difficulties in recognizing one's own imcompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology* 77 (6): 1121-1134.

22. Levy, Gilat. 2007. Decision making in committees: Transparency, reputation, and voting rules. *The American Economic Review* 97 (March): 150-168.

23. Malmendier, Ulrike and Geoffrey Tate. 2005. CEO overconfidence and corporate investment. *The Journal of Finance* 60 (December): 2661-2700.

24. _______. 2008. Who makes acquisitions? CEO overconfidence and the market's reaction. *Journal of Financial Economics* 89 (July): 20-43.

25. McLennan, Andrew. 1998. Consequences of the Condorcet jury theorem for beneficial information aggregation by rational agents. *The American Political Science Review* 92 (June): 413-418.

26. Morris, Stephen. 1996. Speculative investor behavior and learning. *The Quarterly Journal of Economics* 111 (November): 1111-33.

27. Odean, Terrance. 2004. Do investors trade too much? in *Advances in Behavioral Economics*, edited by Colin Camerer, George Loewenstein and Matthew Rabin. Princeton, NJ: Princeton University Press.

28. Persico, Nicola. 2003. Committee design with endogenous information. *Review of Economic Studies* 70:. 1-27.

29. Visser, Bauke and Otto Swank. 2007. On committees of experts. *The Quarterly Journal of Economics* 122 (1): 337-372.

References

[1] Aumann, Robert. 1976. Agreeing to disagree. *The Annals of Statistics* 4 (6): 1236-1239.

[2] Austen-Smith, David. 1987. Interest groups, campaign contributions, and probabilistic voting. *Public Choice* 54 (2): 123-139.Banks, Jeffrey and Joel Sobel. 1987. Equilibrium Selection in Signaling Games. *Econometrica* 55 (May): 647-661.

[3] Austen-Smith, David and Jeffrey Banks. 1996. Information aggregation, rationality, and the Condorcet jury theorem. *The American Political Science Review* 90 (March): 34-45.

[4] Baron, David. 1994. Electoral competition with informed and uninformed voters. *The American Political Science Review* 88 (March): 33-47.

[5] Battaglini, Marco, Rebecca Morton and Thomas Palfrey. 2010. The swing voter's curse in the laboratory. *Review of Economic Studies* 77 (January): 61-89.

[6] Benoit, Jean-Pierre and Juan Dubra. 2009. Overconfidence? Working paper (January). Available at SSRN: http://ssrn.com/abstract=1088746.

[7] Besley, Timothy. 2006. *Principled Agents? The Political Economy of Good Government.* New York, NY: Oxford University Press, Inc.

[8] _______ and Stephen Coate. 1998. Sources of inefficiency in a representative democracy: A dynamic analysis. *American Economic Review* 88 (March): 139-156.

[9] Blanes i Vidal, Jordi and Marc Möller. 2007. When Should Leaders Share Information with Their Subordinates? *Journal of Economics & Management Strategy* 16 (Summer): 251-283.

[10] Bolton, Patrick, Markus Brunnermeier and Laura Veldkamp. 2011. Leadership, Coordination and Corporate Culture. Working Paper (August).

[11] Boot, Arnoud. 1992. Why Hang on to Losers? Divestitures and Takeovers. *Journal of Finance* 47(September): 1401-1423.

[12] Brennan, Geoffrey and James Buchanan. 2000. *The Power to Tax: Analytical Foundation of a Fiscal Constitution.* Library of Economics and Liberty.

[13] Burson, Katherine, Richard Larrick and Joshua Klayman. 2006. Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology* 90 (1): 60-77.

[14] Camerer, Colin and Dan Lovallo. 1999. Overconfidence and excess entry: An experimental approach. *The American Economic Review* 89 (March): 306-318.

[15] Caselli, Francesco and Massimo Morelli. 2004. Bad politicians. *The Journal of Public Economics* 88: 759-782.

[16] Cho, In-Koo and David Kreps. 1987. Signaling Games and Stable Equilibria. *The Quarterly Journal of Economics* 102 (May): 179-221.

[17] Coate, Stephen. 2004. Pareto-improving campaign finance policy. *The American Economic Review* 94 (June): 628-655.

[18] Coate, Stephen and Stephen Morris. 1995. On the form of transfers to special interests. *The Journal of Political Economy* 103 (December): 1210-1235.

[19] Coughlan, Peter. 2000. In defense of unanimous jury verdicts: Mistrials, communication, and strategic voting. *The American Political Science Review* 94 (June): 375-393.

[20] Dries, Bill. 2007. Peete sentenced to 51 months in prison. *Memphis Daily News.* 15 November.

[21] Egan, Tim. 2004. The evangelical vote. *State of the Union.* 9 November.

[22] Ehrlinger, Joyce, Kerri Johnson, Matthew Banner, David Dunning and Justin Kruger. 2008. Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes* 105: 98-121.

[23] Englmaier, Florian. 2004. A Strategic Rationale for Having Overconfident Managers. Working Paper (December).

[24] Fama, Eugene. 1980. Agency Problems and the Theory of the Firm. *Journal of Political Economy* 88 (April): 288-307.

[25] Feddersen, Timothy and Wolfgang Pesendorfer. 1996. The swing voter's curse. *The American Economic Review* 86 (June): 408-424.

[26] _______. 1997. Voting behavior and information aggregation in elections with private information. *Econometrica* 65 (September): 1029-1058.

[27] _______. 1998. Convicting the innocent: The inferiority of unanimous jury verdicts. *The American Political Science Review* 92 (March): 23-35.

[28] _______. 1999. Elections, information aggregation, and strategic voting. *Proceedings of the National Academy of Sciences* 96(September): 10571-10574.

[29] Ferraz, Claudio and Frederico Finan. 2008. Exposing corrupt politicians: The effect of Brazil's publicly released audits on electoral outcomes. *The Quarterly Journal of Economics* 123 (2): 703-745.

[30] Ferreira, Daniel and Marcelo Rezende. 2007. Corporate Strategy and Information Disclosure. *RAND Journal of Economics* 38 (Spring): 164-184.

[31] Gerardi, Dino and Leeat Yariv. 2008. Information acquisition in committees. *Games and Economic Behavior* 62: 436-459.

[32] Gershkov, Alex and Balázs Szentes. 2009. Optimal voting schemes wtih costly information acquisition. *Journal of Economic Theory* 144: 36-68.

[33] Gervais, Simon and Itay Goldstein. 2007. The positive effects of biased self-perception in firms. *Review of Finance* 11(August): 453-496.

[34] Goel, Anand and Anjan Thakor. 2008. Overconfidence, CEO Selection and Corporate Governance. *Journal of Finance* 63(December): 2737-2784.

[35] Grossman, Gene and Elhanan Helpman. 2001. *Special Interest Politics*. Cambridge, MA: MIT Press.

[36] Hart, Oliver and Bengt Holmström. 2002. A Theory of Firm Scope. MIT Department of Economics Working Paper No. 02-42.

[37] Haun, Daniel, Andrea Zeringue, Argie Leach and Angela Foley. 2000. Assessing the competence of specimen-processing personnel. *Laboratory Medicine* 31: 633-637.

[38] Hodges, Brian, Glenn Regehr and Dawn Martin. 2001. Difficulties in recognizing one's own incompetence: Novice physicians who are unskilled and unaware of it. *Academic Medicine* 76: S87-S89.

[39] Holmström, Bengt. 1999. Managerial Incentive Problems: A Dynamic Perspective. *The Review of Economic Studies* 66 (January): 169-182. Special Issue: Contracts.

[40] Kanodia, Chandra, Robert Bushman and John Dickhaut. 1989. Escalation Errors and the Sunk Cost Effect: An Explanation Based on Reputation and Information Asymmetries. *Journal of Accounting Research* 27 (Spring): 59-77.

[41] Komai, Mana, Mark Stegeman and Benjamin Hermalin. 2007. Leadership and Information. *American Economic Review* 97 (June): 944-947.

[42] Krajč, Marian and Andreas Ortmann. 2008. Are the unskilled really that unaware? An alternative explanation. *Journal of Economic Psychology* 29 (5): 724-738.

[43] Krueger, Joachim and Ross Mueller. 2002. Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology* 82: 180-188.

[44] Kruger, Justin and David Dunning. 1999. Unskilled and unaware of it: How difficulties in recognizing one's own imcompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology* 77 (6): 1121-1134.

[45] Lazear, Edward and Sherwin Rosen. 1981. Rank-order Tournaments as Optimum Labor Contracts. *Journal of Political Economy* 89 (October): 841-864.

[46] Leaver, Clare. 2009. Bureaucratic minimal squawk behavior: Theory and evidence from regulatory agencies. *American Economic Review* 99 (June): 572-607.

[47] Levy, Gilat. 2004. Anti-herding and Strategic Consultation. *European Economic Review* 48 (June): 503-525.

[48] Levy, Gilat. 2007. Decision making in committees: Transparency, reputation, and voting rules. *The American Economic Review* 97 (March): 150-168.

[49] Malmendier, Ulrike and Geoffrey Tate. 2005. CEO overconfidence and corporate investment. *The Journal of Finance* 60 (December): 2661-2700.

[50] _______. 2008. Who makes acquisitions? CEO overconfidence and the market's reaction. *Journal of Financial Economics* 89 (July): 20-43.

[51] Maskin, Eric and Jean Tirole. 2004. The politician and the judge: Accountability in government. *The American Economic Review* 94 (September): 1034-1054.

[52] McLennan, Andrew. 1998. Consequences of the Condorcet jury theorem for beneficial information aggregation by rational agents. *The American Political Science Review* 92 (June): 413-418.

[53] Morris, Stephen. 1996. Speculative investor behavior and learning. *The Quarterly Journal of Economics* 111 (November): 1111-33.

[54] Myerson, Roger. 1993. Effectiveness of electoral systems for reducing government corruption: A game-theoretic analysis. *Games and Economic Behavior* 5 (1): 118-132.

[55] Odean, Terrance. 2004. Do investors trade too much? in *Advances in Behavioral Economics*, edited by Colin Camerer, George Loewenstein and Matthew Rabin. Princeton, NJ: Princeton University Press.

[56] Ottaviani, Marco and Peter Sørensen. 2000. Herd Behavior: A Comment. *American Economic Review* 90 (June): 695-704.

[57] Persico, Nicola. 2003. Committee design with endogenous information. *Review of Economic Studies* 70:. 1-27.

[58] Prat, Andrea. 2002. Campaign advertising and voter welfare. *Review of Economic Studies* 69 (4): 997-1017.

[59] Prendergast, Canice and Lars Stole. 1996. Impetuous Youngsters and Jaded Old-timers: Acquiring a Reputation for Learning. *Journal of Political Economy* 104 (December): 1105-1134.

[60] Reinikka, Ritva and Jakob Svensson. 2004. Local capture: Evidence from a central government transfer program in Uganda. *The Quarterly Journal of Economics* 119 (May): 679-705.

[61] Rotemberg, Julio and Garth Saloner. 2000. Visionaries, Managers and Strategic Direction. *RAND Journal of Economics* 31 (Winter): 693-716.

[62] Rundquist, Bary, Gerald Strom, and John Peters. 1977. Corrupt politicians and their electoral support: Some experimental observations. *The American Political Science Review* 71 (September): 954-963.

[63] Scharfstein, David and Jeremy Stein. 1990. Herd Behavior and Investment. *American Economic Review* 80 (June): 465-479.

[64] Smothers, Ronald. 2002. Hudson county freeholder charged with extorting bribes. *The New York Times*. 12 December.

[65] Van den Steen, Eric. 2005. Organizational Beliefs and Managerial Vision. *Journal of Law, Economics, and Organization* 21 (1): 256-283.

[66] Visser, Bauke and Otto Swank. 2007. On committees of experts. *The Quarterly Journal of Economics* 122 (1): 337-372.

[67] Zwiebel, Jeffrey. 1995. Corporate Conservatism and Relative Compensation. *Journal of Political Economy* 103(February): 1-25.