UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Incentives and Information in Multiagent Settings

Permalink
https://escholarship.org/uc/item/58q294ct

Author
Nayeem, Omar Ahmed

Publication Date
2013

Peer reviewed|Thesis/dissertation

Incentives and Information in Multiagent Settings

by

Omar Ahmed Nayeem

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Ahn, Chair
Professor Ernesto Dal Bó
Professor Benjamin Hermalin
Professor John Morgan

Spring 2013

Incentives and Information in Multiagent Settings

Abstract

Incentives and Information in Multiagent Settings

by

Omar Ahmed Nayeem

Doctor of Philosophy in Economics

University of California, Berkeley

Professor David Ahn, Chair

This dissertation comprises three papers, each of which analyzes a mechanism design issue that arises in a setting with multiple agents that need to either acquire or aggregate information for use in a decision. The decision affects all agents as well as a principal, who also plays the role of mechanism designer. The theoretical models that I develop in these papers can be applied to a wide range of diverse settings, but I emphasize applications in the areas of organizational economics and political economics.

The first paper, titled "The Value of 'Useless' Bosses," presents a novel view of the role of middle managers in organizations. Conventional wisdom regarding middle management suggests that a principal that can administer her organization independently has no reason to hire a manager, and that a principal that can benefit from a manager's services should hire one with aligned interests. The paper highlights a channel through which virtually any principal can benefit from the services of a manager, particularly of one whose interests differ. Specifically, when a principal relies on a worker to acquire information for an organizational decision, she can strengthen the worker's incentives by delegating the decision to a "biased" manager. Although casual observation of the game suggests that the manager's position is redundant, delegation benefits the principal. Thus, the paper helps to reconcile the prevalence of middle management with its widespread lamentation. It also illustrates how discord between a manager and a worker can improve an organization's performance. The results are consistent with outcomes from various knowledge-based organizations.

The second paper, titled "Communication and Preference (Mis)alignment in Organizations," conveys insights that are similar to the ones from "The Value of 'Useless' Bosses." Like the previous paper, this one explains the benefits of biased agents (both workers and managers) in organizations. However, unlike the previous paper, this one assumes that an organization's principal—whose time, technical expertise, and attention are limited—relies upon division managers to produce reports, which summarize information acquired by workers, to inform her decisions. Given this assumption, a pressing question for the principal is not whether to appoint a manager, but rather which type of manager to appoint. Note that two types of agency problems can arise in the setting described above. First, workers that

bear private costs for their information acquisition efforts may not exert as much effort as the principal would like. Second, managers that do not share the principal's preferences over decisions can produce false reports. The paper shows that, although preference alignment within the organization may be expected to minimize the principal's losses from agency, the principal may benefit from intraorganizational conflict. In particular, the principal can use a manager's bias to strengthen a worker's incentives to acquire information. Since a manager's incentive to mislead the principal vanishes if the acquired information is of sufficiently high quality, the principal realizes an unambiguous welfare gain by hiring a biased manager. The principal can further enhance her welfare by also hiring a biased worker, whose bias clashes with the manager's.

The third paper, titled "Efficient Electorates," analyzes a social choice setting with pure common values, private noisy information about an unobservable payoff-relevant state of the world, and costless voting. In such a setting, an economic argument in favor of direct democracy is essentially one about information aggregation: if all citizens vote according to their private information—which, on average, is correct—then, in large majority-rule elections, the probability that the welfare-maximizing outcome is implemented is close to one. This argument, formalized first by the Marquis de Condorcet in his celebrated "jury theorem" and later extended to cover more general environments, is an asymptotic result that requires voters' information to be sufficiently uncorrelated. The paper shows that, for a fixed number of sincere voters with shared information sources, direct democracy is often suboptimal. It then considers the problem of appointing an optimal electorate given the allocation of information. In special cases of this framework, the problem can be viewed as the choice of an electorate from a set of individuals that communicate with each other via a social network before the election. It provides a characterization of the optimal electorate for certain classes of networks. Because the optimal electorate is often a proper subset of the full set of agents, representative democracy—even in the absence of voting costs—is often more efficient than direct democracy. As the paper illustrates through various examples, though, the solution to the problem of optimal elector appointment is unstable, and so a general characterization of the optimal electorate is elusive.

With love,
to Akbar and Zakira Nayeem,
who enabled me to begin this journey,
and
to Ayesha Athar,
who helped me to complete it

Contents

Acknowledgments

The quality of a doctoral student's relationship with his or her principal advisor is a crucial—if not the most crucial—determinant of the student's success in his or her program. I am fortunate and extremely grateful to have had the excellent guidance and mentorship of David Ahn during my time as a graduate student. David was a continual source of encouragement and support. He consistently gave honest but constructive feedback and would always make himself available to discuss my progress or obstacles. I especially appreciate the fact that we had a strong enough relationship that I could freely talk to him about my personal life in addition to my research.[1] I also thank him for his immense generosity in providing funding opportunities and writing reference letters whenever I needed them.

My relationships with several other faculty members, besides David, were crucial to my success in completing this degree and finding employment. Santiago Oliveros was an unofficial second advisor for a large chunk of my graduate career. He was never one to sugarcoat his thoughts, and, with his honest and sharp critiques of my drafts, he pushed me to think carefully through the motivations behind my ideas and to let economic issues and insights—not mathematical pedantry and technical sophistication—dictate the results of my papers. The pitch of the first chapter of this dissertation (which was also my job market paper) is due largely to discussions with him. Santiago, along with David, was generous in providing funding whenever I needed it.

John Morgan was another tough, but extremely helpful, critic. In addition to suggesting numerous detailed improvements to my drafts, he helped extensively with the art of pitching and packaging my work and differentiating it from previous literature. His intuition about which framing techniques were likely to produce papers and presentations that would sell well was a remarkable boon. The idea of the second chapter of this dissertation emerged largely out of an extended discussion with John.

Before advising me on research, Ben Hermalin taught me mechanism design and agency theory. From his field course, I gained a solid foundation in the area. Ben provided extremely detailed comments on all facets of my writing and presentation and also pushed me to think especially hard and emphasize the novelty of my contribution, especially in interviews and job talks. I thank him, John, and David for writing reference letters for me and reaching out to their contacts when I went on the job market.

Ernesto Dal Bó served on both my oral exam committee and my final dissertation committee. As a self-described critic and skeptic of applied theory, Ernesto provided feedback that was especially useful for ensuring that my work would have appeal beyond audiences composed mostly of theorists. He suggested several applications for the theoretical results of all three chapters. In particular, he and the other three members of my oral exam committee—Fred Finan, Shachar Kariv, and Chris Shannon—provided valuable feedback on the third chapter of this dissertation.

---

[1]Indeed, as anyone that has been through such a program knows all too well, the two are inextricably linked.

Joe Farrell and Sean Gailmard both provided valuable institutional knowledge that helped me relate the results of my first chapter to real world environments, particularly federal government agencies. Both Joe Farrell and Nate Miller, an alumnus of the PhD program, were very generous in helping me navigate the recruiting process with these agencies while I was on the job market. Bob Anderson, the first professor I had at Berkeley (for the so-called "math camp" that preceded my first fall semester) was very generous in providing funding during the summer following my first year. He pushed me to prepare myself to become a theorist by taking graduate-level courses in the mathematics department, and, in exchange for summer funding, he motivated me to read some important papers in the field of political economics before taking my field courses. He would later recommend talking to David for getting started on a research project, and David eventually became my advisor.

I thank Saurabh Bhargava, Botond Kőszegi, Maciej Kotowski, Mauricio Larrain, SangMok Lee, Jian Li, Matthew Rabin, Gérard Roland, and Felix Várdy for detailed comments on earlier drafts of my dissertation chapters. Others with whom I had fruitful discussions include Joan de Martí, Shaddin Dughmi, Haluk Ergin, Ben Golub, Ben Handel, Ted Miguel, Ulrike Malmendier, Elchanan Mossel, Christos Papadimitriou, Alessandro Pavan, Josh Schwartzstein, and Adam Szeidl.

A number of fellow graduate students made this experience more enjoyable. James Zuberi and I each think of the other as a sidekick. However, after all of the shared experiences—late-night workouts, late(r)-night Denny's runs, Target trips, movies, Hulu watching, and tax returns sessions, to name a few—perhaps it doesn't matter which one was the sidekick. Along with James, Alex Poirier also provided great company. It was refreshing to meet someone that followed NHL hockey as closely as I did, if not more closely. Alex was also a great weightlifting buddy, especially for early-morning workouts, and study partner for probability theory. Mauricio Larrain and I became very close while taking theory and math courses during our second year. Our long, intense, and often draining study sessions eventually led to movies and meals together. Mauricio taught me the importance of discipline and time management for success as a graduate student. Although my performance in those areas still pales in comparison to his, I have learned a lot from him. He was also a very helpful source of advice when I went on the job market, and I thank him for that. Two other former fellow students, to whom I look up almost as big brothers, are Saurabh Bhargava and Prasad Krishnamurthy. Their guidance and advice from my first year onwards—particularly while I was on the job market—was extremely helpful. Fortunately, I also got along splendidly with my various officemates: Ivan Balbuzanov, Jalil Kazemitabar, Mauricio, Alex Rothenberg, Sebastian Stumpner, and James. We shared many laughs during the hours that we spent together in Evans 608-16 (later 668), occasionally prompting our neighbors down the corridor to close their doors. I conjecture that, without each other's company, each of us may have graduated in $n_i - 1$ dull years rather than in n_i memorable years. In addition to Mauricio, Mitch Hoffman, Matteo Maggiori, and Valentina Paredes were study partners that made topology and measure theory—at 8 a.m. three times a week—bearable. Ivan, Tristan Gagnon-Bartsch, Mira Frick, Maciej Kotowski, Matt Leister, Jian Li, Takeshi Murooka, Juan Sebastián Lleras, Aniko Öry, Michèle Müller, and Antonio Rosato—all fel-

low theorists—were great sounding boards for ideas. We all benefited from the theory lunch that David and Haluk organized, and from a series of informal student presentations that Aniko and Michèle organized. Juan Sebastián was a great mentor (and GSI as well as, later, co-GSI) for David's course in the PhD core microeconomics sequence. I also benefited from research meetings with Mohamed Shiliwala, Wasim Shiliwala, and Josh Tasoff, all of whom are coauthors on other projects.

The support provided by the economics department, through all phases of the program, was tremendous. Patrick Allen deserves special recognition for conducting nearly all of the behind-the-scenes work for the graduate program, including admissions, requirements, and placement. The placement chair, Professor David Card, was very proactive in making sure that all job market candidates were on track for placement. Alex Mastrangeli, Emil Schissel, Joe Sibol, and Phil Walz kept the everyday operations of the department running smoothly.

Teaching was a time-consuming but rewarding experience. Through teaching various undergraduate and graduate courses, I came to understand and appreciate the field more. I thank my students for asking thought-provoking questions that would sometimes force me to dig deep to further my understanding of a concept. I wish all of my former students the best in their endeavors.

Before deciding to come to graduate school, I consulted several former professors and teaching assistants from my undergraduate days at Cornell. Special thanks are due to Oleg Chalykh, Leonard Gross, John Hopcroft, and Tim Vogelsang for writing the recommendation letters that got me into my graduate program. Sarwat Jahan, Geeta Koolwal, and Ahmed Mahmud were former teaching assistants that encouraged me to pursue a graduate degree in economics. My uncle, Asad Alam, a professional economist with a PhD, was very encouraging as well. I may not have given serious thought to doing a PhD if not for the influence of two friends from Cornell, Salman Arif and Shaddin Dughmi, both of whom left the workforce and joined PhD programs the year before I did.

Last but certainly not least, I owe an immense amount of gratitude to my family. My parents, Akbar and Zakira Nayeem, instilled in me from an early age a deep appreciation for the value of hard work and honest effort. They seized every opportunity to nurture my curiosity about the world. I owe my academic success and passion to strive for excellence in all aspects of life to their parenting, without which I could not have even started on this journey. The outpouring of moral support that I received from my parents; from my siblings, Sarah and Aamir Nayeem; and from my paternal grandparents, Ghulam and Ayesha Nayeem, when I decided—during a time of economic uncertainty—to leave a stable private sector job with great potential upside to embark on a murky and tortuous path toward a career with uncertain—perhaps less lucrative—payoffs, but about which I was more passionate, was immensely comforting and empowering. Although my maternal grandparents, Nehaluddin and Ishrati Amanullah (may they rest in peace), are unable to witness this achievement, I feel that their prayers and wishes for my success have helped me realize it, and I hope that they would have been proud of me. My wife, Ayesha Athar, came into my life shortly after I had entered the dissertation writing phase. Through her I have acquired a second set of parents—Shahid and Shaista Athar—and a trio of new siblings—Sabuhi, Masood, and

Ahmed Athar—that have welcomed me into their family with open arms and have been genuine well-wishers. Their support has been a blessing. Ayesha moved to Berkeley just as I was preparing to enter the throes of the most difficult segment of this journey: the final push toward completion. She has been by my side for the highest peaks and lowest valleys of my graduate school career. Through it all, her companionship and unwavering encouragement have been remarkable steadying forces for which I am grateful every day. Together, she and I have concluded this long but rewarding journey. The final product follows.

Chapter 1

The Value of "Useless" Bosses

> "[W]e were hiring people just to read the reports of people who had been hired to write reports." —Jack Welch, former CEO of General Electric, on the firm's bloated organizational structure in the early 1980s, as quoted by Gordon (1996)

1.1   Introduction

Few modern archetypes are either as well known or as evocative as that of the maligned middle manager: a self-important and incompetent bureaucrat that impedes productivity. This archetype pervades popular culture (e.g., the *Dilbert* comic strip) and resonates with a common sentiment that underlies the abundance of personal anecdotes about unqualified (and overpaid) bosses. While these images are obvious caricatures of middle managers, a similar—if more tempered—view is present in a prominent and longstanding strand of the management literature that argues that forthcoming technological advances will render the roles of many middle managers redundant.[1] Still, middle managers remain present in most organizations.[2]

A leading rationale for middle management is that an organization's principal, whose time is scarce and whose capacities to acquire and process information are bounded, can benefit from delegating some administrative tasks to middle managers. The negative perceptions described above are based largely upon views that focus exclusively on middle managers' roles in monitoring workers' productivity and in propagating information across hierarchical levels.[3] In the context of such a view, a middle manager provides no marginal value to a

---

[1]This literature is surveyed in Section 1.2.

[2]As of May 2011, an estimated 5.9 million U.S. jobs fell into the category of "management occupations," not including the subcategories of "chief executives" and "legislators." These positions represented roughly 4.6 percent of total U.S. employment. This percentage has risen every year since 2006. (Source: Bureau of Labor Statistics, U.S. Department of Labor, Occupational Employment Statistics, `www.bls.gov/oes/`, Accessed: 3 November 2012.)

[3]To be sure, these two functions are salient ones that have motivated a number of classical models of hierarchies. This work is surveyed in Section 1.2.

principal that is capable of perfectly monitoring and coordinating all aspects of her organization independently. Furthermore, such a view suggests that a middle manager's interests ought to coincide with the principal's.

This paper challenges such views and their implications. It highlights a channel through which a principal—even one that needs no help with the administration of her organization—benefits from the appointment of a middle manager. It also demonstrates that a rational principal, regardless of her administrative capabilities, may choose to appoint a middle manager with beliefs that differ from her own. The manager need not hold any skill-based or informational advantage over the principal; in fact, it is helpful if the manager is perceived as misinformed by others in the organization. Furthermore, casual observation of the organization's functioning suggests that the manager is dispensable. Thus, the paper helps to reconcile the prevalence of middle management with the popular notion that middle managers are redundant or even deleterious to an organization's performance. The paper also illustrates how the principal can exploit ideological clashes within the organization to improve performance. The results are broadly consistent with observed outcomes in such disparate knowledge-based organizations as academic institutions, technology firms, and government agencies.

The above insights emerge from a view of the middle manager as a delegate to whom a principal can assign certain decisions. In particular, when an organizational decision requires unfalsifiable information that a worker is assigned to acquire at private cost, the principal faces an agency problem: a rational worker will underinvest (from the principal's perspective) in information acquisition, even if his interests are otherwise perfectly aligned with the principal's. As I show, the principal can mitigate this agency problem by delegating the decision to a "biased" manager, who demands stronger evidence (as compared to the principal) in favor of a certain action to be motivated to implement that action. Furthermore, in equilibrium, the worker's acquired information determines the manager's choice, which coincides with the choice that the principal would have made given the same information. Thus, an observer of the game that ignores crucial, but perhaps subtle, features (e.g., the endogeneity of information, the distinction between equilibrium behavior and out-of-equilibrium behavior, or the usefulness of a threat of ex post inefficiency in securing an ex ante welfare improvement) of the environment is likely to conclude that the middle manager is redundant or harmful.

The middle manager's function can be illustrated through a thought experiment. Suppose that an expert in climatology and environmental policy has been asked by a dictator to conduct an independent investigation into the question of whether global temperatures are rising due to human activity and also into the related question of whether existing environmental regulations need to be strengthened. The dictator's time is scarce, so she will appoint a deputy to evaluate the expert's report and to use it in choosing a policy. Furthermore, the expert's report is subject to external audit, and he faces severe punishment if he is caught fabricating evidence.

The expert's prior beliefs are that global temperatures are likely to be rising due to human activity, and that tighter environmental regulations probably are necessary. The expert is

concerned with economic efficiency, though: he does not want to impose an undue burden upon consumers or businesses. Consider how his incentives to exert effort in conducting the investigation vary based upon the deputy's prior beliefs. In particular, suppose that the deputy is to be chosen from the pool of individuals shown in Table 1.1.

Name	Description
Alice	Firm believer in global warming hypothesis and passionate advocate of concerted action to avert environmental disaster
Beth	Ex ante neutral and uninformed actor that wants to make optimal decision
Carol	Global warming skeptic that notes methodological flaws in supporting studies but will endorse tighter regulations if (and only if) supporting evidence is compelling
Denise	Outright denier of global warming that rejects hypothesis due to its inconsistency with her religious beliefs

Table 1.1: This table describes the four potential deputies that the dictator in the example may appoint. For concreteness, it may be helpful to think of Alice as former U.S. Vice President Al Gore (now best known as an environmental activist), of Beth as an actor behind the veil of ignorance, of Carol as physicist Richard Muller (a longtime skeptic of the global warming hypothesis that criticized previous studies but recently recanted his skepticism based upon his own study's results (Muller, 2011, 2012)), and of Denise as U.S. Representative John Shimkus (who is known to quote religious texts in disputing the validity of the global warming hypothesis (Samuelsohn, 2010)).

If Alice is appointed as the deputy, then the expert has no incentives to exert effort, since his report is unlikely to affect Alice's beliefs, and since Alice will implement the expert's ex ante preferred action by default. If the deputy is Beth, the expert has stronger incentives to exert effort, since Beth's decision will be based upon the results of the report, and since the expert wants Beth to choose the optimal policy. Now suppose that the dictator appoints Carol, who initially opposes the expert's preferred policy. In this case, the expert must work hard to overcome Carol's bias against his own policy preference. Given his belief that Carol's initial view is incorrect but flexible, the expert is confident that, by exerting enough effort, he will be able to convince Carol to change her position. Thus, the expert's incentives are even stronger when Carol is appointed as the deputy. Finally, if the deputy is Denise, the expert's incentives are destroyed altogether, since he knows that, no matter how hard he works, Denise's decision will remain the same.

As the example illustrates, a principal can improve the quality of an organizational decision by delegating it to a middle manager that is moderately biased against the worker's ex ante preferred alternative. Delegation can be interpreted as a commitment device for the principal: it allows her to implement a decision rule that demands stronger evidence for the implementation of the worker's favored alternative than the principal optimally demands ex post (i.e., once the worker's effort is sunk). Such a decision rule, by definition, is ex post inefficient from the principal's perspective: it prescribes incredible threats. In practice,

commitment to an ex post inefficient decision rule is difficult to achieve.[4] I argue that the principal may achieve such commitment by delegating control of the decision to a manager for whom the decision rule described above is ex post efficient.[5]

Section 1.2 places this paper in the context of previous work and also provides some background on middle management and hierarchies in organizations. Section 1.3 describes the setup, including the basic model, and Section 1.4 states and interprets the main results and presents illustrative examples. Section 1.5 discusses the implications of the main results and describes environments in which these implications are consistent with observed outcomes. Section 1.6 considers salient modifications and extensions (e.g., more general payoff structures and richer state and action spaces) to the basic model. Section 1.7 compares the results of this paper with insights that emerge from prior work, highlights potential directions for future research, and concludes. The main results are proved in Appendix A.1. Other technical results (most of which do not appear in the main text) are stated and proved in Appendix A.2.

1.2   Related and Background Literatures

In this section, I first explore connections between this paper and previous work. I then provide evidence of the negative views of middle managers that I consider in the paper. Finally, I outline some alternate theories regarding the value and role of middle management. These insights come both from the literature on management and from the literature on organizational hierarchies. As I explain, though, these existing theories are beside the point that I make in this paper: that even a principal that is perfectly capable of independent administration can benefit from the appointment of a middle manager, particularly of one that might appear useless—even detrimental—to the organization.

Connections with Previous Work

Two noteworthy papers that provide alternate explanations for the existence of middle management are due, respectively, to Rotemberg and Saloner (2000) and Dessein (2002). I defer detailed discussions of—and comparisons with—these papers until Section 1.7, after establishing and discussing this paper's results. These papers adopt quite different views of the middle manager, though, and neither one confronts the puzzle of the apparently "useless" middle manager.

The manager's bias creates an incentive effect—which the climatology thought experiment of Section 1.1 highlights—that forms the crux of this paper's argument. Che and Kartik (2009) identify a similar effect in an environment that involves two actors: a decision maker and an advisor. They show that a decision maker can benefit by appointing an advisor with whom she has a moderate amount of disagreement. The reason is that, as the thought

---

[4]See Gershkov and Szentes (2009) for a discussion of this point.
[5]I discuss this point in greater detail in Section 1.4.

experiment illustrates, disagreement improves the advisor's incentives to acquire information. The environment of my paper is substantively different, though, in the sense that the principal can delegate decision making and information acquisition to two separate agents: a manager and a worker. In the environment that Che and Kartik consider, the principal retains control of the decision and appoints only an advisor.

Li (2001) obtains a related insight in a group decision making context with costly information acquisition and two alternatives. Li shows that, by committing to a "conservative" decision rule that implements the group's ex ante least preferred action whenever the evidence is slightly—but not strongly—in favor of the ex ante preferred action, members of the group can mitigate the free rider problem that emerges due to the public goods nature of acquired information. Thus, the group achieves a better decision in expectation by adopting this rule. The group's commitment to a conservative decision rule in Li's framework plays a role that is analogous to the principal's delegation of the decision to a biased manager in the framework of this paper.

The insight that, in an environment with endogenous information acquisition, incentives can be improved—and welfare enhanced—through commitment to a decision rule that is not ex post efficient also emerges in a committee voting setting studied by Gershkov and Szentes (2009). Most of their analysis, however, focuses on characterizing the optimal voting scheme that employs an ex post efficient decision rule. In a similar vein, Szalay (2005) shows that a principal that delegates both information acquisition and decision making to an agent with aligned interests can benefit by restricting the agent's discretionary power so that it excludes decisions that are optimal when the acquired information is of low quality.

This paper contributes to a literature that highlights the beneficial effects of disagreement and preference differences between advisors and decision makers.[6] Such insights date back at least to the analysis of Calvert (1985), who considered the problem facing a decision maker that needs to choose among alternatives but also among advisors that hold different biases regarding the alternatives. Such a decision maker may prefer to choose a biased advisor over an unbiased one, since a biased advisor's endorsement of an ex ante unfavorable alternative reveals more information than any endorsement from an unbiased advisor does. More recently, Dewatripont and Tirole (1999) observe that advocacy (in government, law, or business, for example) improves the incentives for interested parties to gather and present information upon which decisions are based. Thus, advocacy can improve the quality of implemented decisions. Similar points are made by Rotemberg and Saloner (1995) (in the context of interdepartment conflict within firms) and by Shin (1998) (in the context of legal procedures). Van den Steen (2010a) identifies a similar effect in the context of post-merger culture clashes in firms. He notes that, while agency problems arise due to culture clashes, information collection and experimentation should be increasing in belief heterogeneity within the firm. Similarly, Prendergast (2007) argues that certain institutions—such as government bureaucracies—function better as a result of the employment of biased individuals, who have stronger incentives to exert costly effort.

---

[6]A number of papers in this area are surveyed by Sobel (2010).

The paper also contributes to the literature on authority and delegation in organizations. In addition to the paper by Dessein (2002), another celebrated work in this area is due to Aghion and Tirole (1997), who study the factors that determine the allocation of "formal" and "real" authority in organizations. An organization, in their model, consists of a principal and agent that face the problem of choosing among projects with unknown payoffs. The principal can choose to either retain or delegate decision making (i.e., formal) authority to an agent. If the actor with formal authority chooses to acquire costly information regarding project payoffs, that actor also assumes real authority, which refers to effective control over decisions. An increase in the agent's real authority provides the agent with better incentives to acquire information and participate in the organization, but it comes at the cost (to the principal) of ceding some degree of control. The main result of my paper, framed in these terms, is that a principal can benefit by ceding formal authority to a manager and real authority to a worker. Hirsch (2011) demonstrates the usefulness of "skeptical" middle managers in political organizations. Hirsch's rationale shares some notable features with the one provided here: the principal can use delegation of policymaking to a middle manager that is more responsive (i.e., prone to change policies after failure) than she is as a commitment device to induce greater effort from an expert agent that is assigned the task of policy implementation. The notion of skepticism is quite different in Hirsch's framework, however. In particular, a principal in Hirsch's environment should appoint a middle manager that is less ideologically extreme than she is. In the setting that I present, though, a principal generally benefits from appointing a more ideologically extreme middle manager.

Finally, this paper complements empirical studies that show that management practices— and, in particular, the quality and role of middle management—have significant effects on firm performance.[7] (Some of this work is discussed in greater detail in Section 1.7.) In a perspective piece, Radner (1992) notes that a high proportion of the industrialized labor force is employed by large firms, particularly in roles that involve some degree of management. Thus, Radner argues, the study of management is important to advancing the general understanding of economic activity.

Background on Middle Management

This paper deals directly with unflattering perceptions of middle managers.[8] In particular, it considers the following three negative views, each listed below with supporting evidence and discussion:

Middle managers advance their private interests. As economic agents, middle managers should be expected to make choices that maximize their own welfare. An empirical study due to Guth and MacMillan (1986) indicates that middle managers intervene

---

[7]For examples, see Wooldridge and Floyd (1990); Floyd and Wooldridge (1997); Bloom and Van Reenen (2007); Bloom, Eifert, Mahajan, McKenzie, and Roberts (2013); and a paper by Lazear, Shaw, and Stanton (2012) that bears a similar title to this paper's.

[8]Many of these perceptions are discussed by Dopson and Stewart (1990).

in (and sometimes even sabotage) decision making processes if their interests are at stake. This view also is acknowledged by Huy (2001), who notes that a middle manager sometimes is imagined as "someone who sabotages others' attempts to change the organization for the better." Furthermore, Floyd and Wooldridge (1996) observe that some widely held stereotypes portray middle managers as "politicians," "spin doctors," "subversives," and "empire builders." More recently, Osterman (2008) notes that, "[b]ecause of layoffs and because top management seems to be protecting itself, middle managers have lost their sense of loyalty to their employers."

Middle managers lack technical skills. Under the assumption that workers' day-to-day tasks are more technical than their managers' tasks, an efficient division of labor requires workers to possess comparative advantages over their managers in technical tasks. However, as noted above, middle managers act to advance their own interests, which may not be well aligned with those of the organization. Thus, it might appear that, unless a manager possesses technical expertise that complements his or her decision making abilities (and, hence, unless the manager can bring to bear on decision making a useful skill that the principal lacks), the principal cannot benefit from appointing her. In this paper, I assume the worst: the manager is no more technically qualified than the principal is. The results, therefore, may be interpreted as characterizations of a lower bound for the manager's value.

There is some evidence that workers possess absolute advantages over their managers in technical tasks. For example, Torrington and Weightman (1982) conducted a series of interviews with British middle managers that indicated a shared belief among the managers that their technical skills had atrophied. Furthermore, an empirical study by Kainen and Boyd (2007) suggests that nontechnical interpersonal (also known as "soft") skills can be more critical than technical skills for success as a middle manager (although their study suggests that technical skills, too, are helpful). This result is consistent with the view of Badawy (1982), who, as Kainen and Boyd point out, felt that technical skills were less important than interpersonal and administrative skills for success as a project manager. Kainen and Boyd also note that Souder (1987) felt that technical skills actually could hurt a manager's ability to entertain unconventional or radical ideas, thus stifling creativity in the organization. Furthermore, as Osterman (2008) points out, there is a perception that "[t]he work of middle management has been de-skilled by restructuring."

Middle managers perform redundant functions. In a well-cited piece, which was perhaps the earliest to put forth such an argument, Leavitt and Whisler (1958) predicted that, by the 1980s, advances in information technology (loosely defined to include tools and methodologies that facilitate information processing and decision making) would have changed organizational structure. Specifically, they expected such advances to simultaneously simplify (and hence lower the requisite skill levels of) many middle management tasks and raise the importance of roles that require creativity and inno-

vation. Through these channels, Leavitt and Whisler argued, the advent of information technology would effect a reduction in the size of middle management. Decades later, Drucker (1988) offered a similar prediction: he believed that, as business activities would become increasingly driven by information and reliant on technology, large businesses would find it optimal to trim their management teams and shorten their hierarchies. In Drucker's view, a large fraction of middle managers served no useful purpose other than to convey noisy information. Thus, he claimed, middle management roles would be rendered obsolete within two decades. More recently, Gratton (2011) voices a similar pessimistic forecast based upon the substitutability of monitoring technology for middle management. She observes that "workers see no value in reporting to someone who simply keeps track of what they do, when much of that can be done by themselves, their peers, or a machine." Huy (2001) notes that the "popular press" and "change-management consultants" together have created a stereotype of middle managers as "intermediaries" that "don't add value." Osterman (2008), too, notes that there is a perception that "[m]iddle managers [. . .] do little useful work for their organizations."

Although middle managers and their importance in organizations have drawn comparatively little interest from researchers, a moderately sized collection of works seeks to identify reasons for the existence of middle management. Discussions of the value of middle management emphasize middle managers' roles as champions of subordinates' ideas, implementers of top managers' strategies, synthesizers of information, and facilitators of organizational change (Floyd and Wooldridge, 1996); as therapists for their subordinates (Huy, 2001); as ambassadors between senior managers and workers (Osterman, 2008); and as coaches and mentors (Gratton, 2011). I do not dispute these characterizations (or the results of any other work that describes the value of middle management) in this paper, but I note that they are distinct from the points that I make here.

Background on Hierarchies

The question regarding the value of middle management is linked to a broader question that pertains to the role of hierarchies in organizations. Hierarchies may be viewed through various lenses, which are not mutually exclusive. I outline some canonical views of hierarchies below.[9]

Hierarchies enforce specialization. One obvious property of a hierarchy is that it induces a well-defined division of labor in an organization. The benefit of the division of labor—namely, to achieve an efficient allocation of tasks—is well established. As asserted by Gulick (1937) in his seminal work on the theory of organization, "[w]ork division is the foundation of organization; indeed, the reason for organization." Ideally,

---

[9]Comprehensive surveys are due to Holmström and Tirole (1989); Radner (1992); and Mookherjee (2006, 2012).

each member of the organization is assigned a role, either as a worker or as a manager, that exploits his or her comparative advantage in advancing the organization's objectives.[10] Crémer (1980) views the organization essentially as a collection of specialized functions in his analysis of the problem of designing an organizational structure optimally. He considers the problem from the perspective of a firm that seeks to cluster its various specialized functions into larger units in such a way that related functions are well coordinated with each other. Similarly, Beggs (2001) considers the organization as a collection of individuals with different abilities in task performance, and he shows that a hierarchical structure—with more skilled workers (whose comparative advantages lie in less common tasks) in higher positions—optimally balances a tradeoff between economizing on skilled labor and reducing costs of delay in completing tasks.

Hierarchies strengthen incentives. Another canonical view of the hierarchy emphasizes its role in providing incentives to workers. For example, in the analyses of Williamson (1967) and Calvo and Wellisz (1978), both of which seek to characterize the optimal size of a firm, the hierarchy defines supervisory relationships that discourage workers from shirking. In another paper, Calvo and Wellisz (1979) use this model to rationalize skewed wage distributions in hierarchical organizations without assuming any differences in skill level or difficulty of tasks. Qian (1994) adopts a similar view. He derives the optimal number of hierarchical levels in an organization as the optimal solution to a tradeoff between, on one hand, the principal's desire to reduce both the loss of control across hierarchical levels and the magnitude of managerial compensation and, on the other hand, the principal's desire to reduce shirking through enhanced monitoring capabilities.

Hierarchies facilitate information aggregation and propagation. According to this view, hierarchies serve to aggregate and propagate information efficiently. For example, Keren and Levhari (1979) derive the optimal span of control (i.e., number of subordinates) for a given tier in a hierarchy in which instructions, which represent production plans, are propagated from the principal to the productive workers via intermediate tiers. Geanakoplos and Milgrom (1991) also adopt this view in their model of an organization that includes managers with different comparative advantages and limited attentive capacities. Managers distribute directives among their subordinates (who also may be managers) and optimally choose to focus on those areas in which their prior abilities confer upon them comparative advantages. Garicano (2000), too, emphasizes the organization's role in propagating information, except that his view of the hierarchy, in a sense, is "bottom up" rather than "top down." He shows that, when the classification of knowledge is difficult, it is optimal for information acquisition to

---

[10]As Bylund (2011) notes in a survey piece, the benefits of specialization have been acknowledged, at latest, since the pioneering work of Smith (1776). Usher (1920) and Dobb (1928) believed, in fact, that firms came into existence to capture the gains that could be realized under the division of labor. This view was firmly rejected by Coase (1937), who argued that the price mechanism of the market economy was sufficient for this purpose.

be organized according to a hierarchical structure—in which the most basic and most commonly needed pieces of information are acquired by members of the bottom tier— that is similar to the one derived by Beggs (2001). The information-based view of organizations is pushed to an extreme by, for example, Radner (1993) and Van Zandt (1999), who model (boundedly rational) managers as information processors with finite capacities and characterize optimal arrangements of these procesors. In a similar spirit, Bolton and Dewatripont (1994) model organizations as communication networks that involve groups of agents that specialize in processing different types of information.

Nontechnical reasons for the prevalence and persistence of hierarchies are discussed by Leavitt (2003). He notes that hierarchies fulfill certain psychological needs by providing order and measures of success. Leavitt also argues that hierarchies have been strengthened, rather than weakened (as some, perhaps even he at the time of writing his coauthored 1958 piece, may have expected) by paradigm shifts and technological changes that have occurred over the past several decades.

The literature surveyed above derives some key properties (e.g., size, wage distribution, span of control, specialization, and managers' abilities) and rationalizes some well-known features of hierarchical organizations. In doing so, it sketches some of the channels through which middle management derives its value. However, the papers in this literature do not consider the undesirable characteristics of middle managers that I outlined above. Thus, the prevalence of middle management and its widespread lamentation have not been reconciled.

1.3 Setup

In an influential piece, Hayek (1945) argued that the central problem of economic organization is the efficient use of specialized and dispersed information. In a similar spirit, I view the organization as a collective enterprise for specialized information acquisition and decision making. This view is similar to ones adopted, for example, by Sah and Stiglitz (1985); Bolton and Dewatripont (1994); Aghion and Tirole (1997); and Garicano (2000). Before developing the model, I describe this view of the organization in greater detail.

View of an Organization

The term *organization*, in this paper, refers to a set of actors (*members*), each of whom can be categorized as exactly one of two types: an actor is either an information gatherer (*worker*) or a decision maker. Each member's type is exogenous and fixed. Every organization includes at least (and possibly only) a *principal*, a decision maker whose objective coincides with the organization's. All decision makers aside from the principal are *managers*, and all members of the organization aside from the principal are *agents*.[11] Figure 1.1 illustrates an example

---

[11]For simplicity, I ignore cases with multiple principals, although some organizations (e.g., corporations with multiple shareholders) are modeled more accurately with multiple principals.
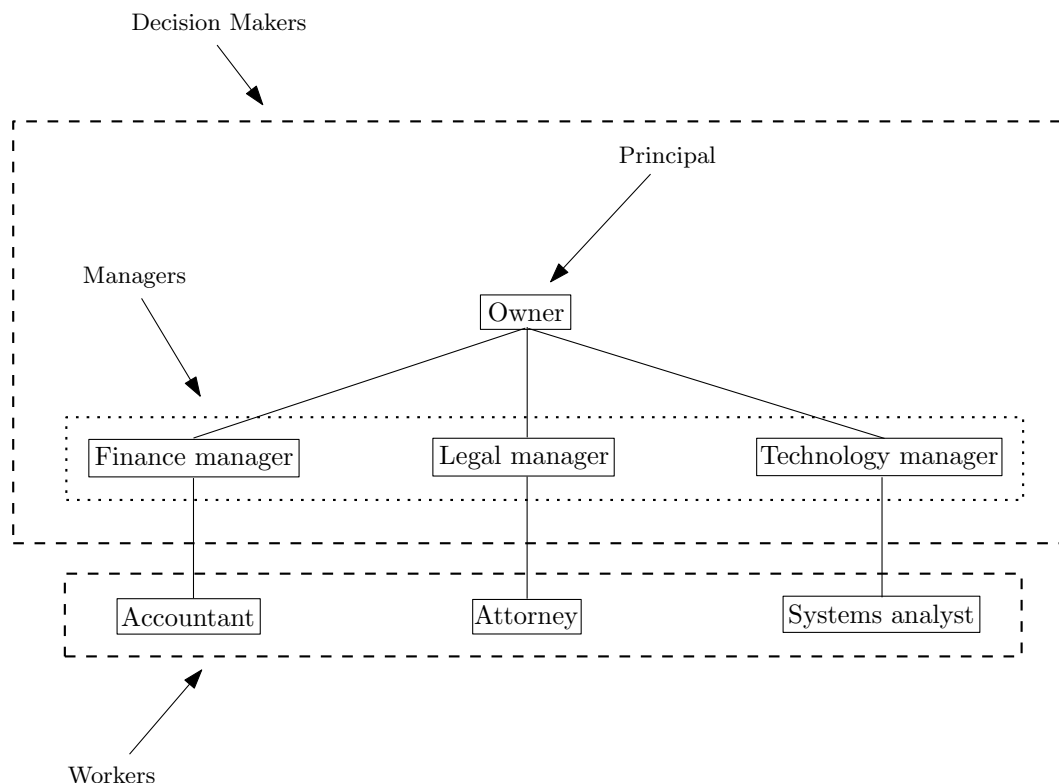
of such an organization.

Figure 1.1: A partial organizational chart for a small firm that has three levels in its hierarchy

The members of the organization play a three-period game of imperfect information, in which their ultimate task is to make a set of decisions. The source of uncertainty is an unobserved random vector, which corresponds to the state of the world. In the first period, the principal assigns each worker the task of acquiring information about at most one component of the state. In addition, she assigns each manager the task of making at most one decision. (The principal retains residual control over any unassigned decisions.) In the second period, each worker that has been assigned a component of the state variable exerts effort to acquire costly private information regarding that component and then makes a public announcement regarding the acquired information. (The announcements need not be either truthful or precise.) In the third period, decision makers choose actions. Each member's welfare depends upon three factors: his or her own action, the set of implemented decisions, and the realization of the state.

This model of the organization, of course, abstracts away from many features that exist in different types of organizations. Notable omissions are transfer payments (which are considered in Section 1.6 and are shown to leave the results largely unchanged), reputational considerations, and production. This simple setup, however, is sufficient to capture the effects with which the paper is concerned.

The Model

The principal of an organization faces a choice between two actions, α and β. For example, suppose that the principal is the proprietor of a small manufacturing firm that needs to expand its productive capacity. As a result, the proprietor is looking to build a new plant, and she is trying to decide between potential sites in Avon (α) and Bedford (β). The payoff from each site depends upon the realization of an unobserved state of the world, which takes either of two values: A (in which case building at Avon will be more profitable) or B (in which case building at Bedford will be more profitable).[12] According to the principal's prior beliefs, the probability that the state is A is $\pi \in (0,1)$. The principal's ex post utility function, defined over the set of action and state pairs, is

$$u : \{\alpha, \beta\} \times \{A, B\} \to \mathbb{R}_-$$

$$(\eta, \omega) \mapsto \begin{cases} -\lambda & \text{if } \eta = \alpha \text{ and } \omega = B, \\ -(1 - \lambda) & \text{if } \eta = \beta \text{ and } \omega = A, \\ 0 & \text{otherwise.} \end{cases} \tag{1.1}$$

This formulation of the utility function is typical in models of decision making in juries.[13] Here, λ is the minimal probability that the principal's beliefs must assign to state A for the principal to prefer α over β. In particular, for a fixed value of π, the higher λ is, the stronger is the principal's ex ante preference toward β.[14]

The principal can appoint any single member of a set of workers to obtain better information regarding the realization of the state.[15] All candidate workers are specialists—in particular, the principal cannot, by herself, obtain additional information about the state. The appointed worker can exert effort to acquire an informative signal regarding the state. When the worker exerts $e \geq 0$ units of effort, the conditional distribution (given the realized state, $\omega \in \{A, B\}$) of the acquired signal, $s \in \{a, b\}$, satisfies

$$\Pr(\{s = a\}|\{\omega = A\}) = \Pr(\{s = b\}|\{\omega = B\}) = \frac{1}{2} + q(e).$$

The term $q(e)$ represents the quality (i.e., informativeness) of the signal that the worker acquires by exerting effort level e. I assume that $q(0) = 0$, $\lim_{e \to \infty} q(e) = \frac{1}{2}$, and $q(\cdot)$ is increasing, strictly concave, and twice differentiable. (For example, let $q(e) \equiv \frac{e}{2e+1}$.) I

---

[12]For convenience, I assume that it is common knowledge both that the firm must choose to build at exactly one of the two sites and that potential plants at the two sites will yield different payoffs.

[13]See, for example, Feddersen and Pesendorfer (1998). In the jury setting, the parameter λ represents the "threshold of reasonable doubt," the minimal probability that a juror's posterior beliefs must assign to the event that a defendant is guilty for the juror to prefer the defendant's conviction over the defendant's acquittal.

[14]See Section 1.6 for a discussion of the generality of utility functions of the form of (2.1).

[15]I use the term "appoint" rather than "hire," because the principal may seek help from existing members of the organization. Furthermore, until Section 1.6, the model abstracts away from transfer payments, outside options, and other considerations that arise in hiring processes.

assume, also, that both the signal's content and its quality are observable.[16] The signal's quality and the worker's effort level are not independently verifiable, though.[17] Effort is costly for the worker: when he exerts e units of effort, he incurs a private cost of $c(e)$, where $c(0) = c'_+(0) = 0$ and $c(\cdot)$ is strictly convex, twice differentiable, and unbounded.[18] (For example, let $c(e) \equiv \frac{e^2}{2}$.) Note that, at all positive effort levels, $c(\cdot)$ is increasing.

In the example, the worker is a financial analyst that can forecast the cashflows that are associated with building a plant at either site. The quality of his forecast depends upon the level of effort that he exerts. He uses his forecast to produce a detailed report that makes both the forecast's result and its accuracy apparent to any member of the organization that reads the report.[19]

The worker's prior beliefs place probability $\pi_W \in (0,1)$ on state A.[20] His ex post utility function is

$$u_W : \{\alpha, \beta\} \times \{A, B\} \to \mathbb{R}_-$$

$$(\eta, \omega) \mapsto \begin{cases} -\lambda_W & \text{if } \eta = \alpha \text{ and } \omega = B, \\ -(1 - \lambda_W) & \text{if } \eta = \beta \text{ and } \omega = A, \\ 0 & \text{otherwise}, \end{cases}$$

---

[16]The signal need not be literally observable; it may be the case, instead, that the worker obtains the signal and then reveals it via a report that conveys its realization and quality. In this case, sufficiently severe penalties for offenses—such as lying or tampering with evidence—that impede truthful revelation can ensure the optimality of truthful revelation. The feasibility of any such penalty relies upon the detectability of the associated offense (e.g., through the existence of an audit procedure that exposes the offense with sufficiently high probability), the verifiability of the occurrence of the offense (e.g., through the accepted legitimacy and transparency of an audit procedure that exposes the offense), and the credibility of the principal's commitment to impose the penalty.

[17]This assumption reflects the fact that the worker's acquired information is specific to the organization; its quality, therefore, is apparent only to members of the organization. An implication of this assumption is that, even in Section 1.6, where I consider transfer payments, the principal cannot write contracts based upon the worker's effort level.

[18]$c'_+(0)$ denotes the right-hand derivative of $c(\cdot)$ evaluated at 0. That is, $c'_+(0) \equiv \lim_{e\downarrow 0} \frac{c(e)}{e}$. (Note that $c'(0)$ does not exist, since 0 is a boundary point of the domain of $c(\cdot)$.)

[19]Note that the report is unfalsifiable. In the example, the proprietor can detect fabricated information in the report (with sufficiently high probability) and commits to firing the analyst (and possibly even taking legal action) if she finds that he has misrepresented his acquired information.

[20]Throughout this paper, I remain agnostic about "correct" beliefs. The point, however, is that, even if both the principal and the worker have initial beliefs that are known to differ from the manager's, the principal may prefer to appoint the manager rather than to retain control over the decision. Furthermore, while it is nonstandard—and may seem dubious in light of canonical results (e.g., ones due to Blackwell and Dubins (1962); Aumann (1976); and Geanakoplos and Polemarchakis (1982)), to assume that actors' prior beliefs differ, open disagreement is observed and acknowledged in a number of settings. Some notable recent attempts have been made to explain this outcome in Bayesian frameworks. For example, Dixit and Weibull (2007) provide conditions under which polarization can occur between parties' beliefs following the revelation of a common signal, and Acemoglu, Chernozhukov, and Yildiz (2007, 2009) present learning foundations that explain the persistence and common knowledge of disagreement. Other papers, besides the ones already cited, that examine the consequences of disagreement in organizational settings are due to Banerjee and Somanathan (2001) and Van den Steen (2005, 2010b).

where $\lambda_W \in (0,1)$. From this formulation of the worker's utility function, it is evident that the worker's interests are at least partially aligned with the principal's: under each state, the worker favors the same action as the principal. However, the worker may hold different beliefs and realize different payoffs from the principal. A candidate worker can be identified by a quadruple of values for the four parameters described above: π_W, λ_W, $q(\cdot)$, and $c(\cdot)$. Such a quadruple is a candidate worker's "type," which sometimes will be denoted by t_W. Let $\mathcal{T}_W$ denote the type space of candidate workers.

Before discussing candidate managers, I define some notions that are necessary for understanding the manager's role in the model. First, I formalize the notion of "bias."

Definition 1.1 (Bias). Consider any member of the organization.

  (i) The member is *unbiased* if, regardless of the signal's quality, she prefers α when the signal's realization is a and β when the signal's realization is b.

 (ii) If, for some positive level of signal quality, the member favors β even when the signal's realization is a, she is *biased in favor of β*.

(iii) If, for some positive level of signal quality, the member favors α even when the signal's realization is b, she is *biased in favor of α*.[21]

Note that the principal is unbiased if and only if $\pi = \lambda$. She is biased in favor of α if and only if $\pi > \lambda$, and she is biased in favor of β if and only if $\pi < \lambda$. (Analogous expressions characterize the worker's bias.) In the context of the manufacturing example, an unbiased actor initially is indifferent between building the plant at Avon and building it at Bedford. Her preferences, therefore, will be determined by the analyst's report. An actor that is biased in favor of building at Avon prefers to build the plant at Avon unless the analyst's report not only indicates that building at Bedford is optimal, but also is of sufficiently high quality to persuade her of the optimality of building at Bedford.

Definition 1.2 (Preference alignment). Two actors have *aligned preferences* if, regardless of the signal's quality and realization, the two actors' preferences over actions coincide.

As Definition 1.2 indicates, the worker's preferences are aligned with the principal's if the worker and principal always agree about which action is, in expectation, better. In the manufacturing example, the analyst and proprietor have aligned preferences if and only if, regardless of the content and quality of the analyst's forecast, the analyst and proprietor agree about which of the two sites is better. It is worth noting that an actor's preferences are aligned with those of an unbiased actor if and only if both actors are unbiased.

---

[21]Although it may not be immediately obvious from the definitions, every member of the organization can be classified into exactly one of the categories "unbiased," "biased in favor of α," and "biased in favor of β." To see this fact, note that, if a member of the organization favors α when she sees an a signal of quality $q \in \left[0, \frac{1}{2}\right)$, she also favors α when she sees an a signal of quality $q' \in \left(q, \frac{1}{2}\right)$. (An analogous statement holds for β and b as well.)

Now return to the model. In addition to appointing the worker, the principal also can appoint a (middle) manager to consider the worker's report and make a decision. Like the principal, the manager is not a specialist that can acquire information regarding the state. However (also like the principal), she can interpret the realized signal and its quality. A candidate manager can be identified by her prior beliefs and preferences, which are described by her values for the parameters $\pi_M, \lambda_M \in (0, 1)$, respectively. As it turns out, however, the set of these pairs is redundant as a type space for candidate managers. Lemma 1.1 shows that, without loss of generality, one can assume that all candidate managers share a common value of λ_M and vary only in their values of π_M.

Lemma 1.1. *Let $p, \ell \in (0, 1)$. For any $\ell' \in (0, 1)$, there exists a unique $p' \in (0, 1)$ such that a manager with attributes $\pi_M = p$ and $\lambda_M = \ell$ and a manager with attributes $\pi_M = p'$ and $\lambda_M = \ell'$ have aligned preferences.*

The result of Lemma 1.1 is illustrated in Figure 1.2. Since the mapping between the worker's signal and a manager's preferences over actions fully characterizes that manager's behavior, two candidate managers with aligned preferences are "equivalent." Lemma 1.1 thus allows, without loss of generality, a normalization of all candidate managers' λ_M parameters.

Assumption 1.1 (Unidimensional manager type space). Each candidate manager has $\lambda_M = \frac{1}{2}$.

Assumption 1.1 allows a candidate manager to be identified by her value for the parameter π_M,[22] which, from now on, will be called her "type." Accordingly, the set of candidate manager types, $\mathcal{T}_M$, will be a subset of $(0, 1)$. The only restriction that I place on $\mathcal{T}_M$ is that $\frac{\pi - \pi\lambda}{\pi + \lambda - 2\pi\lambda}$ (which, as the proof of Lemma 1.1 shows, is the type of candidate manager whose preferences are aligned with the principal's) belongs to $\mathcal{T}_M$. The interpretation is that the principal has the option to retain control of the decision.

Note that a candidate manager's type indicates a degree of skepticism toward the optimality of either action. Definition 1.3 formalizes this notion.

Definition 1.3 (Skepticism). Consider a manager of type π_M.

(i) The manager's *skepticism* is given by $\left| \pi_M - \frac{1}{2} \right|$.

(ii) If $\pi_M < \frac{1}{2}$, the manager is *skeptical regarding α*.

(iii) If $\pi_M > \frac{1}{2}$, the manager is *skeptical regarding β*.

By Definition 1.3, the more "extreme" a manager's prior beliefs are (in the sense of putting very high probability on either of the states), the more skeptical she is. Note that an unbiased manager (who, ex ante, is indifferent between the two actions) has $\pi_M = \frac{1}{2}$ and skepticism 0. Furthermore, a manager that is skeptical regarding either action is biased in favor of the opposite action.

---

[22]The result of Lemma 1.1 and the permissibility of Assumption 1.1 reflect the fact that, with state-dependent utility functions, it is impossible to uniquely identify both utility functions and subjective beliefs from observed choice behavior. See Dekel and Lipman (2010) for an illustration of this point.
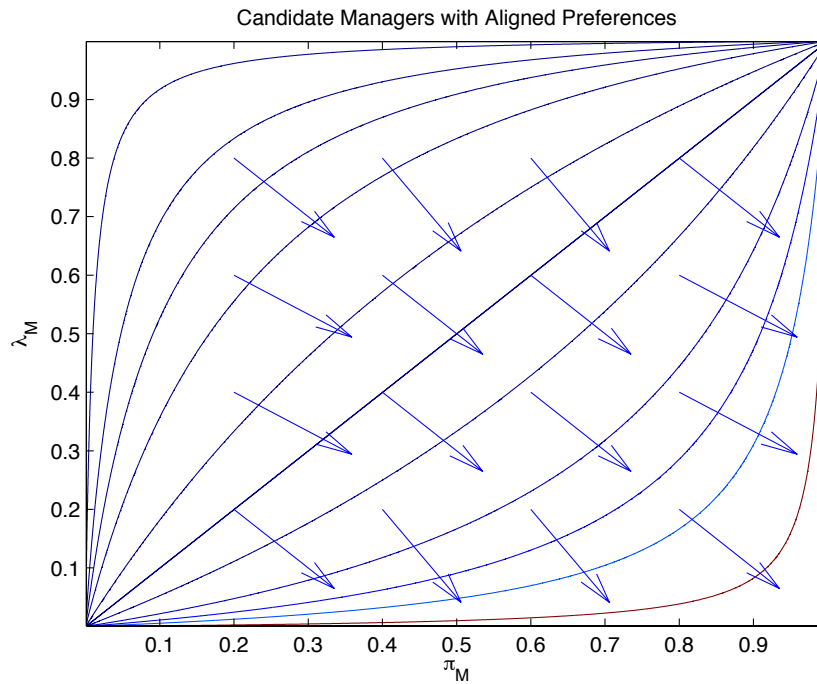
Figure 1.2: Consider the equivalence relation of preference alignment, defined on the set of attribute (i.e., (π_M, λ_M)) pairs of candidate managers. Each curve in this figure is a locus of attribute pairs that represents an equivalence class of the relation, and the arrows point toward curves of attribute pairs that are associated with candidate managers that, ex ante, have stronger biases toward α. Since any two candidate managers whose attributes lie on a common curve have aligned preferences, they will behave identically if appointed. Furthermore, since, for any fixed value of λ_M, each curve contains exactly one point that has this value of λ_M, it is sufficient to fix λ_M at an arbitrary value for all candidate managers, and to then identify them only by their values of π_M. Assumption 1.1 fixes $\lambda_M = \frac{1}{2}$ for all candidate managers.

In the example, the manager is any agent that is familiar with the organization and can use the analyst's report to make a decision. She is not a financial analyst that can conduct her own forecast. Furthermore, she may have her own beliefs and agenda, which are distinct from the proprietor's and the analyst's. For example, the manager might reside in Avon and believe that, if the plant were to be built there, her home's value would fall. Effectively, though, candidate managers are assumed to vary only in their beliefs.

Because the baseline model abstracts away from transfer payments, participation constraints do not arise.[23] Thus, any agent that the principal appoints will participate, though obviously in a rational manner. All features of the game (including the actors' attributes) are common knowledge.[24] Figure 1.3 summarizes the game's timing.

---

[23]Transfer payments and participation constraints are discussed in Section 1.6.

[24]The assumption that actors' types are common knowledge is stark but arguably is less stark than the
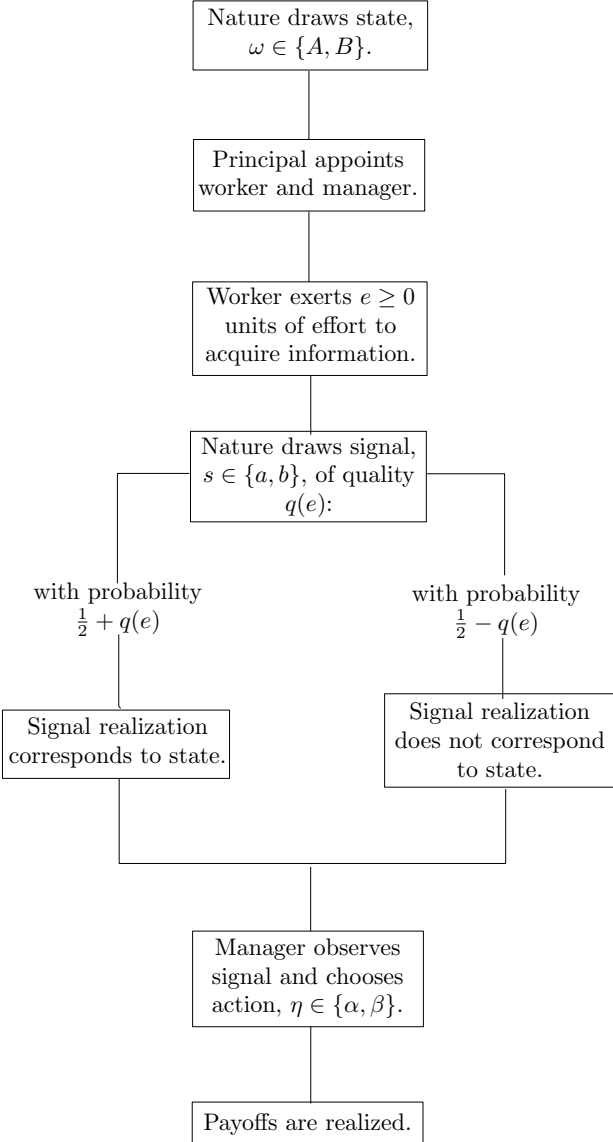
Figure 1.3: The timing of the game

Formal Description of Solution Concept

An obvious desideratum for a solution concept in this game is that each player should choose his or her action optimally given his or her observation of the history, (correct) beliefs of players' future actions, and (private) beliefs about the state of the world. Definitions 1.4, 1.5, and 1.6 together provide a precise characterization of such a solution concept.

---

standard assumption of common prior beliefs.

Definition 1.4 (Strategies)**.** Consider the game shown in Figure 1.3, in which the players include the principal, candidate managers, and candidate workers. An *appointment* refers to an element of the set, $\mathcal{T}_M \times \mathcal{T}_W$, of pairs of manager types and worker types. A *decision rule* is a mapping from the set, $\left[0, \frac{1}{2}\right) \times \{a, b\}$ of signal quality and signal realization pairs, to the set, $\{\alpha, \beta\}$, of actions. An *exertion rule* is a mapping from the set, $\mathcal{T}_M$, of manager types to the set, $\mathbb{R}_+$, of feasible effort levels.

Definition 1.4 defines pure strategies for each type of player in this game. An appointment refers to the principal's choices of a manager and a worker. A decision rule specifies a manager's behavior at the decision making stage. An exertion rule describes a worker's effort level based upon the characteristics of the appointed manager. Together, when specified for each player (including all candidate managers and candidate workers that are not chosen by the principal), these objects define a strategy profile.

Definition 1.5 (Strategy profile)**.** A *strategy profile* in the game depicted in Figure 1.3 refers to a triple that consists of:

- An appointment

- A mapping from $\mathcal{T}_M$ to the set of decision rules

- A mapping from $\mathcal{T}_W$ to the set of exertion rules

In Definition 1.5, a simple appointment is sufficient to describe the behavior of the principal, whose type is fixed and exogenous. The appointed manager and appointed worker, however, have endogenous types, which the principal chooses. Since candidate managers and candidate workers can be identified only by their types, the principal's beliefs about candidate managers' behavior can be summarized by a mapping from the set, $\mathcal{T}_M$, of manager types to the set of decision rules. Similarly, the principal's beliefs about candidate workers' behavior can be summarized by a mapping from $\mathcal{T}_W$ to the set of exertion rules.[25] Now the appropriate solution concept can be defined.

Definition 1.6 (Equilibrium)**.** An *equilibrium* in the game depicted in Figure 1.3 is a strategy profile that satisfies the following conditions:

(i) The appointment maximizes the principal's expected utility, given the anticipated behavior of candidate managers and candidate workers (as captured by the second and third components of the strategy profile).

(ii) The decision rule for each candidate manager maximizes her expected utility. (In cases of indifference, the decision rule selects the action that corresponds to the signal's realization.)

---

[25]Myerson (1998) provides an elaborate discussion of this point in the more general setting of games with population uncertainty, in which the number of players in the game is a random variable, and the players' types also are randomly determined. In Myerson's terminology, the second and third components of the triple described in Definition 1.5 are called "pure strategy functions."

(iii) The exertion rule for each candidate worker maximizes his expected utility, given the anticipated behavior of the appointed manager, as captured by the second component of the strategy profile. (In cases of indifference, the exertion rule selects the highest optimal effort level.)

Embedded in conditions (ii) and (iii) of Definition 1.6 are "tiebreaking" rules for both the manager and the worker. These rules are imposed for technical convenience and are not necessary to obtain the main insights that emerge from the results. I will discuss these rules, and the effects of relaxing them, as they are invoked in the analysis. To emphasize them, and for easy reference, I state them formally as an assumption:

Assumption 1.2 (Tiebreaking rules). If the appointed worker is indifferent among effort levels, he chooses the highest one from the optimal set. If the appointed manager is indifferent between actions, she chooses the one that corresponds to the signal's realization.

Inspection of Definition 1.6 and Figure 1.3 shows that, if an equilibrium exists, it can be found by a generalized backward induction procedure. Hence the equilibria that I have defined prescribe optimal play at each information set for each player, as in perfect Bayesian equilibrium. In particular, this solution concept fulfills the criterion that I noted above. Since much of the paper is devoted to characterizing the effects of the principal's appointment on her welfare and on observed outcomes, though, the analysis will focus mostly on non-equilibrium outcomes, in which the principal's appointment is suboptimal.[26]

The following terms are useful in describing outcomes of the game. They are equivalent and will be used interchangeably to shift the emphasis as needed.

Definition 1.7 (Informing, influencing, and rubberstamping)**.** Fix a quality level for the worker's signal. The signal *informs* an actor's decision if the actor would choose α if the signal's realization were a and would choose β otherwise. If the worker's signal informs an actor, the worker *influences* that actor, and the actor *rubberstamps* the worker's signal.

1.4 Main Results

To aid in the exposition of the results, I begin by considering a simplified version—which I call the "simple case"—of the environment. In the simple case, the principal has attributes $\pi = \frac{1}{2}$ and $\lambda = \frac{1}{2}$, and there is a single worker. This worker's ex ante beliefs and preferences are represented by the parameters $\pi_W = \frac{1}{2}$ and $\lambda_W = \frac{1}{2}$,[27] and his information acquisition capabilities are summarized by the functions $q(\cdot)$ and $c(\cdot)$. Given that only one worker is available, the principal's problem reduces to the optimal appointment of a manager. In the simple case, all possible manager types are available: $\mathcal{T}_M \equiv (0, 1)$. The analysis of the simple case will demonstrate the principal's incentive to delegate the decision to a biased manager. After discussing the simple case, I consider the general case.

---

[26]Another issue is that the existence of equilibrium depends upon the structure of the space of possible appointments. This point is discussed further in Section 1.4.

[27]Note that both the principal and worker are unbiased and have aligned preferences.

The Simple Case: Incentives to Appoint a Manager

I compare two regimes. In the first regime, the principal retains control of the decision. That is, the appointed worker acquires a signal, and the principal uses it to choose an alternative. The outcome of this regime will serve as a baseline that can be compared with the outcome of the second regime, in which the principal delegates the choice of action to a biased manager.

Principal Retains Control

Suppose that the principal retains control of the decision. Note that the worker makes only one choice: the level of effort to expend in acquiring information. Proposition 1.1 characterizes the outcome.

Proposition 1.1. *In the simple case, when the principal retains decision making authority, the worker's effort level, e_r^*, is positive and is characterized by $q'(e_r^*) = 2c'(e_r^*)$. The principal rubberstamps the worker's signal and achieves an expected payoff of $U_r^* = \frac{q(e_r^*)}{2} - \frac{1}{4}$.*

Thus, when the principal retains control over the decision, the worker exerts a positive level of effort, denoted by e_r^*. Since the worker's welfare is affected by the decision, he finds it worthwhile to exert some effort in ensuring that, with sufficiently high probability, the principal makes the optimal choice. In the manufacturing example, the analyst produces an informative report that the proprietor uses to decide whether to build the plant at Avon or at Bedford. Because the proprietor is initially indifferent between the two locations, she bases her decision upon (i.e., rubberstamps) the analyst's report.

Can the unbiased principal do better by delegating control to a biased manager? As I demonstrate in the following section, she can.

Principal Delegates Decision to a Manager

Suppose that the principal appoints a manager, who observes the worker's acquired signal before choosing an alternative. As Proposition 1.1 showed, any signal of positive quality will inform the unbiased principal's decision. How strong a signal is required to inform the manager's decision? Lemma 1.2 provides an exact answer.

Lemma 1.2. *The worker's signal informs the decision of a manager of type π_M if and only if the signal's quality is high enough to overcome the manager's skepticism: $q(e) \geq \left|\pi_M - \frac{1}{2}\right|$.*

As Lemma 1.2 indicates, the worker can influence the manager only by extracting evidence of sufficiently high quality to compensate for the manager's skepticism; otherwise, regardless of the signal's realization, the manager chooses the action that is optimal according to her prior beliefs.[28] Therefore, when the principal appoints a manager of type π_M, the worker

---

[28]The weak inequality in Lemma 1.2 relies upon Assumption 1.2, according to which the manager rubberstamps the signal when she is indifferent between actions. The most reasonable relaxation of this assumption would require the indifferent manager to choose the action toward which she is biased, in which case the inequality in Lemma 1.2 becomes strict.

chooses an effort level, $e_d^*(\pi_M)$, that satisfies

$$e_d^*(\pi_M) \in \arg\max_{e \geq 0} U_W(e\,;\pi_M), \tag{1.2}$$

where the function $U_W(\cdot\,;\pi_M)$ maps the worker's effort to his expected utility, given the manager's decision rule:[29]

$$U_W(e\,;\pi_M) \equiv \begin{cases} -c(e) - \frac{1}{4} & \text{if } e < q^{-1}\left(\left|\pi_M - \frac{1}{2}\right|\right), \\ \frac{q(e)}{2} - c(e) - \frac{1}{4} & \text{otherwise.} \end{cases} \tag{1.3}$$

To understand (1.3), return to the manufacturing firm example. Suppose that the appointed manager is skeptical regarding the optimality of building a plant at Avon. This manager will choose to build the plant at Avon if and only if two conditions are met. First, the analyst's report must indicate that building at Avon is optimal. Second, the analyst's report must be of sufficiently high quality (or, equivalently, the analyst must exert enough effort in producing the report) to undo the manager's skepticism. If the analyst's effort level is too low (i.e., below $q^{-1}\left(\left|\pi_M - \frac{1}{2}\right|\right)$), then his effort is wasted, since his report does not inform the manager's decision. Thus, the analyst's expected utility, as a function of his effort level, rises discontinuously at $q^{-1}\left(\left|\pi_M - \frac{1}{2}\right|\right)$.

This discussion suggests that the principal can strengthen the worker's incentives to acquire information by delegating the decision to a biased manager. Since the worker's cost of effort is unbounded, though, as the worker increases his level of exertion, his cost eventually will exceed the benefit that he obtains from the implementation of the optimal action. Clearly it would be irrational for the worker to choose such a high effort level. In particular, if the manager is so skeptical that an effort level of this magnitude is required to convince her to consider the worker's report, the worker's incentives to acquire information will be destroyed completely, and he will exert no effort.

This point can be illustrated more concretely in the context of the running example. If the appointed manager is extremely skeptical regarding the optimality of building the plant at Avon, then the analyst may find it too costly to produce a forecast of sufficiently high accuracy to (possibly, depending on the result of the forecast) induce the manager to have the plant built at Avon. In particular, the analyst would find it strictly better to exert no effort at all and to essentially gamble that the manager's preferred choice (i.e., building the plant at Bedford) turns out to be optimal. Furthermore, since any amount of effort that is insufficient to convince the manager to build the plant at Avon will be wasted, the analyst exerts no effort in producing the forecast: his report is uninformative.

It should be clear, now, that the choice of manager is a crucial and delicate one for the principal. A "healthy dose of skepticism" can strengthen the worker's incentives to acquire information of high quality. Too much skepticism, though, will destroy the worker's incentives completely. How much skepticism is too much? The critical threshold is defined below:

---

[29]If Assumption 1.2 is relaxed as in Footnote 28, the inequality in (1.3) becomes weak.

Definition 1.8 (Tolerance for skepticism). Let ϕ denote the worker's *tolerance for skepticism*: the maximal level of skepticism that the appointed manager, in the simple case, may hold in order for the worker to exert a positive level of effort (and hence to obtain an informative signal).[30]

With this definition and notation in hand, it is possible to characterize the worker's behavior and the principal's welfare under delegation.[31]

Proposition 1.2. *In the simple case, when the principal appoints a manager of type π_M, the worker's effort level, $e_d^*(\pi_M)$, is determined by the manager's skepticism. In particular,*

$$e_d^*(\pi_M) = \begin{cases} e_r^* & \text{if } 0 \le \left|\pi_M - \frac{1}{2}\right| \le q(e_r^*), \\ q^{-1}\left(\left|\pi_M - \frac{1}{2}\right|\right) & \text{if } q(e_r^*) < \left|\pi_M - \frac{1}{2}\right| \le \phi, \\ 0 & \text{if } \phi < \left|\pi_M - \frac{1}{2}\right| < \frac{1}{2}. \end{cases}$$

If $\left|\pi_M - \frac{1}{2}\right| \le \phi$, the manager will rubberstamp the worker's signal. Otherwise, the manager will implement the action toward which she is biased. The principal's expected utility, $U_d^*(\pi_M)$, is

$$U_d^*(\pi_M) = \frac{q\left(e_d^*(\pi_M)\right)}{2} - \frac{1}{4} = \begin{cases} \frac{q(e_r^*)}{2} - \frac{1}{4} & \text{if } 0 \le \left|\pi_M - \frac{1}{2}\right| \le q(e_r^*), \\ \frac{\left|\pi_M - \frac{1}{2}\right|}{2} - \frac{1}{4} & \text{if } q(e_r^*) < \left|\pi_M - \frac{1}{2}\right| \le \phi, \\ -\frac{1}{4} & \text{if } \phi < \left|\pi_M - \frac{1}{2}\right| < \frac{1}{2}. \end{cases}$$

As Proposition 1.2 indicates, the worker's effort level is determined completely by the appointed manager's skepticism, and there are three substantively distinct cases to consider. Figures 1.4, 1.5, and 1.6, respectively, illustrate the three cases. In each of the figures, $U_W(\cdot\,; \pi_M)$, which represents the worker's expected utility as a (piecewise) function of his effort level, is plotted as the solid curve. The dashed curve depicts the worker's expected utility, when the principal's retains control, for $e \in [0, q^{-1}\left(\left|\pi_M - \frac{1}{2}\right|\right))$. In this range of effort levels, the realized signal is too uninformative to overcome the appointed manager's skepticism.

---

[30] The technical definition of ϕ is the unique positive fixed point of the function $x \mapsto 2c\left(q^{-1}(x)\right)$, defined on $[0, \frac{1}{2})$. Lemma A.1 in Appendix A.2 guarantees the existence and uniqueness of ϕ.

[31] Proposition 1.2 depends upon Assumption 1.2 for two reasons. First, if an indifferent manager does not rubberstamp the signal, then, if the manager's skepticism is in the range $(q(e_r^*), \phi)$, the worker will prefer to exert a slightly higher amount of effort than $q^{-1}\left(\left|\pi_M - \frac{1}{2}\right|\right)$, since (as noted in Footnote 29) the inequality in (1.3) becomes weak. Strictly speaking, the worker's optimal effort level is undefined in this case. Second, if the manager's skepticism is ϕ, Assumption 1.2 guarantees that the worker is indifferent between exerting no effort and exerting $q^{-1}(\phi)$ units of effort (which is sufficient to influence the manager), and that the worker chooses to exert $q^{-1}(\phi)$ units of effort. If, instead, the worker exerts no effort when $\left|\pi_M - \frac{1}{2}\right| = \phi$, then the inequalities in the expressions for $e_d^*(\pi_M)$ and $U_d^*(\pi_M)$ have to be adjusted accordingly. It will become clear shortly that these technical complications do not invalidate the main insights; see Footnote 32.
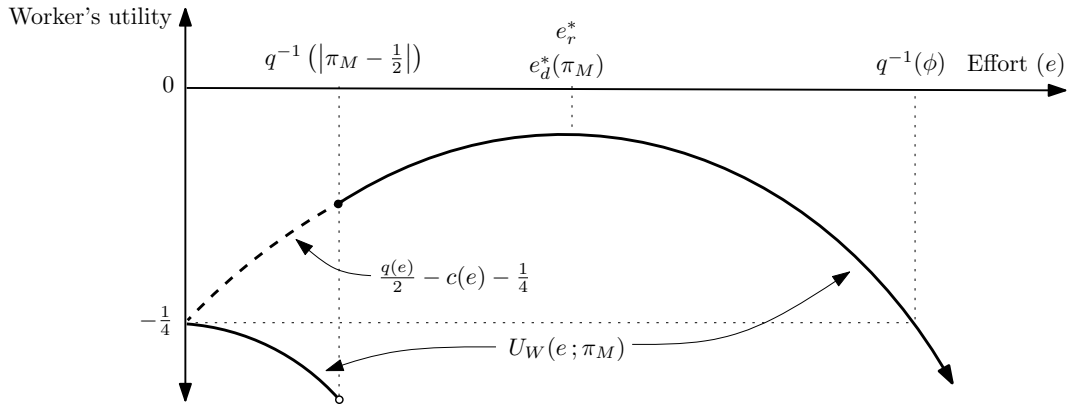
Figure 1.4: The appointment of a mildly skeptical manager (i.e., of one with $0 \leq \left|\pi_M - \frac{1}{2}\right| \leq q(e_r^*)$) does not affect the worker's incentives. The worker's optimal effort level, e_r^*, under the principal's retention of control, results in a signal that is of sufficiently high quality to inform the manager's decision (i.e., $q(e_r^*) \geq \left|\pi_M - \frac{1}{2}\right|$). Thus $e_d^*(\pi_M) = e_r^*$: the worker's effort level is unchanged from the case in which the principal retains control.
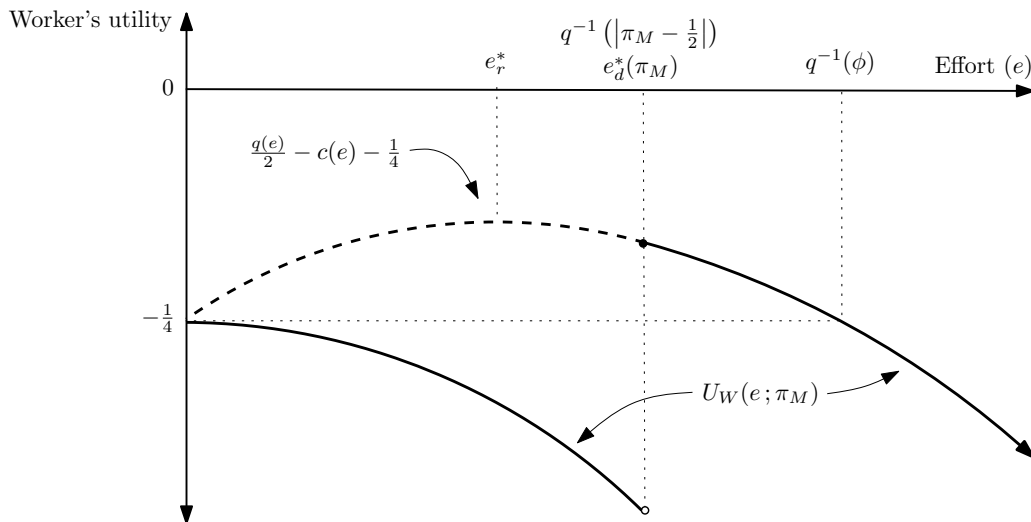
Figure 1.5: The appointment of a moderately skeptical manager (i.e., of one with $q(e_r^*) < \left|\pi_M - \frac{1}{2}\right| \leq \phi$) strengthens the worker's incentives. The worker, in this case, must exert more effort to influence the manager than he would have chosen to exert had the principal retained control. Given that the required level of effort to influence the manager is not prohibitively costly for the worker (i.e., $\left|\pi_M - \frac{1}{2}\right| \leq \phi$), the worker finds it optimal to exert just enough effort to influence the manager. Note that $e_d^*(\pi_M) > e_r^*$: the worker exerts more effort than he does when the principal retains control.

Implications and Discussion

Now I address the main question that arises in the simple case: how should the principal appoint the manager? Corollary 1.1 provides an answer.[32]

---

[32]Corollary 1.1 depends crucially upon the result of Proposition 1.2, which, as noted in Footnote 31, relies upon Assumption 1.2. Suppose that Assumption 1.2 is relaxed, though. In particular, suppose that an
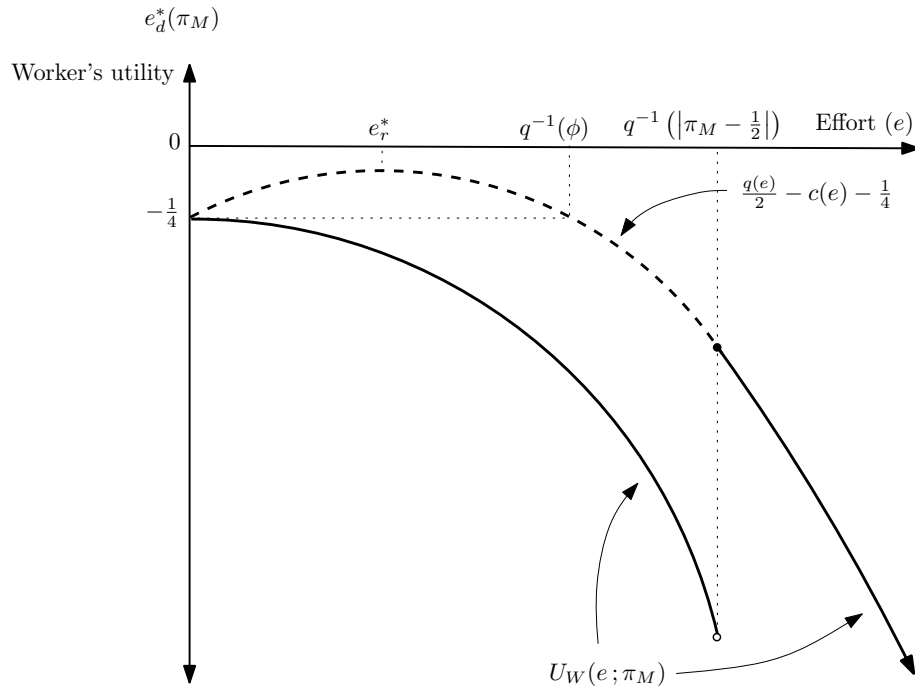
Figure 1.6: The appointment of a severely skeptical manager (i.e., of one with $\phi < \left|\pi_M - \frac{1}{2}\right| < \frac{1}{2}$) destroys the worker's incentives. In this case, the worker will have to undertake such a high level of effort to influence the manager that he prefers to send an uninformative signal and "play the odds" that favor the ex post optimality of the manager's ex ante preferred choice (i.e., $U_W(0\,;\pi_M) > U_W\left(q^{-1}\left(\left|\pi_M - \frac{1}{2}\right|\right)\,;\pi_M\right)$). Hence $e_d^*(\pi_M) = 0$: the worker exerts no effort.

Corollary 1.1. *In the simple case, when the principal can choose among candidate managers of different types, she prefers to appoint a manager of type $\pi_M \in \left[\frac{1}{2} - \phi, \frac{1}{2} - q(e_r^*)\right) \cup \left(\frac{1}{2} + q(e_r^*), \frac{1}{2} + \phi\right]$ than to retain decision making authority. Such a manager will rubberstamp the worker's signal, which would have informed the principal's choice, too, if she had retained control of the decision. In equilibrium, the principal appoints a manager of type $\pi_M \in \left\{\frac{1}{2} - \phi, \frac{1}{2} + \phi\right\}$.*

According to Corollary 1.1, the principal, ex ante, would rather delegate the decision to any moderately skeptical manager than retain decision making authority. Since the manager rubberstamps the worker's signal, a cursory analysis of the game may suggest that

---

indifferent manager chooses the action toward which she is biased, and that, if the worker is indifferent among effort levels, he chooses the lowest among the optimal effort levels. (In particular, when the worker faces a manager of skepticism ϕ, he exerts no effort.) In this case, the principal still can benefit from delegation by appointing a manager whose skepticism is $\phi - \varepsilon$, for some small $\varepsilon > 0$. The worker will prefer to acquire a signal of quality $\phi - \varepsilon + \delta$, where $0 < \delta < \varepsilon$, than to exert no effort. This signal is sufficiently precise to inform the manager's decision, and it is more precise than the signal that the worker would have acquired if the principal had retained control of the decision. Thus, the main insight of this section—that delegation to a biased manager benefits the principal—holds even if Assumption 1.2 is relaxed.
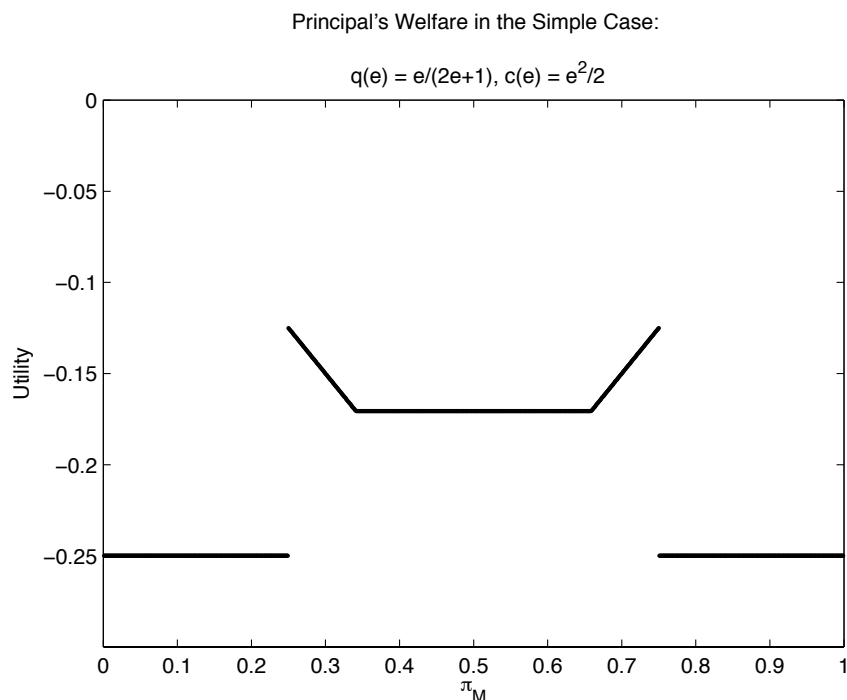
Principal's Welfare in the Simple Case:

$q(e) = e/(2e+1)$, $c(e) = e^2/2$

Figure 1.7: This graph illustrates the results of Proposition 1.2 and Corollary 1.1. Suppose that $q(e) \equiv e/(2e + 1)$ and $c(e) \equiv e^2/2$. Then $q(e_r^*) \approx 0.16$ and $\phi = 0.25$. The principal does best by appointing a manager of skepticism 0.25 (i.e., one with $\pi_M \in \{0.25, 0.75\}$). A manager that is any more skeptical will destroy the worker's incentives and bring the principal's welfare to its minimum.

the manager's strategy is to rubberstamp the signal *unconditionally*, which is exactly what the (unbiased) principal would have done if she had retained control. Thus the manager's position may appear redundant to a casual observer that does not recognize the effect of the manager's skepticism upon the amount of information that the worker acquires.

For delegation to be useful, it is crucial that the principal be unable to strip the manager of her decision making authority after the worker's signal is revealed. Clearly the principal has incentives to make the decision herself at that stage. However, if the principal could take such a step, then, in equilibrium, the worker would anticipate the principal's intervention and rationally would choose to exert e_r^* units of effort. Hence, to strengthen the worker's incentives to exert effort, the principal needs to credibly commit to abide by the manager's choice. Put another way, delegation must serve as an effective commitment device for the principal. An implication is that, off the equilibrium path, the principal must accept decisions that she finds suboptimal (based upon her posterior distribution). For example, in the case of the manufacturing firm, if the analyst were to exert too little effort relative to his equilibrium level, and the skeptical manager were to choose to build the plant at Bedford even upon observing a report that, in the proprietor's view, indicates that the plant ought to be build at Avon, the proprietor cannot overturn the manager's decision to build at Bedford.

While the view of delegation as commitment has been discussed previously in contexts such as monetary policy (Rogoff, 1985) and tax audit policy (Melumad and Mookherjee, 1989), Baker, Gibbons, and Murphy (1999) argue that formal authority in organizations cannot be delegated. In their view, delegation is informal: the principal always has the option to overturn decisions.[33] In light of this view, it might seem implausible that delegation alone provides the type of credible commitment that the principal needs in order to strengthen the worker's incentives. One way for the principal to achieve this commitment is to remain willfully ignorant of the signal. For example, she may require communication to be hierarchical, so that the worker cannot communicate directly with her.[34] This idea is related to an insight due to Crémer (1995), who observed that a principal can benefit from an increase in the cost of a signal that provides information regarding an agent's performance, because the higher cost can make it easier for the principal to commit to threats that she uses to provide incentives.

Corollary 1.1 also states that the principal finds it optimal to appoint a manager of skepticism level ϕ, the highest one in the moderate range. Such a manager will be the most skeptical one whose appointment will not impose too heavy an informational burden upon the worker. When reporting to such a manager, the worker is indifferent between acquiring an uninformative signal and acquiring a signal of sufficiently high quality to inform the manager's decision. By Assumption 1.2, the worker chooses the latter option, which leads the manager to rubberstamp his report. The imposition of a marginally heavier informational burden upon the worker (through the appointment of a slightly more skeptical manager) will destroy the worker's incentives to acquire information.

In the context of the manufacturing firm example, if the proprietor appoints a manager whose prior beliefs are moderately in favor of building either at Avon or at Bedford, she can expect a better outcome than if she retains control of the decision. She does best by appointing a manager whose skepticism level (regarding either building at Avon or building at Bedford) matches the analyst's tolerance for skepticism. When such a manager is entrusted with the decision, the analyst is indifferent between exerting no effort and exerting just enough effort to produce a report that will inform the manager's decision. By Assumption 1.2, the analyst chooses to produce an informative report, which the manager proceeds to rubberstamp. The optimal manager challenges the analyst enough for the organization to achieve maximal benefit from the analyst's skills, but not so much that the analyst gives up entirely.

---

[33]Baker, Gibbons, and Murphy show that delegation can be supported by the principal's reputational concerns. In particular, a principal can use delegation to strengthen an agent's incentives, but only if the agent does not believe that the principal will overturn his decision. Thus, in a setting with repeated interactions, the principal may find it optimal to not intervene (even when intervention is optimal in the stage game) so as to avoid developing a reputation for intervention, which will destroy the future expected benefits of delegation. Of course, in this static setting, there is no scope for reputational considerations.

[34]Friebel and Raith (2004) and Ambrus, Azevedo, and Kamada (2013) observe that communication within many organizations (e.g., the U.S. Armed Forces) follows a hierarchical "chain of command," which is enforced by explicit rules (e.g., open door policies), established norms (e.g., organizational culture), or other features (e.g., information technology or the locations of senior managers' offices).

Finally, as Corollary 1.1 indicates, rubberstamping is not necessarily indicative of the manager's ex ante indifference. On the contrary, a manager that is appointed in equilibrium will be quite opinionated. However, she will rubberstamp the worker's report based upon its high quality.

The Principal's Choice of Worker

As a prelude to the analysis of the general case, suppose that the principal, in addition to choosing a manager, also can choose among workers, all of whom have $\pi_W = \lambda_W = \frac{1}{2}$ but differ in $q(\cdot)$ and $c(\cdot)$. In the example of the manufacturing firm, the proprietor can appoint any member of a pool of financial analysts. All of these financial analysts are unbiased, but they differ in their returns to effort and in their costs of effort. These differences may reflect variations in skill level, years of experience, familiarity with the problem, opportunity costs, or other factors.

A natural question to ask is: what are the principal's preferences over workers from this pool? To answer this question, begin by considering two workers. From Corollary 1.1 and Proposition 1.2, the principal prefers to appoint the worker with the higher tolerance for skepticism. The reason is that, by appointing this worker, the principal allows herself the option to appoint a more skeptical manager—and thus induce the worker to obtain better information—than if she had chosen to appoint the worker with the lower tolerance for skepticism. This argument establishes Corollary 1.2:

Corollary 1.2. *Consider a pool of workers that have $\pi_W = \lambda_W = \frac{1}{2}$ but vary in their returns to effort and costs of effort (i.e., in the attributes $q(\cdot)$ and $c(\cdot)$). The unbiased principal's preference ranking (from most preferred to least preferred) among these workers coincides with the workers' ordering (from highest to lowest) by tolerance for skepticism.*

Among workers that have $\pi_W = \lambda_W = \frac{1}{2}$, a worker's tolerance for skepticism is an indicator of his efficiency in acquiring information, since it provides an upper bound on the quality of information that he is willing to acquire in order to influence the manager. Clearly the worker's efficiency is related to both $q(\cdot)$ and $c(\cdot)$: a highly efficient worker will have steep returns to effort and flat costs of effort. Corollary 1.3 provides a partial characterization of the principal's preferences over workers based on $q(\cdot)$ and $c(\cdot)$.

Corollary 1.3. *Consider a pool of workers that have $\pi_W = \lambda_W = \frac{1}{2}$ but vary in their returns to effort and costs of effort (i.e., in the attributes $q(\cdot)$ and $c(\cdot)$). Define the following partial order among the workers:*

$$W_1 \succeq W_2 \overset{\Delta}{\Longleftrightarrow} \left(c_1 \circ q_1^{-1}\right)''(\cdot) \leq \left(c_2 \circ q_2^{-1}\right)''(\cdot),$$

where W_1 and W_2 are two workers from the pool with attributes of $q_1(\cdot)$ and $c_1(\cdot)$ (for W_1) and $q_2(\cdot)$ and $c_2(\cdot)$ (for W_2), respectively. If members of a subset of the pool can be ranked according to $\succeq$, then the unbiased principal's preference ranking, restricted to that subset, coincides with $\succeq$.

According to Corollary 1.3, the principal prefers workers for whom $(c \circ q^{-1})(\cdot)$ is relatively less convex. This property can be interpreted as a preference for workers that can acquire information relatively efficiently, since workers whose $(c \circ q^{-1})(\cdot)$ functions exhibit low convexity will have less convex $c(\cdot)$ functions (i.e., low costs of effort) or more convex $q(\cdot)$ functions (i.e., high returns to effort). Low costs of effort and high returns to effort are consistent with efficient information acquisition, and the principal clearly should favor more efficient workers: she can induce them to acquire better information than less efficient workers are willing to acquire. The result is illustrated in Figure 1.8. For example, consider
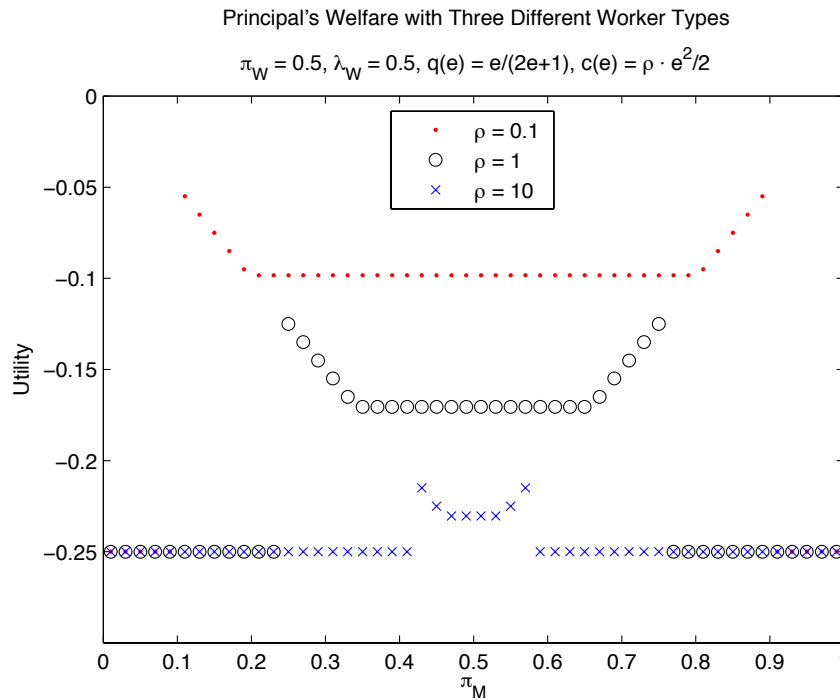
Figure 1.8: This graph illustrates the result of Corollary 1.3. It compares the principal's expected utility, as a function of the appointed manager's type, across three candidate worker types. All three candidate workers share the attributes $\pi_W \equiv 1/2$, $\lambda_W \equiv 1/2$, and $q(e) \equiv e/(2e+1)$. However, they vary in their cost functions through their values for a parameter ρ, where $c(e) \equiv \rho e^2/2$. Higher values of ρ are associated with higher levels of convexity of the function $(c \circ q^{-1})(\cdot)$. As the graph indicates, the principal does better with a worker that has a lower value of ρ for two reasons. First, a less convex $(c \circ q^{-1})(\cdot)$ is associated with a higher tolerance for skepticism, so the associated worker can be induced to acquire more information through the appointment of a more skeptical manager. Second, even for a fixed manager type, a worker with a lower value of ρ will acquire at least as much information as one with a higher value of ρ, and strictly more information if the manager's skepticism does not exceed the former worker's tolerance for skepticism. This relationship arises in part from the fact that e_r^* is higher for the worker with the lower value of ρ, since his cost of acquiring information is lower than the other worker's.

two financial analysts that the proprietor of the manufacturing firm may appoint. Suppose that the analysts are identical in their returns to effort, but that one has a more rapidly rising opportunity cost than the other (e.g., because of other personal commitments). It is intuitively clear that the analyst with the steep cost will have a lower tolerance for skepticism, since a given level of effort yields him the same benefit as it does the other worker, but at a higher cost. Thus the proprietor will be able to appoint a more skeptical manager—and, in doing so, achieve a better expected outcome—if she appoints the analyst with the lower opportunity cost. Similarly, if two analysts have identical costs of effort but vary in their returns to effort (e.g., because of differences in skill), the more productive analyst will have a higher tolerance for skepticism, and the proprietor will prefer to appoint that analyst. In either case, the proprietor prefers to appoint the analyst that is more efficient in acquiring information, because she can induce this analyst to produce a more accurate forecast.

The General Case

Now I allow the principal to have an arbitrary type, and I assume that she has access to a pool of workers. Suppose that the principal's type is $(\pi, \lambda) \in (0, 1)^2$, and that she can choose among workers, including ones whose preferences are not aligned with hers. Recall that candidate workers can be identified by four attributes: π_W (prior beliefs), λ_W (minimal belief on A for α to be preferred), $q(\cdot)$ (returns to effort), and $c(\cdot)$ (cost of effort). For brevity, whenever possible, I use t_W to denote a quadruple of the form $(\pi_W, \lambda_W, q(\cdot), c(\cdot)) \in \mathcal{T}_W$. Unless noted otherwise, I make no assumptions on $\mathcal{T}_W$ other than the ones that $\pi_W, \lambda_W \in (0, 1)$, and that $q(\cdot)$ and $c(\cdot)$ satisfy the properties that were discussed previously. Similarly, I let $\mathcal{T}_M$, the type space of candidate managers, be an arbitrary subset of $(0, 1)$ that contains $\frac{\pi\lambda - \pi}{2\pi\lambda - \pi - \lambda}$ (again, so that the principal can retain control of the decision), and I make no additional assumptions on its structure unless noted otherwise.

Worker Attributes

For stating and interpreting the results, it is helpful to define two attributes in terms of the basic parameters that define a worker's type. These attributes, respectively, indicate the strength of the worker's incentives to acquire information and the severity of the worker's ex ante bias toward either action.

Definition 1.9 (Valuation of information)**.** Consider a worker of type

$$t_W = (\pi_W, \lambda_W, q(\cdot), c(\cdot)).$$

The worker's *valuation of information* is given by $\xi(t_W) \equiv \lambda_W + \pi_W - 2\pi_W\lambda_W \in (0, 1)$.

According to Definition 1.9, a worker's valuation of information is his expected loss conditional on an error (i.e., the manager's choice either of α when the state is B or of β

when the state is A) when the manager rubberstamps his signal.[35] Thus, assuming that the worker exerts enough effort to influence the manager, his private marginal benefit of effort is increasing in his valuation of information (which is independent of his effort level). Figure 1.9 depicts $\xi(\cdot)$ as a function of π_W and λ_W. Going back to the manufacturing firm example,

Figure 1.9: The curves in this figure show the isoquants of $\xi(\cdot)$ in (π_W, λ_W) space. The arrows depict the function's gradient. Thus, each arrow points in the direction of the function's steepest increase from the point of the arrow's origin. Note that a worker with a high valuation of information either believes that the state is likely to be A and suffers a severe loss from the (β, A) error (in which case he has especially strong incentives to induce the manager to choose α) or believes that the state is likely to be B and suffers a severe loss from the (α, B) error (in which case he has especially strong incentives to induce the manager to choose β).

an financial analyst with an extremely high valuation of information might believe strongly that a plant at Avon is likely to be more profitable than one at Bedford and, furthermore, that building the plant at Bedford when it is suboptimal to do so will exact such heavy losses on the firm that his employment would be jeopardized (e.g., because he would be fired as punishment or because the financial losses would force the proprietor to lay off workers).

The second attribute that is helpful to define is the worker's α-tilt (or just tilt), which indicates the strength of his ex ante bias toward α:

---

[35]The rationale behind the name of this attribute can be understood clearly from (1.4), the worker's objective function.

Definition 1.10 (Tilt). Consider a worker of type $t_W = (\pi_W, \lambda_W, q(\cdot), c(\cdot))$. The worker's *tilt* is given by $\tau(t_W) \equiv \pi_W - \lambda_W \in (-1, 1)$.

When $\tau(t_W)$ is close to 1, the worker has a high value of π_W and a low value of λ_W. The first condition signifies a high level of confidence that α is optimal. The second condition is complementary: it represents a high level of belief-independent apprehension regarding the action β. In the example, a financial analyst with a tilt close to 1 might believe that the plant at Avon is much more likely to yield higher profits than the one at Bedford and, furthermore, that a plant mistakenly built at Bedford will be disastrous for the firm.

From this discussion, it may seem that tilt plays the same role as valuation of information in determining a worker's behavior. To see that the role of tilt is distinct, compare Figures 1.9 and 1.10. Although workers with extremely high valuations of information tend to have
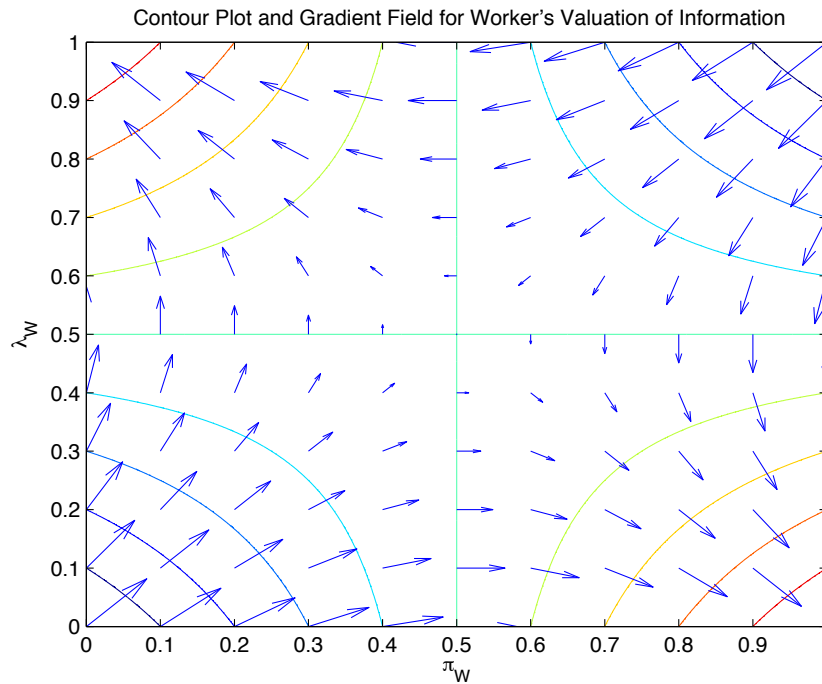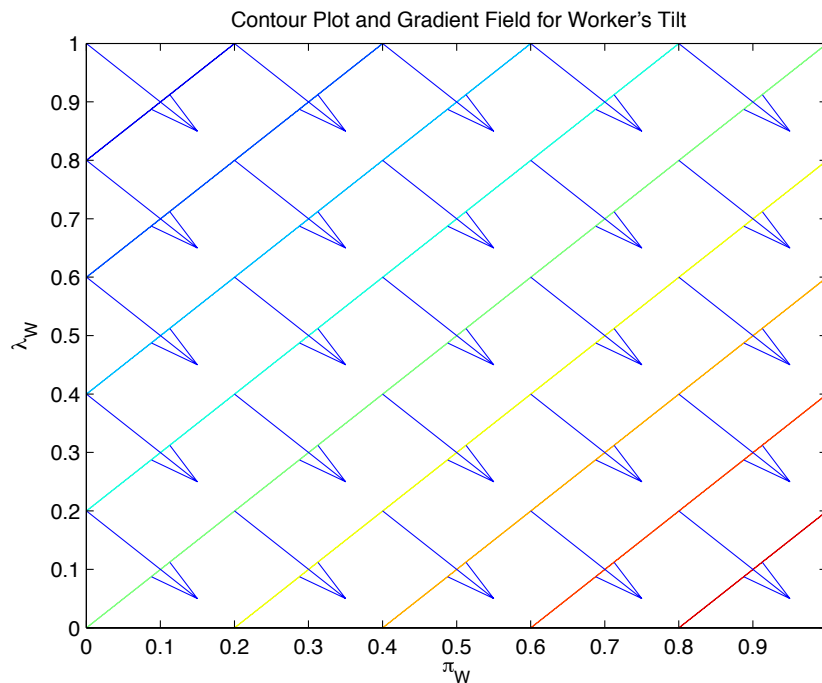

Figure 1.10: The curves in this figure show the isoquants of $\tau(\cdot)$ in (π_W, λ_W) space. The arrows depict the function's gradient. Thus, each arrow points in the direction of the function's steepest increase from the point of the arrow's origin. A comparison with Figure 1.9 shows that valuation of information and tilt are related but distinct concepts.

very high magnitudes of tilt (captured by $|\tau(t_E)|$), the graphs illustrate that the concepts are distinct. For example, an unbiased worker—whose tilt is 0 by definition—may have a valuation of information that is close to 0: consider the case of a worker with $\pi_M = \lambda_M \approx 0$. Such a worker is ex ante indifferent between α and β. This worker believes that the state is B with very high probability. Furthermore, he suffers relatively little from the (α, B)

error, so $u_W(\alpha, B) \approx 0 = u_W(\beta, B)$. Thus, this worker believes that the state of the world in which he is nearly indifferent is very likely to occur, so he has very weak incentives to acquire information. On the other hand, an unbiased worker also can have $\pi_M = \lambda_M = \frac{1}{2}$, which implies a valuation of information of $\frac{1}{2}$. This worker, in contrast to the previous one, is not approximately indifferent regarding the decision in either state of the world, so his incentives to acquire information are stronger. Hence, an unbiased worker (who is ex ante indifferent regarding the decision) could, but need not, have a low valuation of information. A worker with a low valuation of information believes that, with high probability, he will be nearly indifferent regarding the decision ex post.

The discussion above highlights an interpretation of valuation of information as the degree to which a worker expects to be affected, ex post, by the decision. This notion is related to, but distinct from, the magnitude of tilt, which indicates an ex ante bias toward either decision. It also should be clear now that, unlike the manager type space, the worker type space cannot be reduced in dimensionality without a loss of generality.

Using Assumption 1.2 and Definitions 1.9 and 1.10, the expert's expected utility as a function of his effort level can be written as

$$U_W(e\,;\pi_M, t_W) \equiv \begin{cases} -c(e) - \frac{\xi(t_W)}{2} - \frac{\tau(t_W)}{2} & \text{if } e < q^{-1}(|\pi_M - \frac{1}{2}|) \text{ and } \pi_M < \frac{1}{2}, \\ -c(e) - \frac{\xi(t_W)}{2} + \frac{\tau(t_W)}{2} & \text{if } e < q^{-1}(|\pi_M - \frac{1}{2}|) \text{ and } \pi_M > \frac{1}{2}, \quad (1.4) \\ -c(e) - \frac{\xi(t_W)}{2} + \xi(t_W)q(e) & \text{otherwise,} \end{cases}$$

where $t_W = (\pi_W, \lambda_W, q(\cdot), c(\cdot))$. In interpreting (1.4), it is crucial to keep in mind that $q(\cdot)$ and $c(\cdot)$ are part of the worker's type, t_W, and may vary across candidate workers. (1.4) allows the characterization of the worker's effort level and the principal's welfare.

Characterization Results

The results of this section generalize the ones from the simple case of Section 1.4 (in which $\xi(t_W) = \frac{1}{2}$ and $\tau(t_W) = 0$).

Lemma 1.3. *Let $t_W \equiv (\pi_W, \lambda_W, q(\cdot), c(\cdot))$ be a worker type. The function $e \mapsto -c(e) - \frac{\xi(t_W)}{2} + \xi(t_W)q(e)$, defined on $[0, \infty)$, has a unique maximizer, denoted by $\hat{e}(t_W)$, that is positive and is characterized by the condition $c'(\hat{e}(t_W)) = \xi(t_W)q'(\hat{e}(t_W))$.*

In a sense, $\hat{e}(t_W)$ is an analogue of e_r^* from the simple case: $\hat{e}(t_W)$ is the effort level that a worker of type t_W would exert if the manager were to adopt a policy of rubberstamping the worker's signal. Due to this interpretation, I will refer to $\hat{e}(t_W)$ as the "rubberstamping effort level" for a worker of type t_W.

Just as in the simple case, the principal can appoint a moderately skeptical manager to provide stronger incentives to the worker, where the term "moderately skeptical" is defined with respect to a worker-specific threshold that represents the maximal level of skepticism that does not destroy the worker's incentives to exert effort. In the simple case, in which there was only a single worker (who was unbiased), the maximal level of skepticism was given

by the attribute ϕ. An analogue of this attribute exists in this setting, too, but obviously it is a function of the worker's type:

Definition 1.11 (Tolerance for skepticism)**.** Fix a worker type t_W.

(i) Let $\eta \in \{\alpha, \beta\}$ be given. $\phi_\eta(t_W)$ is the worker's *tolerance for skepticism regarding η*, the maximal level of skepticism regarding η that an appointed manager may hold in order for the worker to exert a positive level of effort.

(ii) $\phi(t_W) = \max \{\phi_\alpha(t_W), \phi_\beta(t_W)\}$ is the worker's *tolerance for skepticism*, the maximal level of skepticism that an appointed manager may hold in order for the worker to (possibly) exert a positive level of effort.[36]

The technical definitions of $\phi_\alpha(\cdot)$ and $\phi_\beta(\cdot)$ are provided in Definition A.2 in Appendix A.2. However, Definition 1.11 is sufficient to understand the roles of these attributes, which arise in Proposition 1.3, the main characterization result of the general case. Before stating that result, I state a useful fact:

Lemma 1.4. *Consider a worker of type $t_W \equiv (\pi_W, \lambda_W, q(\cdot), c(\cdot)) \in \mathcal{T}_W$, and suppose that $\phi_\alpha(t_W) = 0$. Then the worker is biased toward β and prefers the outcome in which he shirks and β is implemented unconditionally to any outcome in which he informs the decision. Similarly, if $\phi_\beta(t_W) = 0$, the worker is biased toward α and prefers the outcome in which he shirks and α is implemented unconditionally to any outcome in which he informs the decision.*

Given an action $\eta \in \{\alpha, \beta\}$, a worker with $\phi_\eta(t_W) = 0$ has no tolerance for skepticism regarding η. If he knows that the manager is skeptical regarding η, he shirks and allows the manager to choose the opposite action, which both he and the manager favor ex ante. The formal proof of the result is in Appendix A.1, but it will be helpful here to provide some more intuition to illustrate which types of workers satisfy such a condition. Suppose that $\phi_\beta(t_W) = 0$ and $\pi_M > \frac{1}{2}$. The proof of Lemma 1.4 shows that

$$\tau(t_W) > 2\left[\xi(t_W)q(\hat{e}(t_W)) - c(\hat{e}(t_W))\right], \tag{1.5}$$

from which it follows that

$$\arg\max_{e \geq 0} U_W(e\,;\pi_M, t_W) = \{0\}.$$

That is, the worker finds it optimal to shirk if α will be chosen when he shirks. Such a worker essentially feels coddled by a like-minded manager. See Figure 1.11 for a graphical illustration of the situation. The condition of (1.5) may result from various combinations of the factors that I outline below:

---

[36]For example, if $\phi_\alpha(t_W) = 0.2$ and $\phi_\beta(t_W) = 0.4$, then the worker shirks if the manager's type, π_M, satisfies either $\pi_M < 0.3$ or $\pi_M > 0.9$. Note that any manager with $\left|\pi_M - \frac{1}{2}\right| > 0.4 = \phi(t_W)$ will destroy the worker's incentives, but only some managers with $\left|\pi_M - \frac{1}{2}\right| \leq 0.4$ will destroy the worker's incentives. For example, a manager with $\pi_M = 0.25$ will destroy the worker's incentives, but one with $\pi_M = 0.75$ will not. Both of these managers have skepticism 0.25.
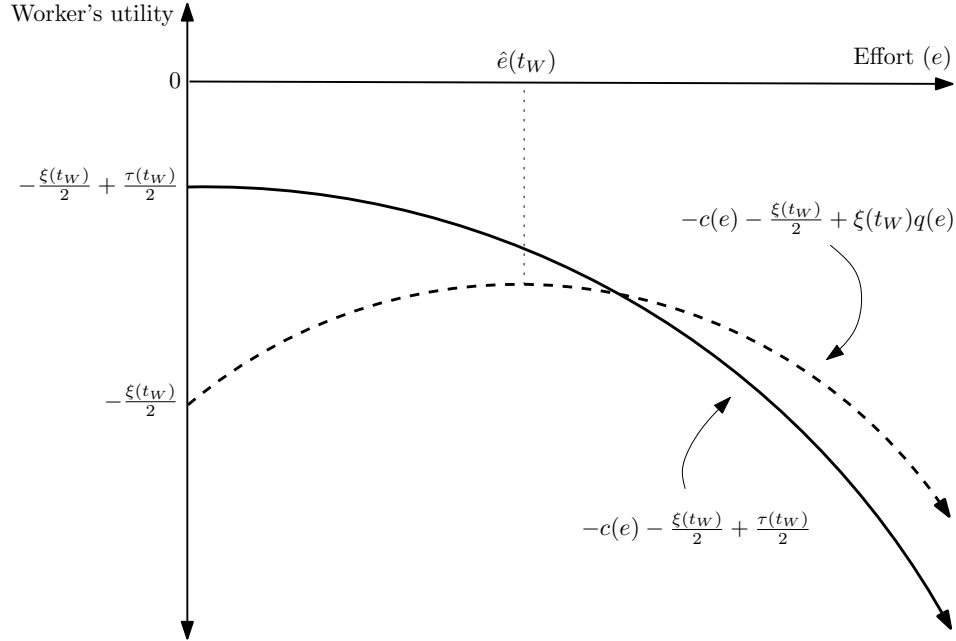
Figure 1.11: When the worker's type satisfies (1.5), he prefers the outcome in which he shirks and α is implemented to any outcome in which his signal informs the decision. Thus, if the manager is biased in favor of α, the worker finds it optimal to shirk.

$\tau(t_W)$ **is large.** In words, the worker's tilt toward α is severe: he is confident that α is optimal (because π_W is high) and, regardless of his prior beliefs, finds the action β very risky (because λ_W is low). Given that the appointed manager, who is biased toward α, does not need to be convinced to choose α, this worker expects to gain little from acquiring information.

$\xi(t_W)$ **is small.** That is, the worker's valuation of information is low; his incentives to acquire information are weak.

$q(\hat{e}(t_W))$ **is small compared to** $c(\hat{e}(t_W))$**.** In this case, either $q'(\cdot)$ diminishes rapidly or $c'(\cdot)$ grows rapidly (or both). That is, $q(\cdot)$ is very concave or $c(\cdot)$ is very convex. In either case, such a worker is not efficient in acquiring information.

Now I state the main characterization result.

Proposition 1.3. *Suppose that the principal's type is* $(\pi, \lambda) \in (0,1)^2$*, and that she appoints a manager of type* $\pi_M \in \mathcal{T}_M$ *and a worker of type* $t_W \equiv (\pi_W, \lambda_W, q(\cdot), c(\cdot)) \in \mathcal{T}_W$*. The principal's appointments can produce any of three types of outcomes, which are determined by a mutually exclusive and exhaustive set of conditions:*

Skepticism does not affect worker's incentives. *Suppose that all of the following conditions hold:*

- $0 \leq \left| \pi_M - \frac{1}{2} \right| \leq q\left(\hat{e}(t_W) \right)$;
- $\pi_M - \frac{1}{2} \leq \phi_\beta(t_W)$;
- $\frac{1}{2} - \pi_M \leq \phi_\alpha(t_W)$.

Then the manager rubberstamps the worker's signal. The worker's effort level and the principal's expected utility, respectively, are

$$e_d^*(\pi_M ; t_W) = \hat{e}(t_W),$$

$$U_d^*(t_W, \pi_M ; \pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q\left(\hat{e}(t_W) \right) \right).$$

Skepticism strengthens worker's incentives. *Suppose that either of the following conditions holds:*

- $q(\hat{e}(t_W)) < \pi_M - \frac{1}{2} \leq \phi_\beta(t_W)$;
- $q(\hat{e}(t_W)) < \frac{1}{2} - \pi_M \leq \phi_\alpha(t_W)$.

Then the manager rubberstamps the worker's signal, and

$$e_d^*(\pi_M ; t_W) = q^{-1}\left(\left| \pi_M - \frac{1}{2} \right| \right),$$

$$U_d^*(t_W, \pi_M ; \pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \left| \pi_M - \frac{1}{2} \right| \right).$$

Skepticism destroys worker's incentives. *Suppose that either of the following conditions holds:*

- $\pi_M - \frac{1}{2} > \phi_\beta(t_W)$;
- $\frac{1}{2} - \pi_M > \phi_\alpha(t_W)$.

Then the manager chooses her preferred action, and

$$e_d^*(\pi_M ; t_W) = 0,$$

$$U_d^*(t_W, \pi_M ; \pi, \lambda) = \begin{cases} \pi\lambda - \lambda & \text{if } \pi_M > \frac{1}{2}, \\ \pi\lambda - \pi & \text{if } \pi_M < \frac{1}{2}. \end{cases}$$

To interpret Proposition 1.3, suppose that the manager's type is $\pi_M \geq \frac{1}{2}$, so that she is not biased in favor of β (or, equivalently, is not skeptical regarding α).[37] Consider, individually, the outcomes that are outlined in the statement of the result:

---

[37]Due to the symmetry of the environment, the interpretation of Proposition 1.3 for cases in which the manager is biased in favor of β is analogous to the one presented here.

Skepticism does not affect worker's incentives. The result states that, if the manager's skepticism regarding β exceeds neither the quality of the signal that the worker produces when facing an unbiased manager nor the worker's tolerance for skepticism regarding β, then the worker's effort level does not change from the case in which he faces an unbiased manager. The reason is that the same amount of effort remains sufficient to influence this (mildly) biased manager.

The condition $\pi_M - \frac{1}{2} \leq \phi_\beta(t_W)$ reflects an important caveat that did not arise in the simple case. In the simple case, the worker was unbiased; thus his tolerance for skepticism (regarding either action) was positive. The worker in the general case need not be unbiased and, in particular, may have no tolerance for skepticism regarding the action against which he is biased. As noted in the discussion following the statement of Lemma 1.4, in this case, the worker's incentives to exert effort vanish if he faces a manager that prefers the same action as he does.

Note that, if the manager is unbiased (i.e., if $\pi_M = \frac{1}{2}$), the three conditions of this case hold. The following two cases, therefore, deal exclusively with biased managers, who have positive skepticism.

Skepticism strengthens worker's incentives. In this case, the manager's skepticism regarding β is moderate: it is strong enough to induce the worker to exert additional effort to influence the manager, but it is not severe enough to destroy the worker's incentives entirely. The intuition is the same as in the simple case. (Note, also, that the case $\phi_\beta(t_W) = 0$ is ruled out explicitly, since $q(\hat{e}(t_W)) > 0$ by the result of Lemma 1.3.)

Skepticism destroys worker's incentives. If $\phi_\beta(t_W) > 0$, the intuition is the same as in the simple case: to influence such a biased manager, the worker needs to exert an excessively high amount of effort. He prefers to shirk and "play the odds" that the manager's preferred action turns out to be optimal. There is one additional case that did not arise with an unbiased worker, though: it may be the case that $\phi_\beta(t_W) = 0$. Then, even if the manager's skepticism regarding β is very slight, the worker finds it more appealing to shirk and let the manager implement α than to exert any effort in influencing the manager.

The principal, of course, can retain control of the decision. In the simple case, the principal was unbiased and thus favored rubberstamping any signal. In this case, though, she generally requires a certain threshold level of signal informativeness to be convinced to abandon her ex ante preferred action.

Lemma 1.5. *For any principal type $(\pi, \lambda) \in (0,1)^2$, given $q \in \left(\frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda}, \frac{1}{2} \right)$,[38] the principal prefers the decision to be informed by a signal of quality q than for either action to be*

---

[38]To see that $\frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda} < \frac{1}{2}$, observe that

$$2\pi + 2\lambda - 4\pi\lambda = 2\pi \cdot (1 - \lambda) + 2\lambda \cdot (1 - \pi) > 0,$$

implemented unconditionally. On the other hand, if $q \in \left[0, \frac{|\pi-\lambda|}{2\pi+2\lambda-4\pi\lambda}\right)$, the principal prefers the unconditional implementation of her ex ante preferred action.

Note that $\frac{|\pi-\lambda|}{2\pi+2\lambda-4\pi\lambda}$ is similar to the manager's skepticism, as it provides a lower bound for the set of quality levels of signals that inform the principal's decision.[39] This observation motivates the following definition.

Definition 1.12 (Informational standard)**.** Consider a principal of type (π, λ). The principal's *informational standard*, given by $\frac{|\pi-\lambda|}{2\pi+2\lambda-4\pi\lambda}$, is the minimal quality of signal that she prefers to rubberstamp rather than to unconditionally implement her preferred action.

Corollary 1.4 is an analogue of Proposition 1.3 that characterizes the outcomes that may arise when the principal retains control of the decision.

Corollary 1.4. *Suppose that the principal's type is $(\pi, \lambda) \in (0,1)^2$, and that she appoints a worker of type $t_W \equiv (\pi_W, \lambda_W, q(\cdot), c(\cdot)) \in \mathcal{T}_W$ but retains decision making authority. The principal's appointment of the worker can produce any of three types of outcomes, which are determined by a mutually exclusive and exhaustive set of conditions:*

Informational standard does not affect worker's incentives. *Suppose that all of the following conditions hold:*

- $0 \leq \frac{|\pi-\lambda|}{2\pi+2\lambda-4\pi\lambda} \leq q\left(\hat{e}(t_W)\right);$

- $\frac{\pi-\lambda}{2\pi+2\lambda-4\pi\lambda} \leq \phi_\beta(t_W);$

- $\frac{\lambda-\pi}{2\pi+2\lambda-4\pi\lambda} \leq \phi_\alpha(t_W).$

Then the principal rubberstamps the worker's signal. The worker's effort level and the principal's expected utility, respectively, are

$$e_r^*(\pi, \lambda \,; t_W) = \hat{e}(t_W),$$

$$U_r^*(t_W \,; \pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q\left(\hat{e}(t_W)\right)\right).$$

Informational standard strengthens worker's incentives. *Suppose that either of the following conditions holds:*

---

and so

$$\frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda} < \frac{\pi + \lambda \cdot (1 - 2\pi)}{2\pi + 2\lambda - 4\pi\lambda}$$
$$= \frac{1}{2}.$$

Similarly, $\frac{\lambda-\pi}{2\pi+2\lambda-4\pi\lambda} < \frac{1}{2}$.

[39]It is worth noting that the skepticism of a manager of type $\frac{\pi\lambda-\pi}{2\pi\lambda-\pi-\lambda}$ (whose preferences are aligned with those of a principal of type (π, λ)) is $\frac{|\pi-\lambda|}{2\pi+2\lambda-4\pi\lambda}$.

- $q(\hat{e}(t_W)) < \frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda} \leq \phi_\beta(t_W)$;
- $q(\hat{e}(t_W)) < \frac{\lambda - \pi}{2\pi + 2\lambda - 4\pi\lambda} \leq \phi_\alpha(t_W)$.

Then the principal rubberstamps the worker's signal, and

$$e_r^*(\pi, \lambda \, ; t_W) = q^{-1}\left(\frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda} \right),$$

$$U_r^*(t_W \, ; \pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda} \right).$$

Informational standard destroys worker's incentives. *Suppose that either of the following conditions holds:*

- $\frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda} > \phi_\beta(t_W)$;
- $\frac{\lambda - \pi}{2\pi + 2\lambda - 4\pi\lambda} > \phi_\alpha(t_W)$.

Then the principal chooses her preferred action, and

$$e_r^*(\pi, \lambda \, ; t_W) = 0,$$

$$U_r^*(t_W \, ; \pi, \lambda) = \begin{cases} \pi\lambda - \lambda & \text{if } \pi > \lambda, \\ \pi\lambda - \pi & \text{if } \pi < \lambda. \end{cases}$$

Illustrative Examples

Proposition 1.3 describes the principal's welfare as a function of the types of the appointed worker and manager. Although the result is an obvious generalization of Proposition 1.2, the relationships between the agents' types and the principal's welfare are not so straightforward in the general case as they are in the simple case. In this section, I present and interpret several examples that highlight various considerations that underlie the principal's welfare maximization problem, which is analyzed in Section 1.4.

One key difference is that, in the general case, the principal need not be unbiased. Thus the symmetry in her payoffs that arises in the simple case—and which is evident in Figures 1.7 and 1.8—need not appear in the general case. In particular, if the principal makes an appointment that she expects to lead to the implementation of α, her payoff might not be the same as under an appointment that she expects to lead to the implementation of β. Furthermore, it might be the case that the principal prefers the unconditional implementation of her ex ante preferred action to any outcome in which the manager rubberstamps the worker's signal. Figures 1.12, 1.13, and 1.14 illustrate these points. They fix a worker type and show how the principal's welfare depends upon the manager's type.

The examples depicted by Figures 1.15, 1.16, and 1.17, in some sense, conduct an opposite exercise. In these figures, a manager type is fixed, and the (unbiased) principal's welfare is

plotted as a function of λ_W, while holding the worker's other attributes fixed.[40] As the figures show, the relationship between λ_W and the principal's welfare is non-monotonic, due to the non-monotonicity of the relationship between λ_W and $\xi(\pi_W, \lambda_W, q(\cdot), c(\cdot))$.
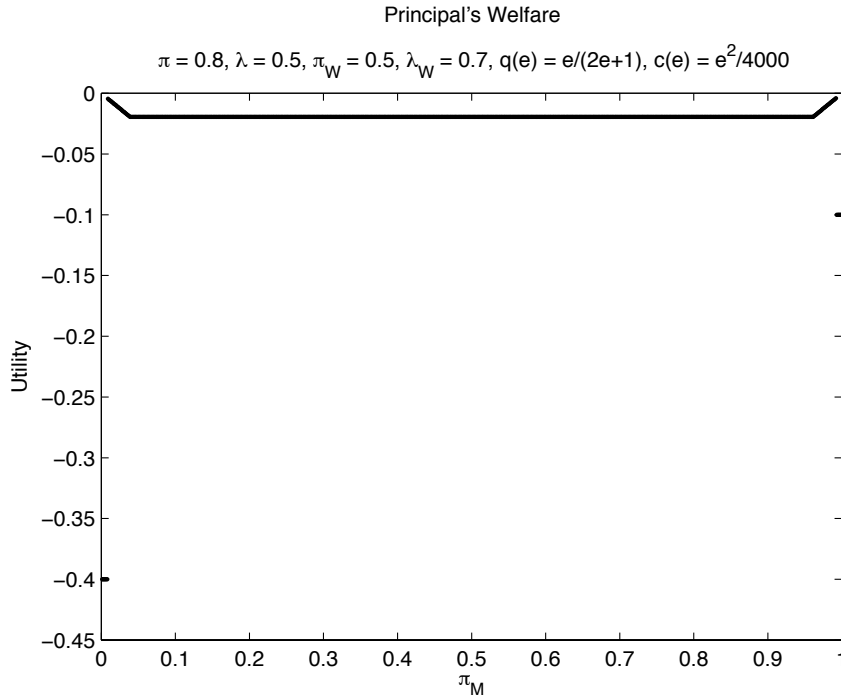
Figure 1.12: Let $t_W = (0.5, 0.7, q(\cdot), c(\cdot))$, where $q(e) = e/(2e + 1)$ and $c(e) = e^2/4000$. Then $0.49 < \phi_\alpha(t_W) < \phi_\beta(t_W) < 0.5$, and $q(\hat{e}(t_W)) \approx 0.46$. The worker in this situation is biased in favor of β but has positive tolerance for skepticism regarding α. Due to the relatively low convexity of the worker's cost function, both of his tolerances for skepticism are very close to $\frac{1}{2}$, and his rubberstamping effort level also is very high. Therefore it takes an extremely skeptical manager (i.e., one with $\pi_M < 0.04$ or $\pi_M > 0.96$) to strengthen his incentives, and an even more skeptical one (i.e., one with $\pi_M < 0.01$ or $\pi_M > 0.99$) to destroy them. Note that, since the principal is biased in favor of α, she prefers the unconditional implementation of α (which gives her an expected payoff of -0.1) to that of β (which gives her an expected payoff of -0.4).

The Principal's Problem

The examples of Section 1.4 demonstrate that the principal's welfare can change with the types of the appointed manager and worker in surprising ways. This section considers the

---
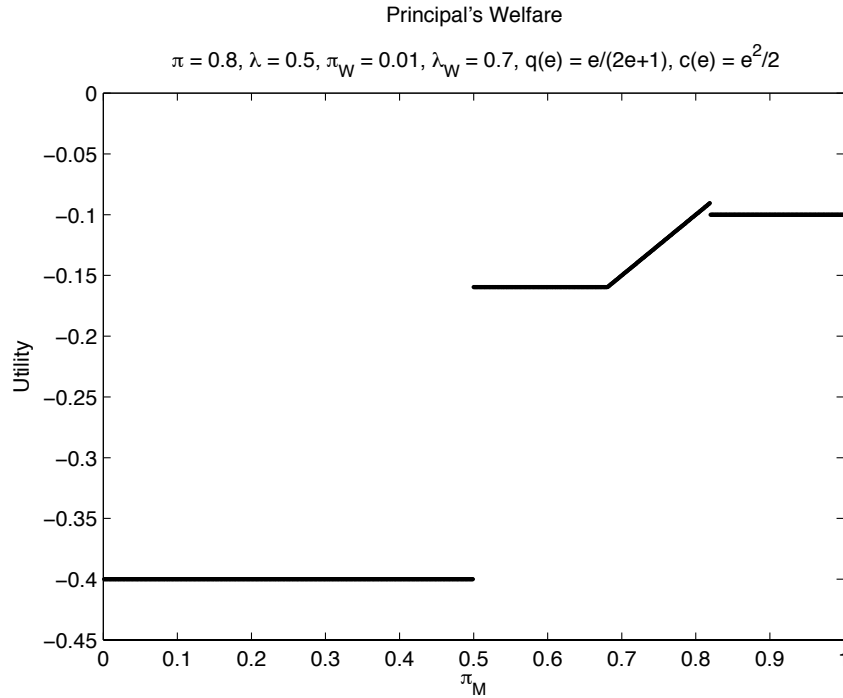
Figure 1.13: This worker's cost function, $c(e) = e^2/2$, is much more convex than that of the worker of Figure 1.12. As a result, this worker's rubberstamping effort level is lower. This worker is strongly biased in favor of β and has zero tolerance for skepticism regarding α. Thus, whenever a manager of type $\pi_M < \frac{1}{2}$ is appointed, the worker shirks, and the manager chooses β. By appointing a manager that is sufficiently skeptical regarding β, the principal can achieve an even higher expected payoff than she achieves from the unconditional implementation of α. However, once the manager's skepticism regarding β exceeds $\phi_\beta(t_W) \approx 0.32$ (i.e., once $\pi_M > 0.82$), the worker's incentives are destroyed.

principal's welfare maximization problem:

$$\max_{(t_W, \pi_M) \in \mathcal{T}_W \times \mathcal{T}_M} U_d^*(t_W, \pi_M \, ; \pi, \lambda). \tag{1.6}$$

A solution to (1.6)—and hence an equilibrium—clearly exists if both $\mathcal{T}_W$ and $\mathcal{T}_M$ are finite. When this condition does not hold, though, an equilibrium need not exist.[41] Furthermore, when an equilibrium exists, it need not be unique. Corollaries 1.5 and 1.6 illustrate examples of situations in which multiple equilibria, all of which yield the same expected payoff to the

---

[41] For example, suppose that $\phi(\mathcal{T}_W) = (0, 1)$ and $\mathcal{T}_M = (0, 1)$. Note that $\max \phi(\mathcal{T}_W)$ does not exist. Hence, given any worker, the principal can find another worker with a higher tolerance for skepticism (i.e., with the ability to acquire better information). Furthermore, given that any conceivable type of manager can be appointed, the principal can appoint a manager that will take maximal advantage of this latter worker's tolerance for skepticism. Hence, given any pair of worker and manager, the principal can find another pair that yields her a higher expected utility.

Principal's Welfare

$\pi = 0.8$, $\lambda = 0.5$, $\pi_W = 0.99$, $\lambda_W = 0.7$, $q(e) = e/(2e+1)$, $c(e) = e^2/2$
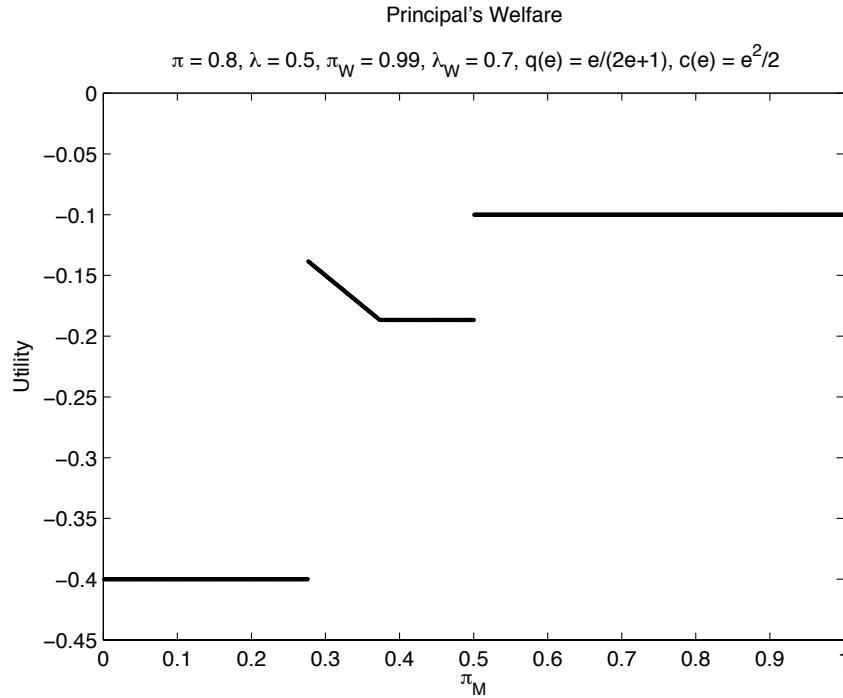
Figure 1.14: This worker is biased in favor of α. For this worker's type, $q(\hat{e}(t_W)) \approx 0.13$, $\phi_\alpha(t_W) \approx 0.22$, and $\phi_\beta(t_W) = 0$. Hence, the worker shirks when $\pi_M < 0.28$ or $\pi_M > 0.5$. Candidate managers with $0.28 \leq \pi_M < 0.37$ strengthen the worker's incentives. However, this worker is unwilling to acquire information of sufficiently high quality to meet the principal's informational standard, so the principal prefers the unconditional implementation of α, which she can achieve by retaining control of the decision.

principal, exist. Both of these results describe situations that involve rather extreme types of workers, whose incentives to acquire effort remain fixed across all available candidate managers. The types are described in Definitions 1.13 and 1.14, respectively.

Definition 1.13 (Self-driven worker)**.** Suppose that the type, $t_W \equiv (\pi_W, \lambda_W, q(\cdot), c(\cdot))$, of a candidate worker satisfies the following three conditions:

- $q(\hat{e}(t_W)) \geq \left| \pi_M - \frac{1}{2} \right|$ for all $\pi_M \in \mathcal{T}_M$;

- $\phi_\alpha(t_W) > 0$;

- $\phi_\beta(t_W) > 0$.

This candidate worker is *self-driven*.

Note that a candidate worker is self-driven with respect to the space, $\mathcal{T}_M$, of available managers. This feature might seem odd if self-drivenness is viewed as an intrinsic individual
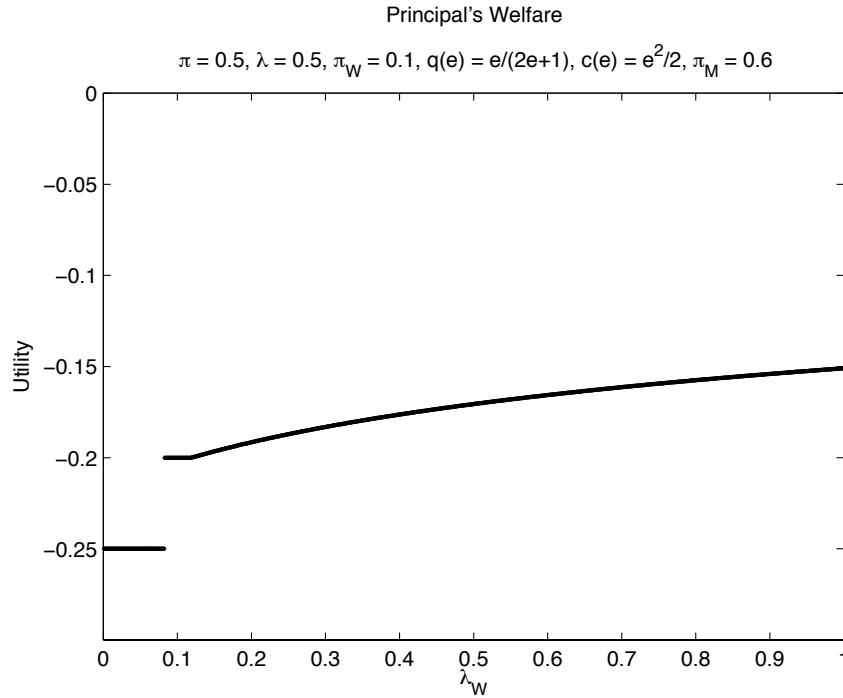
Figure 1.15: In this figure, all parameters are fixed except for λ_W. When $\lambda_W < 0.09$, the worker has no tolerance for skepticism regarding β. For $0.09 \leq \lambda_W \leq 0.11$, the worker's tolerance for skepticism regarding β exceeds the manager's skepticism, which, in turn, exceeds $q(\hat{e}(t_W))$. Thus, the manager strengthens the worker's incentives to acquire information, and the worker acquires a signal of quality 0.1, which is the manager's skepticism. As λ_W increases beyond 0.11, though, the worker acquires more information, since, as $\xi(t_W)$ increases, $q(\hat{e}(t_W))$ overtakes the manager's skepticism and continues to increase (as does the principal's welfare). That is, for $\lambda_W > 0.11$, the manager has no effect on the worker's incentives.

trait. However, an individual's self-drivenness arguably is determined at least partly by his environment. For example, a very talented researcher working in a lackluster environment in which no manager is insightful enough to challenge his ideas will not adjust his effort level based upon his manager's identity. However, if the researcher is moved to a more stimulating environment in which managers are talented and frequently challenge their workers, he is likely to work harder to overcome a manager's skepticism. Thus, a worker that is self-driven in one organization (a feature of which is $\mathcal{T}_M$) need not be self-driven in all organizations.

Just as a worker may be self-driven in the organization, a worker alternatively may be underqualified for the organization, in the sense that any candidate manager will be too skeptical for the worker, and hence the worker will shirk regardless of which manager is appointed:

Definition 1.14 (Underqualified worker)**.** A candidate worker of type $t_W \in \mathcal{T}_W$ is *un-*

Principal's Welfare

$\pi = 0.5$, $\lambda = 0.5$, $\pi_W = 0.6$, $q(e) = e/(2e+1)$, $c(e) = e^2/2$, $\pi_M = 0.6$
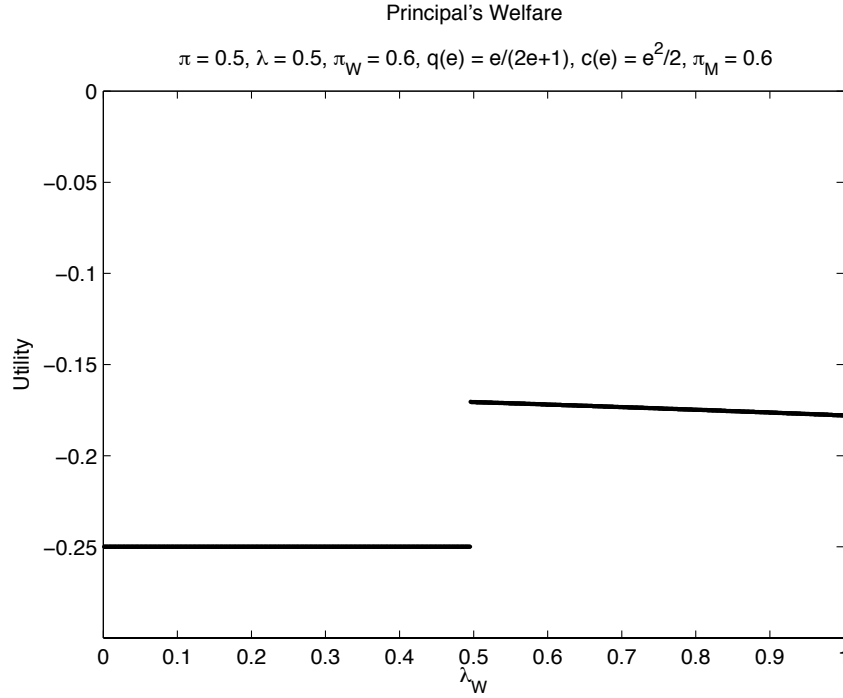
Figure 1.16: For candidate workers with the given parameters, whenever $\lambda_W < 0.5$, $\phi_\beta(t_W) = 0$ holds. Because the manager is skeptical regarding β, such workers shirk if appointed. Once $\lambda_W \geq 0.5$, though, $\phi_\beta(t_W)$ exceeds the manager's skepticism regarding β, as does $q(\hat{e}(t_W))$. Thus, for $\lambda_W \geq 0.5$, the manager has no effect on the worker's incentives. Counterintuitively, the worker's effort level (and the principal's welfare) is declining as λ_W increases beyond 0.5, despite the worker becoming more biased toward β (which the manager opposes ex ante). The reason is that $\xi(t_W)$ is decreasing. Thus, in spite of increasing disagreement between the manager and worker, the worker's information acquisition incentives are weakened as λ_W grows beyond 0.5, due to the fact that his valuation of information falls.

derqualified if, for every candidate manager type $\pi_M \in \mathcal{T}_M$, either $\phi_\alpha(t_W) < \frac{1}{2} - \pi_M$ or $\phi_\beta(t_W) < \pi_M - \frac{1}{2}$.

Corollaries 1.5 and 1.6 describe non-unique equilibria that arise in situations that involve self-driven and underqualified workers, respectively.

Corollary 1.5. *Suppose that*

$$\max_{t_W = (\pi_W, \lambda_W, q(\cdot), c(\cdot)) \in \mathcal{T}_W} q(\hat{e}(t_W))$$

exists and that there exists a self-driven candidate worker, of type $t_W^* = (\pi_W^*, \lambda_W^*, q^*(\cdot), c^*(\cdot))$, such that

$$t_W^* \in \arg\max_{t_W = (\pi_W, \lambda_W, q(\cdot), c(\cdot)) \in \mathcal{T}_W} q(\hat{e}(t_W)).$$

Principal's Welfare

$\pi = 0.5$, $\lambda = 0.5$, $\pi_W = 0.9$, $q(e) = e/(2e+1)$, $c(e) = e^2/2$, $\pi_M = 0.6$
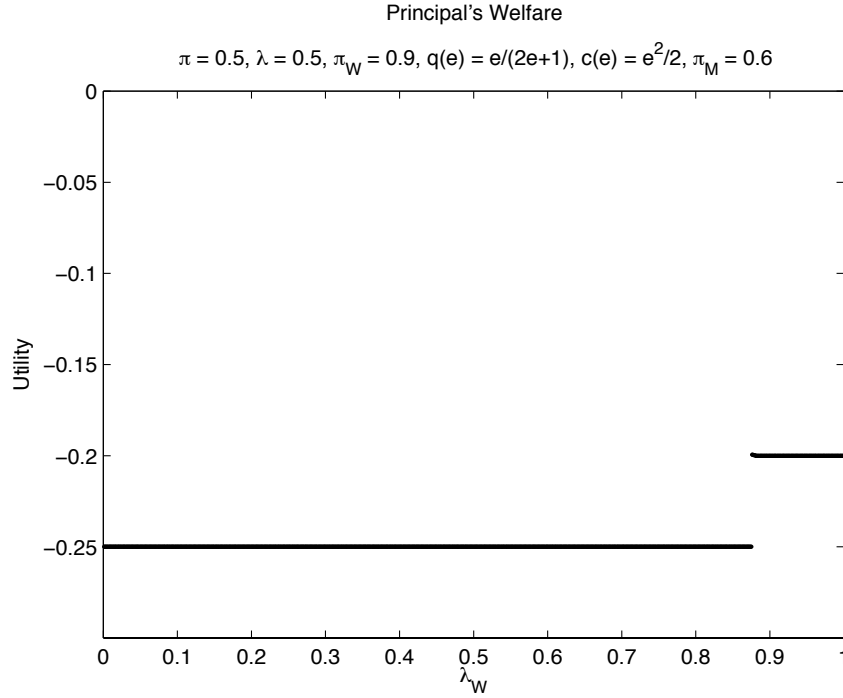
Figure 1.17: Candidate workers with $\lambda_W < 0.88$ have no tolerance for skepticism regarding β. Because the manager is skeptical regarding β, these workers shirk. Once $\lambda_W \geq 0.88$, though, $\phi_\beta(t_W)$ exceeds the manager's skepticism regarding β, so the worker no longer shirks. For $0.88 \leq \lambda_W < 0.89$, $q(\hat{e}(t_W))$ exceeds the manager's skepticism, so the manager does not affect the worker's incentives. For $\lambda_W \geq 0.89$, though, the manager's skepticism lies between $q(\hat{e}(t_W))$ and $\phi_\beta(t_W)$, so the worker obtains a signal that is strong enough to offset the manager's skepticism. In particular, the principal's welfare is fixed for these values of λ_W. Just as in the example of Figure 1.16, making the worker disagree more strongly with the manager does not necessarily improve the principal's welfare, since it lowers the worker's valuation of information.

For any candidate manager π_M^*, (t_W^*, π_M^*) is a solution to (1.6).

Corollary 1.6. *If all candidate workers are underqualified, the principal should retain control of the decision. In particular, let*

$$t_W^* \in \mathcal{T}_W,$$

$$\pi_M^* \equiv \frac{\pi\lambda - \pi}{2\pi\lambda - \pi - \lambda}.$$

The pair (t_W^*, π_M^*) is a solution to (1.6).

The result of Corollary 1.5 is intuitive: if there is a "superstar" candidate worker that is self-driven and can be expected to acquire better information than all other candidate workers, the principal should appoint him as the worker. Since the worker is self-driven, the

choice of appointed manager is irrelevant. Assuming that more than one candidate manager is available, the irrelevance of the manager leads to a multiplicity of equilibria. However, all equilibria yield the same expected utility to the principal.

Note that, while self-driven workers provide obvious value to the organization, they also render managers useless in providing incentives. Based upon this discussion, one might expect organizations with many self-driven workers to have flatter hierarchies. Indeed, flat hierarchies are observed in some startup firms, particularly in the technology sector.[42] One explanation is that, since such firms tend to be small, the principal does not need to delegate oversight responsibilities and other tasks to middle managers. Another explanation—which is supported by Corollary 1.5—is that the self-selection of talented workers into the inherently less stable environment of a startup indicates that these workers are passionate about the startup's vision and do not need (or, indeed, respond to) the pushes of managers that the firm could hire to provide extra incentives.

In the situation described by Corollary 1.6, all candidate workers are underqualified. When framed in such terms, this situation might seem highly unrealistic, but it can arise when each candidate manager holds such an extreme position regarding the decision that she will be unresponsive to any evidence that even the most resourceful worker would be willing to collect. In such a situation, the principal finds it optimal to make the decision herself: the worker plays no role in the decision. Thus, since the choice of worker is irrelevant, there are multiple equilibria. Once again, though, the principal achieves the same expected utility under all equilibria.

The key insight of this paper—that managers strengthen workers' incentives—is reflected in the equilibria described by Corollary 1.7, which makes use of Definition 1.15:

Definition 1.15 (Sufficiency to strengthen incentives). The pool, $\mathcal{T}_M$, of candidate managers is *sufficient to strengthen incentives* if, for all $t_W \equiv (\pi_W, \lambda_W, q(\cdot), c(\cdot)) \in \mathcal{T}_W$, there exists $\pi_M \in \mathcal{T}_M$ such that either of the following two conditions holds:

- $q(\hat{e}(t_W)) < \pi_M - \frac{1}{2} \leq \phi_\beta(t_W)$;

- $q(\hat{e}(t_W)) < \frac{1}{2} - \pi_M \leq \phi_\alpha(t_W)$.

Note that, if the pool of candidate managers is sufficient to strengthen incentives, no candidate worker is self-driven.

Corollary 1.7. *Suppose that the pool of candidate managers is sufficient to strengthen incentives. Let $\mathcal{T}^*$ denote the set of pairs of candidate workers and managers in which the candidate manager would strengthen the candidate worker's incentives if both were appointed:*

$$\left\{ (t_W, \pi_M) \in \mathcal{T}_W \times \mathcal{T}_M : q(\hat{e}(t_W)) < \pi_M - \frac{1}{2} \leq \phi_\beta(t_W) \text{ or } q(\hat{e}(t_W)) < \frac{1}{2} - \pi_M \leq \phi_\alpha(t_W) \right\}.$$

---

[42]See Teece (1996) and Silverman (2012) for evidence and examples.

If

$$\max_{(t_W, \pi_M) \in \mathcal{T}^*} \left| \pi_M - \frac{1}{2} \right|$$

exists and is no less than the principal's informational standard, then any pair

$$(t_W^*, \pi_M^*) \in \underset{(t_W, \pi_M) \in \mathcal{T}^*}{\arg\max} \left| \pi_M - \frac{1}{2} \right|$$

is a solution to (1.6).

The equilibria of Corollary 1.7 involve a manager that is maximally skeptical among the pool of candidate managers that strengthen the incentives of at least one worker. Since the pool of candidate managers is sufficient to strengthen incentives, every worker, when facing this manager, will either increase his effort level beyond his rubberstamping level (because his incentives are strengthened) or shirk (because his incentives are destroyed). Pairing this manager with any worker whose incentives she strengthens will maximize the principal's welfare.

Even when no solution to the principal's problem (and hence no equilibrium) exists, however, the punchline result—that the principal typically benefits from delegation—can hold:

Corollary 1.8. *Suppose that the pool of candidate managers is sufficient to strengthen incentives and that there exist $t_W \in \mathcal{T}_W$ and $\pi_M \in \mathcal{T}_M$ such that either of the following conditions holds:*

- $\frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda} < \pi_M - \frac{1}{2} \leq \phi_\beta(t_W).$

- $\frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda} < \frac{1}{2} - \pi_M \leq \phi_\alpha(t_W).$

It is suboptimal for the principal to retain control of the decision.

Corollary 1.8 asserts that the principal benefits from delegation when three conditions hold. First, the pool of candidate managers is sufficient to strengthen incentives: given any worker, there exists a manager that can increase the worker's effort from his rubberstamping effort level. Second, there exists a worker that is willing to acquire sufficiently precise information that the principal prefers the decision to be made on the basis of the acquired information than to have her ex ante preferred action implemented. Third, there exists a manager that can strengthen this worker's incentives to a greater extent than the principal can. An implication of this result is that, provided that the candidate pools of managers and workers are sufficiently rich—in the sense that extremely skeptical managers and workers with very high tolerances for skepticism can be found—the principal should not retain control of the decision.

The final main result, Corollary 1.9, demonstrates that the principal can exploit disagreement between the manager and worker.

Corollary 1.9. *Suppose that $\mathcal{T}_M$ is a dense subset of $(0,1)$. Consider the worker and manager pair (t_W, π_M), where $\tau(t_W) \cdot \left(\pi_M - \frac{1}{2}\right) > 0$ and $\phi(t_W) > \frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda}$. There exists $\pi'_M \in \mathcal{T}_M$ such that $\tau(t_W) \cdot \left(\pi'_M - \frac{1}{2}\right) < 0$ and $U_d^*(t_W, \pi_M; \pi, \lambda) < U_d^*(t_W, \pi'_M; \pi, \lambda)$.*

In words, the result states that, if the pool of candidate managers is very rich—or, in technical terms, a dense subset of $(0,1)$—then, given any worker and manager that prefer the same action ex ante, the principal can increase her own expected welfare by keeping the worker but replacing the manager with one whose bias opposes the worker's. The intuition is that the worker can be induced to "fight harder" if he has to overcome skepticism regarding his preferred action. In particular, if the manager is initially skeptical regarding the action that the worker prefers, the worker will be willing to acquire more information to overcome the manager's skepticism than he would if the manager's skepticism were against the worker's ex ante less preferred alternative. This result echoes the insights of Rotemberg and Saloner (1995) and Dewatripont and Tirole (1999): that clashes in interests can be helpful due to their positive effects upon conflicting parties' incentives for information acquisition. Furthermore, the result implies that the principal generally prefers to delegate the decision to a manager (particularly to one whose bias clashes with the worker's) rather than to the worker directly, as the latter option is equivalent to delegating the decision to a manager whose preferences are aligned with the worker's. In particular, the manager's role is not redundant: the principal does better by separating the tasks of information acquisition and decision making than by leaving both in the hands of one agent.

1.5 Evidence and Implications

As noted in Section 1.3, the model emphasizes information acquisition and decision making as the essential activities of the organization. Thus, one ought to expect that the model's predictions will apply best to organizations in which workers' tasks involve research, creativity, and innovation. A normative implication of the model is that managers in such organizations should be skeptical regarding new ideas, so that only the truly outstanding ones are implemented. Indeed, Sutton (2010) argues that, in environments in which innovation is important, a manager must "kill off all the bad ideas [...] and most of the good ideas, too." He relates an incident in which, in an address to Yahoo! managers, Apple co-founder Steve Jobs—a universally recognized paragon of entrepreneurship, creativity, and innovation—once asserted that "killing good ideas" is difficult but also is a sign of a great company.

Software companies are salient examples of organizations in which workers' tasks involve creativity and innovation. In a recent empirical study, Mollick (2011) estimates the impact of middle managers on firm performance variation in the computer game industry. Specifically, he employs a multiple membership cross-classified multilevel model to separate the effects of designers (whom he classifies as "innovators," based upon the nature of their tasks), producers ("middle managers"), and firms on variance in game revenue. Mollick finds that,

of these three contributors, middle managers have the largest effect on revenue variance. He notes that it is counterintuitive that middle managers are "critical to firm performance even in highly innovative industries," in which one might expect that the innovators would play the most crucial roles in determining firm performance, and he concludes that the result "suggests the need for further research into the mechanisms by which middle managers influence firm performance" (Mollick, 2011). This paper delineates one mechanism through which middle managers can affect firm performance directly in innovative industries: by providing the incentives that determine workers' levels of effort and performance.

In defense of a recent spate of layoffs—especially of middle managers—at Microsoft, Matt Rosoff of *Business Insider* attributes much of the firm's struggles to its "bloated, inefficient [...] corps of middle managers" that are "too scared or uninsightful to challenge Steve Ballmer and other senior leaders" (Rosoff, 2011). The view that an organization derives little value from middle managers that are unable to challenge senior leaders is consistent with the message of this paper: that the principal prefers a manager that can credibly implement a decision rule that differs from the one that she would employ if she retained control.

The theoretical effect of the manager's bias on the worker's incentives forms the crux of this paper's results. Wang and Wong (2012) present experimental evidence that supports a similar theory. They argue that a firm can benefit from a manager's (possibly irrational) escalation of commitment to a particular project. The intuition is that, given uncertainty about whether the firm will choose to complete the project, employees may underinvest in costly human capital investments, which, on one hand, will increase the likelihood of the project's success but, on the other hand, will be wasted if the project is abandoned. The appointment of a manager that has a tendency to "overcommit" to the completion of the project will improve employees' incentives to invest, thus raising the likelihood of the project's success and ultimately benefiting the firm. Just as in this paper, the manager serves as a useful commitment device for the firm. In two separate experiments involving, respectively, business undergraduates and part-time MBA students, Wang and Wong find support for the positive effect of managerial commitment bias on employee incentives.

One environment in which a widely acknowledged difference in values and beliefs exists between members of the productive layer and members of the administrative layer is that of government agencies. It is commonly observed that, due in part to self-selection into their profession, bureaucrats (i.e., workers in government agencies) tend to be overzealous— compared to the average member of society, whose interests these agencies purportedly seek to advance—regarding the agency's work (Gailmard and Patty, 2007; Prendergast, 2007). To offset this bias, relatively neutral overseers are appointed, ostensibly to play the roles of "devil's advocates."[43] One example of such an appointment is that of Joseph P. Kennedy as the first chairman of the Securities and Exchange Commission. As Gailmard and Patty (2012) note,

---

[43]I am grateful to Joseph Farrell for pointing out this feature of appointments to oversight positions in federal antitrust agencies.

> Kennedy was a shocking choice among New Deal reformers and their press allies. He had made a fortune in the 1920s by, among other activities, engaging in some of the very stock market manipulations excoriated in the Pecora hearings in the Senate leading up to the 1933 Securities Act. Though not a Wall Street insider on the order of J. P. Morgan, Kennedy was regarded by New Dealers as a prominent symbol of exactly what was wrong with financial institutions. He was the antithesis of a zealous crusader who might put securities trading on a sound moral footing; his appointment was denounced by liberals in Congress and the press as letting the fox guard the henhouse.

Gailmard and Patty (2012) observe that an appointment like that of Kennedy gives a regulatory agency policy preferences that lie between those of the government and those of the regulated private entities. Such an appointment serves a strategic purpose for the government, which needs to elicit information from private entities and to use that information in setting regulatory policy vis-à-vis those entities. The government, however, cannot credibly commit to abstain from using the elicited information to set policy in a manner that advances its own interests (and thus hurts the interests of the private entities). Thus, if the government attempts to elicit information and retains full discretion over policy, it will not succeed in extracting much information from the private entities and thus will not be able to set policy optimally. On the other hand, by delegating both tasks to the agency, which holds intermediate policy preferences, the government can achieve its policy preferences to a greater extent, since the private entities "trust" the agency more than they trust the government.[44] I note, however, that such an appointment also can be justified via the previously discussed "devil's advocate" argument, which is consistent with this paper's message.

The basic idea that disagreement and competing interests can improve organizations' performance has been posited in other contexts as well. Historians and political commentators, for example, widely extol former U.S. President Abraham Lincoln's appointment of a "team of rivals" to his Cabinet as a bold and shrewd decision that contributed to his success as a president (Goodwin, 2005). Meanwhile, *New York Times* columnist Roger Cohen (2012) attributes part of the difficulty that President Barack Obama faced in fulfilling public expectations regarding the impact of his first term in office to the president's failure to fill his Cabinet positions in a similar fashion. These examples illustrate the widespread appeal of a notion that the inclusion of biased individuals on a decision making committee—or, more generally, that diversity in teams—can improve outcomes through its positive effects on information acquisition and aggregation.[45] The paper's results help to formalize and support this notion.

---

[44]This type of argument is closely related to the analysis of Dessein (2002).

[45]See Gerardi and Yariv (2008) and Cai (2009) for formal treatments of these types of environments.

1.6 Extensions

In this section, I consider two extensions—more general payoff structures and richer spaces of states, actions, and signals—to the model of Section 1.3 and study their effects on the results of Section 1.4. I also show that the insights of Section 1.4 are robust to a modification of the game in which the manager is responsible for choosing the worker.

Accommodating More General Payoff Structures

Utility functions of the form of (1.1) are used in the model to represent the interests of each type of actor. These functions (and, more generally, the actors' type spaces) might appear overly restrictive, especially since they do not seem to accommodate transfer payments, which obviously are present in many different types of organizations. Although this apparent limitation does not create inconsistencies within the model (which abstracts away from transfer payments), it is natural to ask how the results might change if the model were to be modified to allow transfer payments. In this case, the principal may have to use transfer payments to satisfy participation constraints for both agents—especially for the worker, whose task carries an explicit cost—based upon the values of their outside options. Furthermore, the principal may be willing to provide payment to the worker to compensate him, at least partially, for his cost of information acquisition. In this section, I show that the utility functions that I use in the model are more general than they might appear. I also show that the model accommodates certain types of transfer payments, and I demonstrate how the principal can use these payments to her advantage.

Consider an actor whose payoffs are shown in Table 1.2. This actor has a weak preference

State

	A	B
α	$\overline{v}_\alpha$	$\underline{v}_\alpha$
β	$\underline{v}_\beta$	$\overline{v}_\beta$

Action

Table 1.2: This matrix illustrates a general payoff structure. Throughout the analysis, maintain the assumption that $\overline{v}_\alpha > \underline{v}_\beta$ and $\overline{v}_\beta > \underline{v}_\alpha$ (i.e., that the decision is nontrivial).

for α if and only if the probability, p_A, that her beliefs assign to state A satisfies

$$p_A \geq \frac{\overline{v}_\beta - \underline{v}_\alpha}{\overline{v}_\alpha - \underline{v}_\alpha + \overline{v}_\beta - \underline{v}_\beta}. \tag{1.7}$$

The condition in (1.7) also would characterize a weak preference for α if the principal's payoffs were, instead, the ones shown in Table 1.3. Fact 1.1 is a direct consequence of this analysis.

State

		A	B
Action	α	0	$-\frac{\overline{v}_\beta - \underline{v}_\alpha}{\overline{v}_\alpha - \underline{v}_\alpha + \overline{v}_\beta - \underline{v}_\beta}$
	β	$-\frac{\overline{v}_\alpha - \underline{v}_\beta}{\overline{v}_\alpha - \underline{v}_\alpha + \overline{v}_\beta - \underline{v}_\beta}$	0

Table 1.3: An actor with this payoff structure and an actor whose payoff structure is shown in Table 1.2 have aligned preferences.

Fact 1.1. Consider an actor whose prior beliefs place probability $p_A \in (0, 1)$ on state A and whose payoffs are shown in Table 1.2.

(i) As a principal, such an actor has type (π, λ), where

$$\pi \equiv p_A,$$

$$\lambda \equiv \frac{\overline{v}_\beta - \underline{v}_\alpha}{\overline{v}_\alpha - \underline{v}_\alpha + \overline{v}_\beta - \underline{v}_\beta}.$$

(ii) As a manager, such an actor has type π_M, where

$$\pi_M \equiv \frac{p_A \cdot \left(\frac{\overline{v}_\alpha - \underline{v}_\beta}{\overline{v}_\alpha - \underline{v}_\alpha + \overline{v}_\beta - \underline{v}_\beta}\right)}{p_A + (1 - 2p_A) \cdot \frac{\overline{v}_\beta - \underline{v}_\alpha}{\overline{v}_\alpha - \underline{v}_\alpha + \overline{v}_\beta - \underline{v}_\beta}}.$$

A similar analysis can be done for workers. The only difference is that workers incur costs by exerting effort. Consider, again, an actor whose prior beliefs place probability $p_A \in (0, 1)$ on state A and whose payoffs are shown in Table 1.2. Suppose that this actor can acquire information, and that his returns to effort and costs of effort, respectively, are summarized by the functions $q(\cdot)$ and $\kappa(\cdot)$. (The cost function, $\kappa(\cdot)$, carries the same units as the actor's payoffs and satisfies the same properties that $c(\cdot)$ was assumed to satisfy.) Suppose, also, that the appointed manager's type is π_M. Then, if this actor were to be appointed as a worker, his expected payoff, as a function of his effort level, would be

$$V(e\,;\pi_M, p_A, \mathbf{v}, q(\cdot), \kappa(\cdot)) \equiv$$
$$\begin{cases} -\kappa(e) & + \overline{v}_\alpha p_A + \underline{v}_\alpha \cdot (1 - p_A) & \text{if } \pi_M - \frac{1}{2} > q(e), \\ -\kappa(e) & + \underline{v}_\beta p_A + \overline{v}_\beta \cdot (1 - p_A) & \text{if } \frac{1}{2} - \pi_M > q(e), \\ -\kappa(e) & + [\overline{v}_\alpha p_A + \overline{v}_\beta \cdot (1 - p_A)] \cdot \left(\frac{1}{2} + q(e)\right) & \\ & + [\underline{v}_\beta p_A + \underline{v}_\alpha \cdot (1 - p_A)] \cdot \left(\frac{1}{2} - q(e)\right) & \text{otherwise.} \end{cases}$$

Now, define the following:

$$c(\cdot) \equiv \frac{\kappa(\cdot)}{\overline{v}_\alpha - \underline{v}_\alpha + \overline{v}_\beta - \underline{v}_\beta},$$

$$\lambda_W \equiv \frac{\overline{v}_\beta - \underline{v}_\alpha}{\overline{v}_\alpha - \underline{v}_\alpha + \overline{v}_\beta - \underline{v}_\beta},$$

$$t_W \equiv (p_A, \lambda_W, q(\cdot), c(\cdot)).$$

It is straightforward to verify that

$$V(\cdot\,; \pi_M, p_A, \mathbf{v}, q(\cdot), \kappa(\cdot)) = U_W(\cdot\,; \pi_M, t_W) \cdot \left(\overline{v}_\alpha - \underline{v}_\alpha + \overline{v}_\beta - \underline{v}_\beta\right) + \overline{v}_\alpha p_A + \overline{v}_\beta \cdot (1 - p_A), \quad (1.8)$$

where $U_W(\cdot\,; \pi_M, t_W)$ is defined in (1.4). (1.8) implies that

$$\arg\max_{e \geq 0} V(e\,; \pi_M, p_A, \mathbf{v}, q(\cdot), \kappa(\cdot)) = \arg\max_{e \geq 0} U_W(e\,; \pi_M, t_W).$$

This analysis establishes Fact 1.2.

Fact 1.2. Consider an actor whose prior beliefs place probability $p_A \in (0, 1)$ on state A and whose payoffs are shown in Table 1.2. Suppose that this actor can acquire information, and that his returns to effort and costs of effort, respectively, are summarized by the functions $q(\cdot)$ and $\kappa(\cdot)$. As a worker, such an actor has type t_W, where

$$t_W \equiv \left(p_A, \frac{\overline{v}_\beta - \underline{v}_\alpha}{\overline{v}_\alpha - \underline{v}_\alpha + \overline{v}_\beta - \underline{v}_\beta}, q(\cdot), \frac{\kappa(\cdot)}{\overline{v}_\alpha - \underline{v}_\alpha + \overline{v}_\beta - \underline{v}_\beta}\right).$$

Facts 1.1 and 1.2, together, show that the utility functions and type spaces that are used in the model are not as restrictive as they might appear to be. They are sufficiently rich to capture payoff structures in which the decision is nontrivial for all actors. It is crucial, though, that all of the actors agree on the optimal action (conditional on the state).

This analysis also illustrates how the model can accommodate certain types of transfer payments. Since the worker's effort level is unverifiable, contracts that specify transfer payments based upon the worker's effort level are unenforceable and thus unavailable to the principal. However, if realized outcomes (i.e., action and state pairs) are verifiable, the principal can credibly commit to providing transfer payments to the worker and manager based upon the realized outcome. Such payments can be built into the payoffs. For example, if the principal needs to induce participation by the worker and manager, she can provide transfers that will bring each actor's expected payoff above the payoff that the actor obtains from his or her outside option. The modified payoffs (including transfers) will determine the agents' types in the model.

Aside from allowing the principal to satisfy participation constraints, the provision of transfers also allows her to manipulate the agents' types. For example, consider an actor that, in the absence of transfer payments, is indifferent regarding the decision, with payoffs

as shown in Table 1.4.[46] (Note that, for this actor, the requirement that the decision is nontrivial is not satisfied.) The principal can provide outcome-contingent transfer payments

State

	A	B
α	v_A	v_B
β	v_A	v_B

Action

Table 1.4: An actor with this payoff matrix is indifferent regarding the decision.

that reward this actor when either of the "good" outcomes (i.e., (α, A) or (β, B)) is realized and punishes her (or rewards her less) otherwise. Such a contract essentially is an incentive or bonus scheme. Table 1.5 shows an example. This contract breaks the actor's indifference.

State

	A	B
α	$v_A + \overline{T}$	$v_B + \underline{T}$
β	$v_A + \underline{T}$	$v_B + \overline{T}$

Action

Table 1.5: The principal can use transfer payments, $\overline{T}$ and $\underline{T}$ (where $\overline{T} > \underline{T}$), to break the indifference of the actor with payoffs shown in Table 1.4.

In particular, the actor's modified payoff structure (with transfer payments) satisfies the assumption that the decision is nontrivial. Therefore, the actor can be appointed as a manager or worker, and the results of Section 1.4 apply.

Corollary 1.9 shows that the principal typically can benefit from differences of opinion between the manager and worker. As the analysis of this section thus far has shown, transfer payments help to determine agents' types, including their biases. Thus, the principal can use transfer payments to, in effect, "create discord" between a manager and a worker, and thus to improve the quality of decision making.[47] For example, consider two actors that, in the absence of transfers, achieve payoffs of 0 under all outcomes. Suppose that only one of the two actors can acquire information, and that both actors hold prior beliefs that assign probability $\frac{1}{2}$ to each state. The principal can offer these actors transfer schemes that give them the payoffs shown in Tables 1.6 and 1.7. Then the principal can appoint these actors

---

[46]Indeed, it seems implausible that an agent will care at all about the decision unless his or her wage depends upon the outcome.

[47]As the examples of Figures 1.16 and 1.17 show, however, determining the optimal level of discord is not trivial.

State

		A	B
Action	α	1	0
	β	0	2

Table 1.6: An actor with this payoff structure and uniform prior beliefs is biased in favor of β.

State

		A	B
Action	α	2	0
	β	0	1

Table 1.7: An actor with this payoff structure and uniform prior beliefs is biased in favor of α.

as a worker and manager, respectively. Their contracts lead them to have opposite biases, from which the principal can benefit.

The principal can use transfer payments to adjust a worker's tolerance for skepticism, too, since (as Fact 1.2 illustrates), transfer payments can strengthen a worker's incentives through the following three channels:

- Increasing the importance that he places on the optimality of the decision (i.e., increasing his valuation of information);

- Intensifying his bias in favor of a particular action (i.e., increasing the magnitude of his tilt);

- Subsidizing his information acquisition cost (i.e., scaling down his cost function).

It also should be noted that, by providing outcome-based transfers, the principal changes her own type as well. In particular, suppose that the principal's payoffs (without transfers) are as shown in Table 1.2, and that, with the transfers, her payoffs are as shown in Table 1.8. Based upon these modified payoffs, the principal's λ parameter is (by Fact 1.1)

$$\frac{\overline{v}_\beta - \underline{v}_\alpha + \left(\underline{T}_\alpha^W + \underline{T}_\alpha^M - \overline{T}_\beta^W - \overline{T}_\beta^M\right)}{\overline{v}_\alpha - \underline{v}_\alpha + \overline{v}_\beta - \underline{v}_\beta + \left(\underline{T}_\alpha^W + \underline{T}_\alpha^M + \underline{T}_\beta^W + \underline{T}_\beta^M - \overline{T}_\alpha^W - \overline{T}_\alpha^M - \overline{T}_\beta^W - \overline{T}_\beta^M\right)}.$$

Thus, the change in the principal's type will be negligible so long as the following condition holds:

State

	A	B
α	$\overline{v}_\alpha - \overline{T}_\alpha^W - \overline{T}_\alpha^M$	$\underline{v}_\alpha - \underline{T}_\alpha^W - \underline{T}_\alpha^M$
β	$\underline{v}_\beta - \underline{T}_\beta^W - \underline{T}_\beta^M$	$\overline{v}_\beta - \overline{T}_\beta^W - \overline{T}_\beta^M$

Action (to the left of the α, β rows)

Table 1.8: This matrix illustrates a payoff structure that the principal may face as a result of payment contracts that she signs with the worker and manager. If (1.9) holds, the transfer payments will have, at most, a small effect upon her induced type.

$$\frac{\overline{v}_\beta - \underline{v}_\alpha}{\overline{v}_\alpha - \underline{v}_\alpha + \overline{v}_\beta - \underline{v}_\beta} \approx$$

$$\frac{\overline{v}_\beta - \underline{v}_\alpha + \left(\underline{T}_\alpha^W + \underline{T}_\alpha^M - \overline{T}_\beta^W - \overline{T}_\beta^M\right)}{\overline{v}_\alpha - \underline{v}_\alpha + \overline{v}_\beta - \underline{v}_\beta + \left(\underline{T}_\alpha^W + \underline{T}_\alpha^M + \underline{T}_\beta^W + \underline{T}_\beta^M - \overline{T}_\alpha^W - \overline{T}_\alpha^M - \overline{T}_\beta^W - \overline{T}_\beta^M\right)}. \quad (1.9)$$

Note that (1.9) holds under the (reasonable) assumption that the transfer payments are small in magnitude compared to the principal's stake in making the correct decision.

In summary, the model is rich enough to capture the effects of outcome-contingent transfers. The principal can use such transfers as instruments to improve the quality of decision making. In light of this discussion, the results of Section 1.4 can be interpreted as illustrations of how, given a fixed incentive scheme based upon outcome-contingent transfer payments to the worker, the principal can further strengthen the worker's incentives by delegating the decision to a manager.

Alternate Information Structures

In the binary model that has been developed and analyzed, an ex ante conflict of interest between the worker and manager can strengthen the worker's incentives to acquire information. For this effect to arise, the acquired information must help to align the manager's preferences over actions with the worker's; otherwise the worker gains nothing from acquiring information. In light of this observation, it is clear that, if the worker and manager disagree about the joint distribution of the state and signal, or if, given a state, the worker and manager disagree about which action is optimal, then the worker's incentives to acquire information need not be strengthened by an ex ante difference between his interests and the manager's interests. Two features of the model—namely, the assumption of common knowledge regarding the signal structure and the similarity among all actors' utility functions—rule out these conditions, which may be viewed, respectively, as extreme forms of skepticism and partisanship.

A more subtle point is that, with a richer action space, a biased manager might make a decision that is ex post inefficient (from the principal's perspective), even upon observing

an arbitrarily informative (but noisy) signal. In this case, the appointment of a biased manager has two competing effects upon the principal's expected welfare. The first effect—the strengthening of the worker's incentives—is helpful to the principal. The second effect—loss of control—is harmful to the principal. In particular, given any imperfect signal, a biased manager will choose an action that the principal deems suboptimal, and the suboptimality (from the principal's perspective) of the manager's preferred action might be increasing in the magnitude of her bias. Based upon the relative strengths of these two effects, the principal may prefer to retain decision making authority than to delegate it to a biased agent.

In the binary model, ex post inefficient decisions are not made in equilibrium, since, in equilibrium, the worker's acquired information is sufficiently precise to overcome the manager's bias. In particular, although the principal and manager assign different expected utilities to the two actions after observing the signal, they agree upon which action is better in expectation. Thus, their interim preferences over actions are aligned, and so the principal does not suffer from loss of control. In a richer environment, though, the difference in interim preferences over actions may persist for arbitrarily informative signals.

A Continuous Environment

For a concrete example, suppose that the state space and action space are both $\mathbb{R}$, and that all actors share the ex post utility function $u(\eta, \omega) \equiv -(\eta - \omega)^2$. Both the principal and the (lone) worker believe that the state, ω, is normally distributed with mean 0 and variance 1. Candidate managers also believe that ω is normally distributed with variance 1, but they assign different means to the distribution. A given candidate manager can be identified by the mean, μ_M, that she assigns to the distribution of the state. I refer to μ_M as the manager's bias.

The worker can exert $e \geq 0$ units of effort to acquire a signal, s, that is distributed normally with mean ω and variance $\frac{1}{e}$. The cost (to the worker) of acquiring such a signal is $\frac{e^2}{2}$. His ex post utility when he expends e units of effort, action η is chosen, and ω is the underlying state, is $u(\eta, \omega) - \frac{e^2}{2}$. All features of the setting are common knowledge.

Suppose that the worker exerts effort e and obtains a signal s. Given the distributional assumptions, a manager of bias μ_M believes that ω is normally distributed with mean $\frac{\mu_M + es}{1+e}$ and variance $\frac{1}{1+e}$.[48] Both the principal and the worker, on the other hand, believe that ω is normally distributed with mean $\frac{es}{1+e}$ and variance $\frac{1}{1+e}$. Let $F(\cdot)$ denote the cdf associated with the principal's and worker's shared posterior beliefs regarding ω.

The manager's choice of action is $\frac{\mu_M + es}{1+e}$. The principal's expected utility can be computed as

$$\int -\left(\frac{\mu_M + es}{1+e} - \omega\right)^2 dF(\omega) = \int -\left(\frac{\mu_M}{1+e} + \left[\frac{es}{1+e} - \omega\right]\right)^2 dF(\omega)$$

---

[48]See DeGroot (1970) for a derivation.

$$= -\underbrace{\int \left(\frac{\mu_M}{1+e}\right)^2 dF(\omega)}_{=\frac{\mu_M^2}{(1+e)^2}} - \underbrace{\int 2 \cdot \frac{\mu_M}{1+e} \cdot \left(\frac{es}{1+e} - \omega\right) dF(\omega)}_{=0}$$

$$- \underbrace{\int \left(\frac{es}{1+e} - \omega\right)^2 dF(\omega)}_{=\frac{1}{1+e}}$$

$$= -\frac{\mu_M^2 + 1 + e}{(1+e)^2}.$$

The worker's expected utility, then, is

$$-\frac{\mu_M^2 + 1 + e}{(1+e)^2} - \frac{e^2}{2}.$$

Given the manager's bias, μ_M, the worker solves the problem

$$e_d^*(\mu_M) \in \arg\max_{e \geq 0} -\frac{\mu_M^2 + 1 + e}{(1+e)^2} - \frac{e^2}{2}.$$

The principal's problem is to choose an optimal level of bias:

$$\mu_M^* \in \arg\max_{\mu_M \in \mathbb{R}} -\frac{\mu_M^2 + 1 + e_d^*(\mu_M)}{(1 + e_d^*(\mu_M))^2}. \tag{1.10}$$

Note that, for any $\mu_M \in \mathbb{R}$, $e_d^*(\mu_M) = e_d^*(-\mu_M)$. Therefore the maximand in (1.10) is an even function of μ_M; only the magnitude, $|\mu_M|$, of the manager's bias is relevant. Figure 1.18 plots the worker's effort level and principal's expected welfare as functions of $|\mu_M|$. Evidently, in this environment, the principal prefers to retain decision making authority than to delegate it to a biased manager, despite the stronger incentives to acquire information that delegation provides the worker. In terms of the preceding discussion, the negative effect, upon the principal's welfare, of the loss of control to the biased manager overpowers the positive effect of the worker's improved incentives.

A Finite Environment

The suboptimality of delegation in the continuous environment described in Section 1.6 raises questions about the robustness of the result that the principal can benefit from delegation. Here, I show that delegation can be useful in finite environments that are more general than the binary setting of the baseline model.[49] To illustrate this point as simply and cleanly as possible, I present an analogue of the simple case of Section 1.4.

---

[49]In most applications, spaces of actions, states, and signals are finite due to resource constraints (e.g., a limited supply of money with fixed denominations), technological constraints (e.g., imperfect measurement equipment), or cognitive constraints. In light of the analysis of this section, the inferiority of delegation in the continuous environment should not raise severe concerns regarding the applicability of the results to real-world settings.
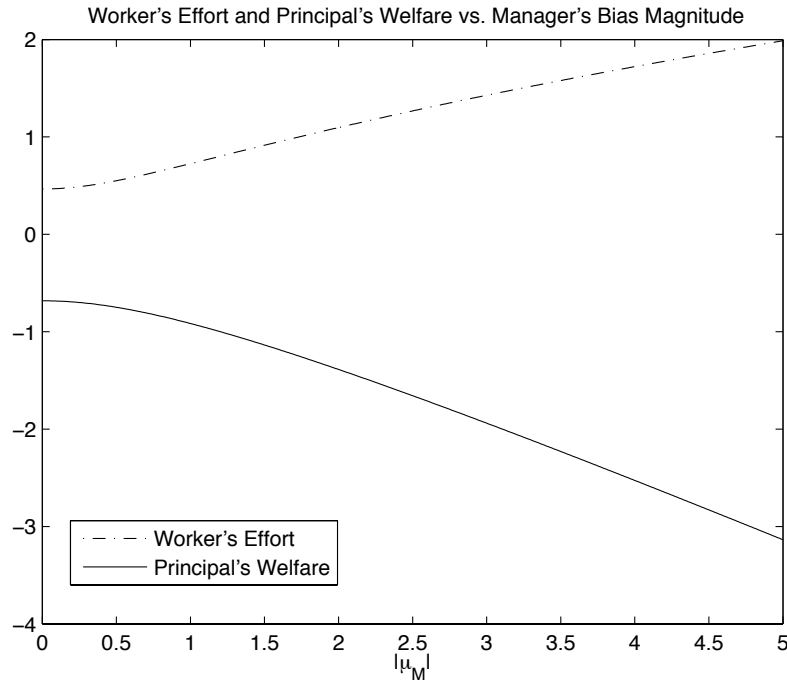
Figure 1.18: In this environment, the principal prefers to retain decision making authority, despite the fact that delegation to a biased manager strengthens the worker's incentives.

Consider an environment with state space $\{\omega_1, \ldots, \omega_N\}$, signal space $\{\sigma_1, \ldots, \sigma_N\}$, and action space $\{\alpha_1, \ldots, \alpha_N\}$. Nondegenerate beliefs in this environment can be described by elements of the $(N-1)$-dimensional open simplex,

$$\left(\Delta^N\right)^{\circ} \equiv \left\{ \boldsymbol{\pi} \in (0,1)^N : \sum_{i=1}^{N} \pi_i = 1 \right\}.$$

That is, if an actor's prior beliefs are given by $\boldsymbol{\pi} \in \left(\Delta^N\right)^{\circ}$, then, for each $i \in \{1, \ldots, N\}$, she assigns probability π_i to the state ω_i.

There is a single worker. Both he and the principal hold uniform prior beliefs, given by $\left(\frac{1}{N}, \ldots, \frac{1}{N}\right)$, over the state space. Candidate managers' beliefs, however, can vary. In particular, suppose that all candidate managers assign probabilities of at least $\frac{1}{N}$ to π_1 and believe that the states in $\{\omega_2, \ldots, \omega_N\}$ are equally likely. Thus, the pool of candidate managers can be represented by $\mathcal{T}_M \equiv \left[\frac{1}{N}, 1\right)$, and a candidate manager of type $\pi_M \in \mathcal{T}_M$ holds beliefs given by $\left(\pi_M, \frac{1-\pi_M}{N-1}, \ldots, \frac{1-\pi_M}{N-1}\right)$. All actors have the following ex post utility function:

$$u(\alpha_i, \omega_j) = \begin{cases} -1 & \text{if } i \neq j, \\ 0 & \text{otherwise.} \end{cases} \tag{1.11}$$

When the worker exerts e units of effort, he incurs a cost of $c(e)$ and obtains a signal of quality $q(e)$,[50] where, for each $i, j \in \{1, \ldots, N\}$,

$$\Pr\left(\sigma_i | \omega_j\right) = \begin{cases} \frac{1}{N} + (N-1)q(e) & \text{if } i = j, \\ \frac{1}{N} - q(e) & \text{otherwise.} \end{cases}$$

Just as in the binary model, the worker's problem is to choose an effort level that maximizes his expected utility. Given the form of the common utility function, the manager will choose the action that corresponds to the mode of her posterior beliefs over the state space. (For example, if the manager's posterior beliefs place highest weight on ω_2, she chooses α_2.) Note that the mode of the manager's posterior belief will be either ω_1 or the state that corresponds to the signal's realization.[51] Hence, the manager will choose either α_1 or the action that corresponds to the signal's realization.

Let $j \in \{2, \ldots, N\}$. From the perspective of a manager of type π_M,

$$\Pr\{\sigma = \sigma_j\} = \pi_M \cdot \left[\frac{1}{N} - q(e)\right]$$
$$+ \frac{1 - \pi_M}{N - 1} \cdot \left[\frac{1}{N} + (N-1)q(e) + (N-2) \cdot \left(\frac{1}{N} - q(e)\right)\right],$$
$$\Pr\left(\{\omega = \omega_1\} \,|\, \{\sigma = \sigma_j\}\right) = \frac{\pi_M \cdot \left[\frac{1}{N} - q(e)\right]}{\Pr\{\sigma = \sigma_j\}},$$
$$\Pr\left(\{\omega = \omega_j\} \,|\, \{\sigma = \sigma_j\}\right) = \frac{\frac{1-\pi_M}{N-1} \cdot \left[\frac{1}{N} + (N-1)q(e)\right]}{\Pr\{\sigma = \sigma_j\}}.$$

Such a manager will rubberstamp the worker's signal if and only if $\Pr\left(\{\omega = \omega_1\} \,|\, \{\sigma = \sigma_j\}\right) \le \Pr\left(\{\omega = \omega_j\} \,|\, \{\sigma = \sigma_j\}\right)$ for each $j \in \{2, \ldots, N\}$. This condition is equivalent to

$$q(e) \ge \frac{\pi_M - 1/N}{N - 1}.$$

Thus, the worker's objective function can be written as

$$U_W(e\,;\pi_M) \equiv \begin{cases} -c(e) - \frac{N-1}{N} & \text{if } e < q^{-1}\left(\frac{\pi_M - 1/N}{N-1}\right), \\ -c(e) - \frac{N-1}{N} + (N-1)q(e) & \text{otherwise.} \end{cases} \tag{1.12}$$

The two branches of (1.12) are shown in Figure 1.19. Note the similarity between (1.3)

---

[50]I assume that $c(\cdot)$ satisfies the same assumptions as in the binary model. In this model, $q(0) = 0$, $\lim_{e \to \infty} q(e) = \frac{1}{N}$, and $q(\cdot)$ is increasing, strictly concave, and twice differentiable. (An example is $q(e) \equiv \frac{e}{Ne+1}$.)

[51]For technical convenience, assume that, if the mode is not unique, then the manager chooses the action that corresponds to the signal realization (e.g., α_2 if the realization is σ_2). Just as in the baseline model, the usefulness of delegation to a biased manager does not rely on this assumption.
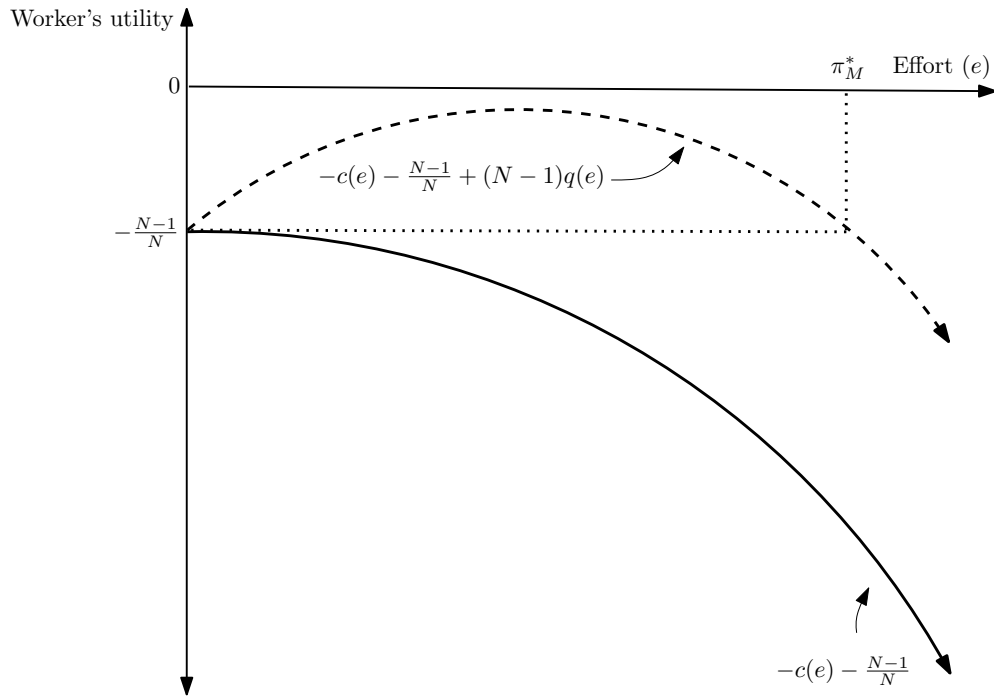
Figure 1.19: The worker's problem in the simple finite environment is similar to that from Section 1.4. The principal does best by choosing a manager of type π_M^*.

and (1.12). A similar analysis to that of Section 1.4 shows that, even in this environment, delegation to a manager that holds beliefs that differ from both the principal's and worker's is useful. An optimal manager will be one of type π_M^*, where π_M^* is the unique positive fixed point of the function $x \mapsto \frac{c\left(q^{-1}(x)\right)}{N-1}$.

Allowing the Manager to Appoint the Worker

In many organizations, managers—rather than principals—exercise direct control over the hiring of workers. A natural question to ask, therefore, is whether the preceding results change in a version of the game (depicted in Figure 1.20) in which the manager appoints the worker. As it turns out, this modification to the game does not change the results substantively. An argument in support of this fact relies on two key observations regarding the original game. The first observation is that the principal achieves no benefit from delegating the decision to a manager whose skepticism does not exceed her own informational standard. This observation is stated formally in Lemma 1.6.

Lemma 1.6. *Consider a candidate manager whose skepticism is no greater than the principal's informational standard, i.e.,*

$$\left| \pi_M - \frac{1}{2} \right| \leq \frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda}.$$
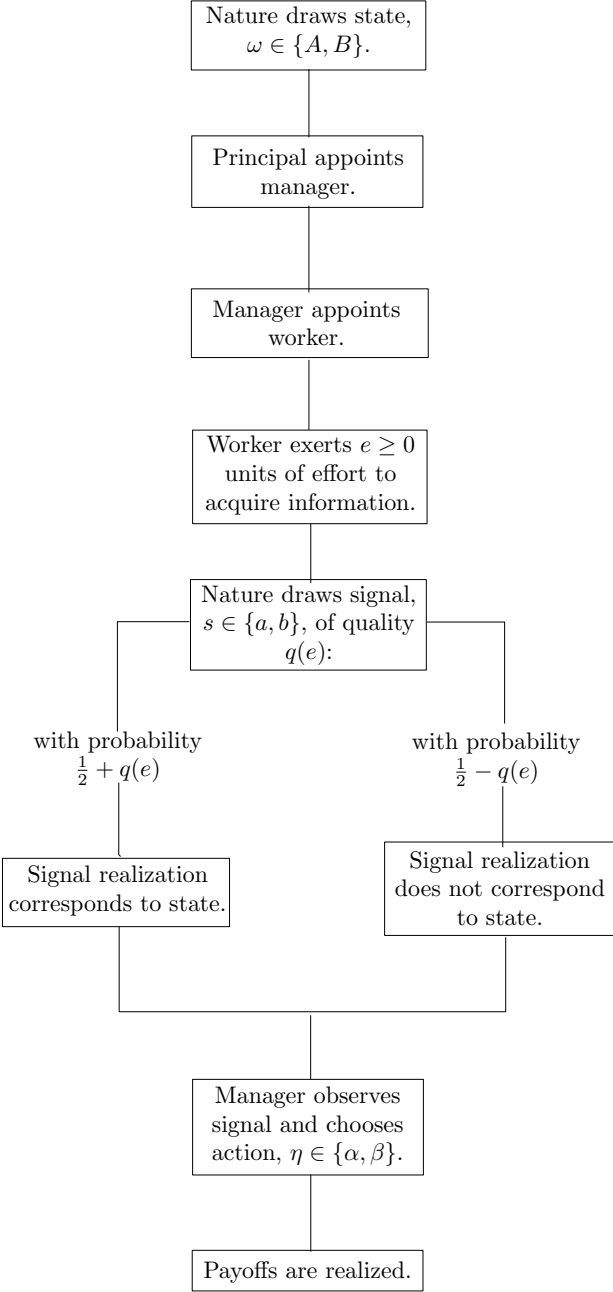
Figure 1.20: A modified version of the game in which the manager appoints the worker

Then, for any candidate worker type, t_W, $U_d^*(t_W, \pi_M \,; \pi, \lambda) \leq U_r^*(t_W \,; \pi, \lambda)$.

The intuition behind Lemma 1.6 is straightforward: in this framework, the middle manager derives value only through the positive effect of her skepticism on a worker's incentives. Thus, a middle manager that cannot provide additional skepticism is not useful to the prin-

cipal; the principal does at least as well by retaining control of the decision as she does by delegating control to such a manager.

The second observation is that a manager that is more skeptical than the principal, if given the additional responsibility of appointing a worker, does not have a strict incentive to appoint a worker that is suboptimal from the principal's perspective. To state this observation formally, it is helpful to define some notation.

Definition 1.16. Consider a manager of type π_M and a candidate worker of type t_W. Let $U_M^*(t_W ; \pi_M)$ denote the manager's *payoff from t_W*, her expected utility when she has control of the decision and the candidate worker of type t_W is appointed.

In the modified game, a candidate manager's payoffs from different candidate worker types define her preferences over candidate workers and, thus, determine her choice of worker. The second observation now can be stated precisely in the form of Lemma 1.7:

Lemma 1.7. *Consider a candidate manager whose skepticism is greater than the principal's informational standard, i.e.,*

$$\left| \pi_M - \frac{1}{2} \right| > \frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda}.$$

Let t_W and t_W' be two worker types. Suppose that $U_d^*(t_W, \pi_M ; \pi, \lambda) \geq U_d^*(t_W', \pi_M ; \pi, \lambda)$. Then $U_M^*(t_W ; \pi_M) \geq U_M^*(t_W' ; \pi_M)$.

Lemma 1.7 asserts that, given a manager that is "more biased" than the principal (in the sense that the manager's skepticism exceeds the principal's informational standard), the principal's (weak) preference relation over candidate worker types is a subset of the manager's. In particular, if the principal cedes control of the worker's appointment to a manager with a stronger bias than her own, she need not worry that the manager will have a strict incentive to appoint a different candidate worker from the one that she would have appointed if she had retained control of the worker's appointment.

Taken together, Lemmas 1.6 and 1.7 imply that the insights from the analysis of the original game of Figure 1.3 carry over to the modified game of Figure 1.20. In particular, any middle manager that allows the principal to achieve a strict welfare increase in the original game can be trusted, in the modified game, to appoint the worker in a manner that is consistent with the principal's preferences:

Fact 1.3. Any welfare increase that the principal can realize through delegation of the decision to a manager remains achievable, through delegation to the same manager, if the manager is allowed to appoint the worker.

1.7 Conclusion

In this paper, I have highlighted a previously unexplored channel through which middle managers—particularly ones whose characteristics and behavior may suggest that their roles

are wasteful—provide value to their organizations. The main insights are closely related to previously posited ideas regarding the benefits of disagreement and delegation in organizations and of commitment to ex post inefficient decision rules. One important contribution of this paper lies in the fact that it helps to reconcile two seemingly contradictory facts: the prevalence of middle management and its widespread lamentation. The paper builds a model of information acquisition and decision making in an organization to illustrate the fact that, by employing a manager whose interests do not coincide with hers, the principal of an organization can commit herself to a decision rule that will force a worker to exert more effort. When the principal appoints a worker and manager appropriately, the manager will rubberstamp the worker's report. A cursory look at the outcome might suggest that the manager's role is dispensable. However, such a conclusion ignores the fact that the manager's appointment strengthens the worker's incentives and, therefore, enhances the principal's welfare. The effect on the worker's incentives is especially pronounced when the principal appoints a manager whose bias clashes with the worker's. Thus, the results help to reconcile the pervasiveness of middle management with the popularity of negative stereotypes of middle managers. Evidence from various organizations—particularly ones in which knowledge production and innovation are important—is consistent with the model's predictions.

The results of this paper supplement existing answers to the question that I posed initially: "What is the value of middle management?" Another answer, which has a similar flavor to the one that I have proposed here, is due to Rotemberg and Saloner (2000). They argue that a firm can improve its employees' incentives to exert effort in innovation by both rewarding employees for implemented ideas and committing itself to a clear business strategy. The purpose of the commitment is to allow an employee to infer the types of projects that are likely to be implemented and, hence, to reduce the uncertainty that is associated with the returns to his or her effort. To achieve and demonstrate such a commitment, the firm can employ a "visionary" CEO, who holds a clear (but not overwhelming) bias toward certain types of projects. The firm can benefit further by employing autonomous and unbiased middle managers, thereby ensuring that profitable opportunities that fall outside the scope of the CEO's vision are explored and possibly implemented as well. Rotemberg and Saloner argue, thus, that the presence of middle managers provides incentives for employees to exert effort in exploring areas that may be profitable, even if the CEO does not favor those areas initially.

Although both the paper by Rotemberg and Saloner and my paper illustrate that a conflict of interest between decision makers in the organization can improve incentives for members of the productive (i.e., bottom) level of the hierarchy to exert costly effort, the two papers' results arise from very different views of the role of middle management. Rotemberg and Saloner assume that middle managers share the objective of maximizing the profitability of the organization, and they focus on the nature of the CEO's bias. In contrast, I begin from the premise that a middle manager's objectives are not necessarily well aligned with the organization's (or, equivalently, with the principal's), and I show that such a biased middle manager can advance the organization's objectives more than one that shares the organiza-

tion's objectives. Furthermore, I focus primarily on the nature of the middle manager's bias and also analyze the effects of the worker's bias.

Another rationale for the existence of middle management is proposed by Dessein (2002), who considers a receiver-sender (or, equivalently, principal-agent) environment with misalignment of interests (as in the canonical model of Crawford and Sobel (1982)). Dessein shows that, rather than either attempting to elicit the agent's (sender's) private information directly or delegating decision making authority to the agent, the principal (receiver) may prefer to delegate decision making authority to a third party (e.g., a middle manager) with intermediate preferences. This result holds whenever the conflict of interests is neither too small (so that delegation directly to the agent is not optimal) nor too large (so that the agency problem will not be too severe).[52]

Dessein's prediction is qualitatively different from the prediction of this paper. In this paper, the principal wants the middle manager to have a greater conflict of interest with the worker than she herself does. The key feature that accounts for this difference in predictions is that, whereas Dessein focuses on communication in an environment with exogenous information, I focus on incentives in an environment without strategic communication. Bias is helpful for strengthening incentives but is harmful for information transmission. The relative suitability of each model depends crucially upon key features—in particular, the nature of information acquisition and the modes of communication—of the environment that one wishes to study.

Open questions remain for future research. For example, organizations typically include multiple hierarchical levels, each with multiple agents. The existing literature on hierarchies rationalizes many of the observed structural features of organizations, but it would be useful to characterize optimal organizational structure in a generalization of this model that adheres to the framework described in Section 1.3. How can a principal optimally structure a team composed of multiple workers and managers with different specialties and abilities? The main insight—that delegation to biased managers is helpful for the principal—should persist in such a setting, but other insights regarding organizational structure may emerge. The model also does not contain much scope for strategic behavior by the worker. For example, there is no strategic communication of information. One interpretation of this feature is that audit technologies, which can detect any attempts to forge or tamper with acquired information, are sufficiently reliable—and punishments sufficiently harsh—to discourage the worker from employing any communication strategy other than truthful and precise revelation of the acquired signal. Admittedly, though, the assumption might be unreasonable in some settings. Furthermore, the qualities of both types of signals are determined identically by the worker's effort level. In principle, though, a worker could acquire information in such

---

[52]A similar result is due to Harris and Raviv (2005), who show that an increase in the difference in interests between a CEO and a division manager regarding an investment decision can lead to an *increase* in the CEO's willingness to delegate the decision to the division manager, since the change in the CEO's loss of control from delegating the decision might be lower than the change in his or her gain from avoiding noisy communication.

a way that one signal realization is more informative than the other.[53] In the context of the manufacturing firm example, the financial analyst could produce a forecast in which one recommendation (e.g., build at Avon) comes with a greater degree of confidence than the other recommendation. In either case, an increase in the worker's scope for strategic behavior is likely to temper the principal's incentives to appoint biased workers, who can profit by manipulating the decision. I conjecture, though, that, since biased workers can be induced to exert more effort, the principal's incentives to appoint them will not be destroyed altogether.

---

[53]Kamenica and Gentzkow (2011) study the worker's problem of choosing an optimal signal in this type of setting.

Chapter 2

Communication and Preference (Mis)alignment in Organizations

2.1 Introduction

An important function that division managers in firms and other organizations fulfill is the transmission of information from subordinates to superiors. For example, consider a situation in which a firm's owner needs to decide whether to continue funding a risky project. Given the scarcity of her time and her lack of familiarity with the project's finer details, it would be inefficient for the owner to elicit raw (and most likely opaque) information from frontline workers, let alone for her to attempt to acquire the information by herself. She would do better to consult division managers, who can elicit the information from their own subordinates and produce executive summaries that the owner can use in making a decision.

A principal (e.g., a firm's owner) in such a setting typically faces two types of agency problems. The first type is an incentive problem with workers: if information acquisition costs are privately borne by workers and unverifiable (and thus cannot be subsidized credibly through enforceable contracts), then workers will acquire less information than the principal would like.[1] The second type is a communication problem with managers: since the principal does not observe the acquired information, a manager can misrepresent it and thereby exploit her informational advantage to manipulate the principal's decision. Standard intuition suggests that the principal's total loss from these two agency problems is minimized when all parties' interests over organizational decisions are aligned.[2] The paper shows that, in contrast to the standard intuition, misalignment of preferences can be optimal from the principal's perspective. In particular, the principal benefits by "playing off" the opposing biases of a manager and worker in a persuasion game. In fact, even when a principal and

---

[1]Even if information costs can be partially subsidized, if acquiring perfect information is prohibitively costly (i.e., requires a worker to incur a higher cost than the maximum that the owner is willing to pay), then the worker will acquire less information than the principal would like.

[2]For example, Crawford and Sobel (1982) and Dessein (2002) illustrate that preference alignment is beneficial in settings with unverifiable and costless (i.e., cheap talk) communication.

worker have aligned preferences, the principal prefers to consult a manager with different preferences rather than one with identical preferences.

To grasp the intuition, consider an environment that comprises three actors: a principal, a manager, and a worker. The worker is assigned the task of acquiring a signal, at a cost that increases in the signal's precision, regarding the realization of the state of the world. The manager observes the signal and transmits a message (which corresponds to an executive summary or recommendation) to the principal. The principal then makes a decision. When all three parties have aligned interests, the principal faces no communication problem with the manager, but she does face an incentive problem with the worker. Suppose, however, that the manager's ex ante preferences over the decision clash (to a moderate extent) with the worker's. Then, in equilibrium, the worker acquires information that is more precise than in the previous case and, in particular, is sufficiently precise to overcome the manager's bias against the worker's preferred action. Furthermore, given the acquired information, the principal's preferences over actions are aligned with the manager's. Thus, the manager's incentives to mislead the principal vanish, and the principal suffers no loss from the communication problem. On the other hand, she mitigates the incentive problem and thus achieves an unambiguous improvement in her welfare.

The results and insights derived here echo those of Chapter 1, which illustrates how a middle manager that appears useless can provide value to an organization's principal. In particular, even if the principal's time, attention, and administrative capabilities are unlimited, she may benefit from delegating control of certain decisions to a manager that holds different beliefs or preferences from her own. Delegation to this type of manager strengthens the incentives of a worker that is responsible for acquiring information that is used to make a decision. This effect may not be apparent to a casual observer, though, and the middle manager may appear useless, or even detrimental, to the principal. The primary difference between the two papers is that, in the present paper's environment, the principal retains control of the decision. Thus, although she does not need to worry about an agency loss associated with ceding control to the manager, she may need to worry about strategic communication by the manager.

A well-established insight in previous literature is that, in situations that involve communication among parties with conflicting interests, the introduction of a neutral and nonstrategic mediator that privately collects information from players and submits recommendations to decision makers can increase efficiency.[3] A recent paper by Ivanov (2010) extends this result to a cheap talk environment with a single sender and single receiver (à la Crawford and Sobel (1982)) in which the mediator holds interests over the decision and forms her recommendations strategically. A decision maker in such a setting can appoint a strategic intermediary, with preferences that differ from her own, and obtain the same optimal payoff as she would had she appointed a nonstrategic, neutral intermediary. Just as in this paper, the decision maker finds it useful to select an intermediary with a bias that clashes with the expert's. Similarly, Che and Kartik (2009) illustrate that a decision maker may deliber-

---

[3]See, for example, Goltsman, Hörner, Pavlov, and Squintani (2009).

ately seek advice from a biased expert due to the positive effect of the bias on the expert's incentives to acquire information.

Section 2.2 presents a model of information acquisition, communication, and decision making in an organization along the lines sketched above. Section 2.3 develops the solution concept, highlights some examples that convey the intuition, and presents the main results (with proofs). Section 2.4 discusses the main results, describes some directions for future research, and concludes. Technical results are stated and proved in Appendix B.

2.2 Model

The principal faces a choice between two actions, α and β, and is uncertain about which action is optimal. Formally, there are two states of the world: A (in which α is optimal) and B (in which β is optimal). Assume that each state is equally likely. The principal's payoffs, defined over action and state pairs, are given by the following function over action and state pairs, where $p \in (0, 1)$:[4]

$$u : \{\alpha, \beta\} \times \{A, B\} \to \mathbb{R}_-$$

$$(d, \omega) \mapsto \begin{cases} -p & \text{if } d = \alpha \text{ and } \omega = B, \\ -(1-p) & \text{if } d = \beta \text{ and } \omega = A, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

As Lemma 2.1 will show, the parameter p in (2.1) represents the minimal probability that the principal's posterior beliefs need to place on state A for the principal to favor the action α. Hence, the principal is ex ante unbiased between α and β if and only if $p = \frac{1}{2}$, and $\left| p - \frac{1}{2} \right|$ denotes the degree of the principal's bias. I will refer to p as the principal's *preference parameter*. Observe that, the lower p is, the stronger the principal's ex ante preference for α.

To better inform the decision, the principal recruits the help of two agents: a worker and a manager.[5] The worker is chosen from a pool of candidate workers, each of whom is a specialist that can acquire information regarding the state. Similarly, the manager is chosen from a pool of candidate managers, each of whom can evaluate the information acquired by the worker and use it in preparing an executive summary for the principal. Specialization makes this arrangement efficient: neither the principal nor a manager holds a comparative advantage in acquiring the information, so a worker's participation is useful. Similarly, the

---

[4]Chapter 1 illustrates the fact that utility functions of the form of (2.1) are general enough to capture any situation that involves a nontrivial binary decision, and, furthermore, that the assumption that the two states are equally likely does not carry a substantive loss of generality.

[5]I assume that an agent that is recruited by the principal cannot decline to participate. Under the (very plausible) condition that the principal can satisfy the two agents' participation constraints with payments that are relatively small compared to her stake in the decision, this assumption is innocuous. See Chapter 1 for a more detailed discussion.

principal cannot interpret the raw information that a worker acquires, and a worker cannot
produce a sufficiently high-level report for the principal's use, so a manager's position is not
redundant. Finally, neither a worker nor a manager possesses the principal's vision of the
organization and grasp of strategy, so the principal must make the decision (although the
other actors may hold private preferences over the decision).

The appointed worker acquires information through a technology (common to all workers)
that provides a signal—a random variable that takes values in $\{a, b\}$—at a private cost to
him.[6] In particular, a worker can obtain a signal, s, of quality $q \in \left[0, \frac{1}{2}\right)$ at cost $c(q)$, where

$$\Pr\left(\{s = a\} \mid \{\omega = A\}\right) = \Pr\left(\{s = b\} \mid \{\omega = B\}\right) = \frac{1}{2} + q.$$

I assume that $c : \left[0, \frac{1}{2}\right) \to \mathbb{R}_+$ satisfies the following conditions:

- $c(0) = 0$;

- $\lim_{q \downarrow 0} c'(q) = 0$;

- $\lim_{q \downarrow 0} c''(q) > 0$;

- $\lim_{q \uparrow \frac{1}{2}} c(q) = \infty$;

- $c(\cdot)$ is strictly convex and twice differentiable.

An example of such a function is $c(q) \equiv \frac{q^2}{1-2q}$. Its graph is shown in Figure 2.1.

The appointed manager observes the signal's realization and quality and produces a
report, $r \in \left\{\hat{\alpha}, \hat{\beta}\right\}$. Finally, the principal receives the report and makes a decision, $d \in \{\alpha, \beta\}$. The game's timing is illustrated in Figure 2.2. Candidate workers and candidate
managers vary in their preferences over decisions. In particular, let a worker of type $w \in (0, 1)$
have preference parameter w (i.e., a payoff function of the form of (2.1), but with w in place
of p). Similarly, let a manager of type $m \in \left(0, \frac{1}{2}\right]$ have preference parameter m.[7] Let
$W \subseteq (0, 1)$ and $M \subseteq \left(0, \frac{1}{2}\right]$, respectively, denote the pools of available candidate workers
and candidate managers.

2.3   Analysis

This section develops the solution concept, discusses a couple of illustrative examples, and
then presents the main results.

---

[6]Chapter 1 allows workers to differ in their skill levels as well, but such a feature is not essential for the
results of this paper.

[7]The assumption that all candidate managers favor α (at least weakly) ex ante is without loss of general-
ity, as the symmetry of the environment allows the results to be readily extensible to cases in which candidate
managers may favor β. Furthermore, in some environments, the assumption can reflect a systematic bias in
managers' incentives (for example, against ventures that make managers look especially bad if unsuccessful).
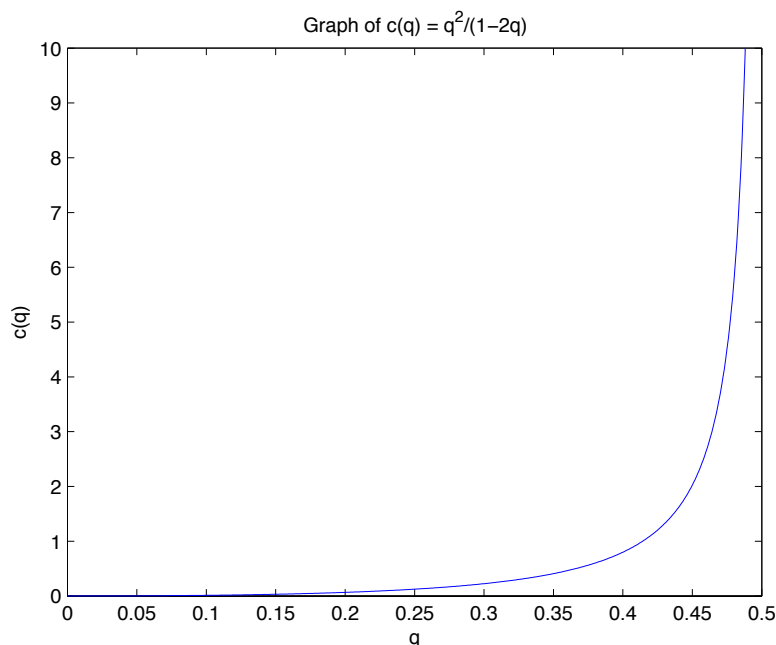
Figure 2.1: Graph of the function $c(q) \equiv \frac{q^2}{1-2q}$

Solution Concept

Note that all candidate workers and candidate managers are players in this game. Since each of these agents can be identified only by his or her type, an agent's strategy maps his or her type and information set to an action. In particular, a pure strategy profile is a quintuple of the form $(w, m, \iota, \rho, \delta)$, where:

- $w \in W$ is the type of appointed worker.

- $m \in M$ is the type of appointed manager.

- ι is an *information acquisition function*, a mapping from $W \times M$ to $\left[0, \frac{1}{2}\right)$. In particular, for each worker type, w', and each manager type, m', $\iota(w', m')$ is the quality of signal that a worker of type w' chooses to acquire (under the strategy ι) when a manager of type m' is appointed;

- ρ is a *recommendation function*, a mapping from $W \times M \times \{a, b\} \times \left[0, \frac{1}{2}\right)$ to $\left\{\hat{\alpha}, \hat{\beta}\right\}$. Suppose that the worker's type is w', the manager's type is m', the signal's realization is s, and the signal's quality is q. Then $\rho(w', m', s, q)$ denotes the report that a manager of type m' makes (under the strategy ρ) when a worker of type w' acquires a signal of realization s and quality q.
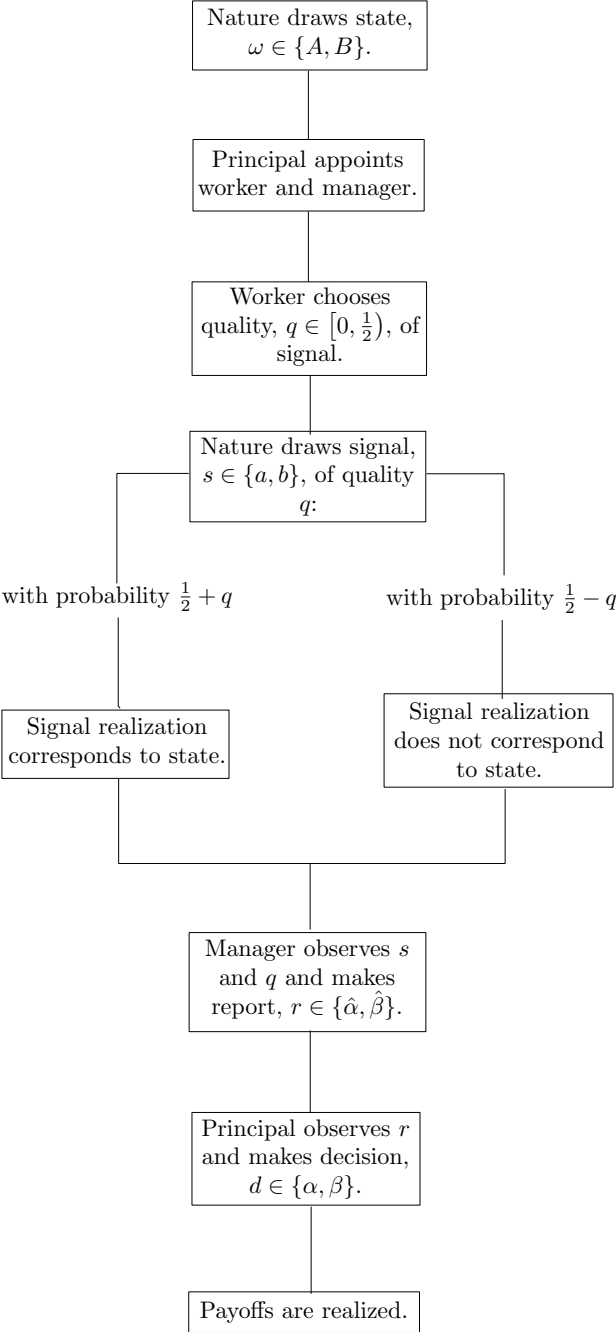
Figure 2.2: The timing of the game

- δ is a *decision function*, a mapping from $W \times M \times \left\{\hat{\alpha}, \hat{\beta}\right\}$ to $\{\alpha, \beta\}$. When the worker's type is w', the manager's type is m', and the recommendation is r, the principal's decision (under δ) is $\delta(w', m', r)$.

A worker's role in this environment is to obtain information for the decision. Lemma 2.1 describes how actors' preferences over actions respond to signals. This result is helpful in defining the solution concept.

Lemma 2.1. *Suppose that an actor with preference parameter λ observes a signal of realization s and quality q.*

   (i) The actor strictly prefers α if and only if either of the following conditions holds:

      R_1: $s = a$ and $\lambda - \frac{1}{2} < q$;
      R_1': $s = b$ and $q < \frac{1}{2} - \lambda$.

  (ii) The actor strictly prefers β if and only if either of the following conditions holds:

      R_2: $s = a$ and $q < \lambda - \frac{1}{2}$;
      R_2': $s = b$ and $\frac{1}{2} - \lambda < q$.

 (iii) The actor is indifferent between α and β if and only if either of the following conditions holds:

      R_3: $s = a$ and $q = \lambda - \frac{1}{2}$;
      R_3': $s = b$ and $q = \frac{1}{2} - \lambda$.

Proof. Let L_1 denote the condition in which the actor strictly prefers α, L_2 denote the one in which the actor strictly prefers β, and L_3 denote the one in which the actor is indifferent between α and β. It is clear that L_1, L_2, and L_3 are mutually exclusive and exhaustive, as are the six conditions R_1 through R_3'. Thus it will suffice to show that L_i implies either R_i or R_i' for each $i \in \{1, 2, 3\}$.

Consider statement (i), and suppose that the actor strictly prefers α. There are two possible cases:

Case 1: $s = a$. In this case, the actor's strict preference for α implies that

$$-\lambda \cdot \left(\frac{1}{2} - q \right) > -(1 - \lambda) \cdot \left(\frac{1}{2} + q \right),$$

   which is equivalent to

$$q > \lambda - \frac{1}{2}.$$

Case 2: $s = b$. Here, a strict preference for α implies that

$$-(1 - \lambda) \cdot \left(\frac{1}{2} - q \right) > -\lambda \cdot \left(\frac{1}{2} + q \right),$$

   which is equivalent to

$$q < \frac{1}{2} - \lambda.$$

Thus L_1 implies that either R_1 or R'_1 holds. Analogous arguments establish the fact that, for each $i \in \{2, 3\}$, L_i implies that either R_i or R'_i holds. $\qquad\square$

The solution concept, described explicitly in Definition 2.1, is a straightforward adaptation of perfect Bayesian equilibrium to this environment.

Definition 2.1. An *equilibrium* is a pure strategy profile, $(w^*, m^*, \iota^*, \rho^*, \delta^*)$, that satisfies the following conditions:

Optimal appointment: Given ι^*, ρ^*, and δ^*, w^* and m^* maximize the principal's expected utility. That is,

$$(w^*, m^*) \in \underset{\substack{w \in W, \\ m \in M}}{\arg\min} \sum_{\substack{r \in \{\hat{\alpha}, \hat{\beta}\}, \\ s \in \{a, b\}}} f_{\rho^*}(r; w, m, s, \iota^*(w, m))$$

$$\cdot \left[g_{\delta^*}(\alpha; r, w, m) \cdot h(s; B, \iota^*(w, m)) \cdot p + g_{\delta^*}(\beta; r, w, m) \cdot h(s; A, \iota^*(w, m)) \cdot (1 - p) \right],$$

where

- For any recommendation function, ρ; recommendation, r; signal realization, s; signal quality, q; candidate worker type, w; and candidate manager type, m,

$$f_\rho(r; w, m, s, q) \equiv \begin{cases} 1 & \text{if } \rho(w, m, s, q) = r, \\ 0 & \text{otherwise,} \end{cases}$$

- For any decision function, δ; decision, d; recommendation, r; candidate worker type, w; and candidate manager type, m,

$$g_\delta(d; r, w, m) \equiv \begin{cases} 1 & \text{if } \delta(r, w, m) = d, \\ 0 & \text{otherwise.} \end{cases}$$

- For any signal realization, s; state, ω; and signal quality, q,

$$h(s; \omega, q) \equiv \begin{cases} \frac{1}{2} + q & \text{if } (s, \omega) \in \{(a, A), (b, B)\}, \\ \frac{1}{2} - q & \text{otherwise.} \end{cases}$$

Optimal information acquisition: For each candidate worker type, $w \in W$; and candidate manager type, $m \in M$, given $\rho^*(m, w, \cdot, \cdot)$ and δ^*, $\iota^*(w, m)$ maximizes the w-type worker's expected utility. That is, for all $w \in W, m \in M$,

$$\iota^*(w, m) \in \underset{q \in [0, \frac{1}{2})}{\arg\min} c(q) + \sum_{\substack{r \in \{\hat{\alpha}, \hat{\beta}\}, \\ s \in \{a, b\}}} f_{\rho^*}(r; s, q, w, m)$$

$$\cdot \left[g_{\delta^*}(\alpha; r, w, m) \cdot h(s; B, q) \cdot w + g_{\delta^*}(\beta; r, w, m) \cdot h(s; A, q) \cdot (1 - w) \right]$$

Optimal recommendation: For each candidate worker type, $w \in W$; candidate manager type, $m \in M$; signal realization, $s \in \{a, b\}$; and signal quality, $q \in [0, \frac{1}{2})$, given δ^*, $\rho^*(w, m, s, q)$ maximizes the m-type manager's expected utility. That is,

- If there exists $r \in \{\hat{\alpha}, \hat{\beta}\}$ such that $\delta^*(w, m, r) = \alpha$, then $\delta^*(w, m, \rho^*(w, m, s, q)) = \alpha$ whenever $s = a$ or $q < \frac{1}{2} - m$.

- If there exists $r \in \{\hat{\alpha}, \hat{\beta}\}$ such that $\delta^*(w, m, r) = \beta$, then $\delta^*(w, m, \rho^*(w, m, s, q)) = \beta$ whenever $s = b$ and $q \geq \frac{1}{2} - m$.

Optimal decision making: For each candidate worker type, $w \in W$; candidate manager type, $m \in M$; and recommendation, $r \in \{\hat{\alpha}, \hat{\beta}\}$, given $\iota^*(w, m)$ and $\rho^*(w, m, \cdot, \cdot)$, $\delta^*(w, m, r)$ maximizes the principal's expected utility. That is, given a recommendation, r; candidate worker type, w; and candidate manager type, m,

- If there exists a unique signal realization, s, such that $\rho^*(w, m, s, \iota^*(w, m)) = r$, then

$$\delta^*(w, m, r) = $$
$$\begin{cases} \alpha & \text{if } \left[s = a \text{ and } p - \frac{1}{2} \leq \iota^*(w, m)\right] \text{ or } \left[s = b \text{ and } \frac{1}{2} - p > \iota^*(w, m)\right], \\ \beta & \text{otherwise.} \end{cases}$$

- If $\rho^*(w, m, a, \iota^*(w, m)) = \rho^*(w, m, b, \iota^*(w, m))$, then, for each $r \in \{\hat{\alpha}, \hat{\beta}\}$,

$$\delta^*(w, m, r) = \begin{cases} \alpha & \text{if } p < \frac{1}{2} \text{ or } [p = \frac{1}{2} \text{ and } r = \hat{\alpha}], \\ \beta & \text{otherwise.} \end{cases}$$

Several points regarding the definition are worth noting. First, the messages, $\hat{\alpha}$ and $\hat{\beta}$, are allowed to assume their meanings endogenously within the equilibrium. Second, a manager that is indifferent between α and β (based upon the observed play of the game and her beliefs about others' actions) acts in accordance with the signal's realization. (Similarly, the principal acts in accordance with her inference about the signal's realization).[8] Third, when the principal expects the manager to send an uninformative recommendation (i.e., one that she believes the manager will send regardless of the signal realization), she makes a decision in accordance with her prior beliefs.[9]

As is often the case in cheap talk games, some equilibria of this game are uninformative.

---

[8]This assumption provides technical convenience but is not crucial for the results. See Chapter 1 for details.

[9]Note that, if the principal expects the manager to send a certain recommendation regardless of the signal realization and instead receives a different recommendation, she still treats it as though it is uninformative. The rationale for this behavior is that, if the principal expects the worker to acquire information

Definition 2.2. A *babbling equilibrium* is an equilibrium in which the principal implements her ex ante preferred choice regardless of the appointed manager's recommendation.

As the name suggests, no information is transmitted in such an equilibrium. In particular, the worker understands that his action has no effect on the decision, so he acquires a free (and uninformative) signal. Conversely, if the principal anticipates this behavior and expects the manager's recommendation to be uninformative, she chooses an action based purely upon her prior beliefs regarding the state.

Proposition 2.1. *A babbling equilibrium always exists.*

Proof. Let W and M be given, and choose any $w^* \in W$ and $m^* \in M$. Set the remaining components of the strategy profile as follows:

- $\iota^*(w, m) \equiv 0$ for all $(w, m) \in W \times M$.

- $\rho^*(w, m, s, q) = \hat{\alpha}$ for all $(w, m, s, q) \in W \times M \times \left\{\hat{\alpha}, \hat{\beta}\right\} \times \left[0, \frac{1}{2}\right)$.

- For all $(w, m, r) \in W \times M \times \left\{\hat{\alpha}, \hat{\beta}\right\}$,

$$\delta^*(w, m, r) \equiv \begin{cases} \alpha & \text{if } p \le \frac{1}{2}, \\ \beta & \text{otherwise.} \end{cases}$$

Observe that $(w^*, m^*, \iota^*, \rho^*, \delta^*)$ is a babbling equilibrium in which no worker acquires an informative signal, every middle manager submits an uninformative recommendation of $\hat{\alpha}$, and the principal chooses her ex ante preferred action.[10] In particular, given ι^*, the principal's beliefs regarding the state do not change in response to the manager's recommendation; the principal expects the worker's signal—and, by extension, the manager's recommendation—to be completely uninformative. Given the principal's beliefs, neither a worker nor a manager can influence the decision. Thus, no worker has an incentive to acquire an informative signal, and no manager has an incentive to provide a recommendation other than $\hat{\alpha}$. $\square$

Babbling equilibria represent complete organizational failures, in which the principal achieves no benefit from her appointments of a worker and manager and makes the decision independently. From this point, I focus on the more plausible (and more interesting) class of informative (i.e., non-babbling) equilibria:

Definition 2.3. An *informative equilibrium* is an equilibrium in which the principal's decision depends on the appointed manager's recommendation.

---

according to ι^* and also to anticipate that the manager's response to ι^* will be to provide an uninformative recommendation, then the principal should expect the worker to acquire a costless (and uninformative) signal. The principal thus believes that an unexpected recommendation is due merely to an inconsequential deviation by the manager rather than to a costly deviation by the worker.

[10]In the case of indifference, the principal chooses α, but she could just as well choose β.

Remark 2.1. Any equilibrium is of either the babbling or the informative variety.

Suppose, for the moment, that an appointed worker is certain that his signal will determine the decision (i.e., that the manager will submit a recommendation that is based upon the signal and that the principal will rubberstamp the recommendation). In this case, the worker's choice of signal quality is a solution to the following minimization problem:

$$\min_{q \in \left[0, \frac{1}{2}\right)} c(q) - \frac{q}{2} + \frac{1}{4}. \tag{2.2}$$

Since the objective function in (2.2) is differentiable, decreasing at 0 (but eventually increasing), and strictly convex, the minimizer, q^*, is unique and is characterized by the first-order condition $c'(q^*) = \frac{1}{2}$. q^* represents the quality of signal that the worker would acquire if the decision were to be made based purely on the signal.

Definition 2.4. Let q^* denote the unique solution to the first-order condition, $c'(q^*) = \frac{1}{2}$. Call q^* the *worker-optimal signal quality under rubberstamping.*

Because the realization of the worker's signal determines the decision in the situation considered above, it might appear that the worker's incentives to acquire information are strongest in that situation, and that q^* is the highest signal quality that a worker could be motivated to acquire. As it turns out, however, a worker can be motivated to acquire a signal of higher quality if he has to work to overcome the bias of a manager whose recommendations the principal implements.

Suppose that the worker anticipates that the manager will incorporate the acquired information into her recommendation—which the principal rubberstamps—if and only if the information is of sufficiently high quality. In particular, suppose that the manager's type is m, and that both the worker and manager anticipate that the principal will rubberstamp the manager's recommendation. Given Lemma 2.1 and the fact that $M \subseteq \left(0, \frac{1}{2}\right]$, the manager's recommendation will be $\hat{\alpha}$ if $q < \frac{1}{2} - m$ and will be based on the signal otherwise.[11] Thus, the worker's problem is

$$\min_{q \in \left[0, \frac{1}{2}\right)} \kappa(q; w, m), \tag{2.3}$$

where

$$\kappa(q; w, m) \equiv \begin{cases} c(q) + \frac{w}{2} & \text{if } q \in \left[0, \frac{1}{2} - m\right), \\ c(q) - \frac{q}{2} + \frac{1}{4} & \text{if } q \in \left[\frac{1}{2} - m, \frac{1}{2}\right). \end{cases}$$

Figures 2.3, 2.4, and 2.5 plot the two branches of $\kappa(\cdot; w, m)$ for different values of w. In the figures, the term $\overline{q}(w)$ represents the maximal quality of signal that a worker of type w can be induced to acquire. A technical characterization of $\overline{q}(w)$ is provided in Definition 2.5. First, a preliminary result:

---

[11]In accordance with Definition 2.1, I assume that, when indifferent between the two actions that her recommendation may induce, the manager will make a recommendation in line with the signal's realization.

Lemma 2.2. *Let $w \in W$ be given, and consider the function*

$$\varphi_w : \left[0, \frac{1}{2}\right) \to \mathbb{R},$$

$$q \mapsto 2c(q) - \left(w - \frac{1}{2}\right).$$

φ_w has at most two fixed points.

Proof. Fix $w \in W$, and consider the following function:

$$\psi_w : \left[0, \frac{1}{2}\right) \to \mathbb{R},$$

$$q \mapsto \varphi_w(q) - q.$$

Observe that $q \in \left[0, \frac{1}{2}\right)$ is a fixed point of φ_w if and only if $\psi_w(q) = 0$. Since $\psi'_w(q) = 0$ if and only if $q = q^*$, Rolle's Theorem implies that $\psi_w(q) = 0$ for no more than two values of q. The result follows. $\qquad\square$

Note that, because the set of fixed points of φ_w is finite, it is either empty or has a well defined maximum. This point is important for the following definition.

Definition 2.5. Define

$$Q_w \equiv \left\{ q \in \left[0, \frac{1}{2}\right) : \varphi_w(q) = q \right\},$$

and

$$\bar{q}(w) \equiv \begin{cases} 0 & \text{if } Q_w = \varnothing, \\ \max Q_w & \text{otherwise.} \end{cases}$$

In words, if φ_w has any fixed points, $\bar{q}(w)$ is its largest fixed point. Otherwise, $\bar{q}(w)$ is set to 0. Call $\bar{q}(w)$ the *maximal signal quality* for the worker of type w.

As its name indicates, $\bar{q}(w)$ is the highest quality of signal that a worker of type w is willing to acquire under the belief that the principal will rubberstamp the manager's recommendation.[12] Figures 2.3, 2.4, and 2.5 illustrate this point.

As the figures suggest, a worker's information acquisition decision depends crucially upon the manager's type. Lemma 2.3 characterizes this decision.

Lemma 2.3. *Suppose that the worker's type is w, that the manager's type is m, and that the worker believes that the principal will follow the manager's recommendation. Assume the following:*

---

[12]There is a slight caveat worth noting here. If the manager is unbiased (i.e., has $m = \frac{1}{2}$) and always sends a recommendation that corresponds to the signal, then a worker with $\bar{q}(w) = 0$ will want to acquire a signal of quality q^*. This scenario is a knife-edge case, though; for even a manager of arbitrarily small (but positive) bias, such a worker will not acquire an informative signal.
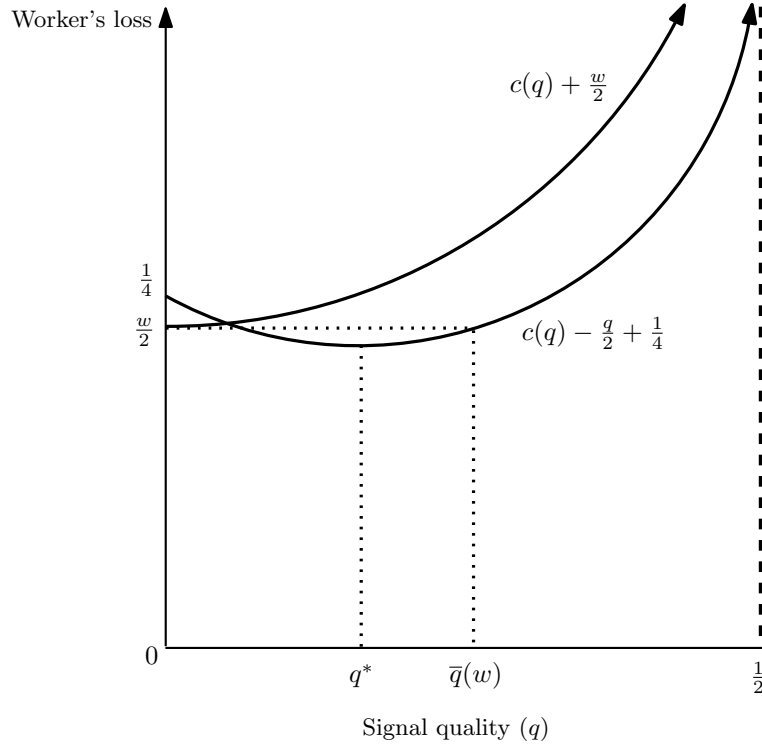
Figure 2.3: Suppose that $w \in \left(2c(q^*) - q^* + \frac{1}{2}, \frac{1}{2}\right)$. In this case, the worker has a mild ex ante preference for α. If the worker believes that the principal will implement the manager's recommendation, then, provided that the manager is not overly biased in favor of α (i.e., as long as $\frac{1}{2} - m \leq \overline{q}(w)$), the worker will find it worthwhile to investigate the alternative by acquiring information.

- *When indifferent among two or more optimal signal qualities, the worker chooses the highest among the ones that are optimal for him.*

- *When indifferent between actions, the manager recommends the one that corresponds to the worker's signal to the principal.*

The quality, $\hat{q}(w, m)$, of the signal that the worker acquires depends upon the manager's type as follows:

Case 1: $w < 2c(q^*) - q^* + \frac{1}{2}$. *In this case,*

$$\hat{q}(w, m) = \begin{cases} q^* & \text{if } m = \frac{1}{2}, \\ 0 & \text{if } 0 < m < \frac{1}{2}. \end{cases}$$
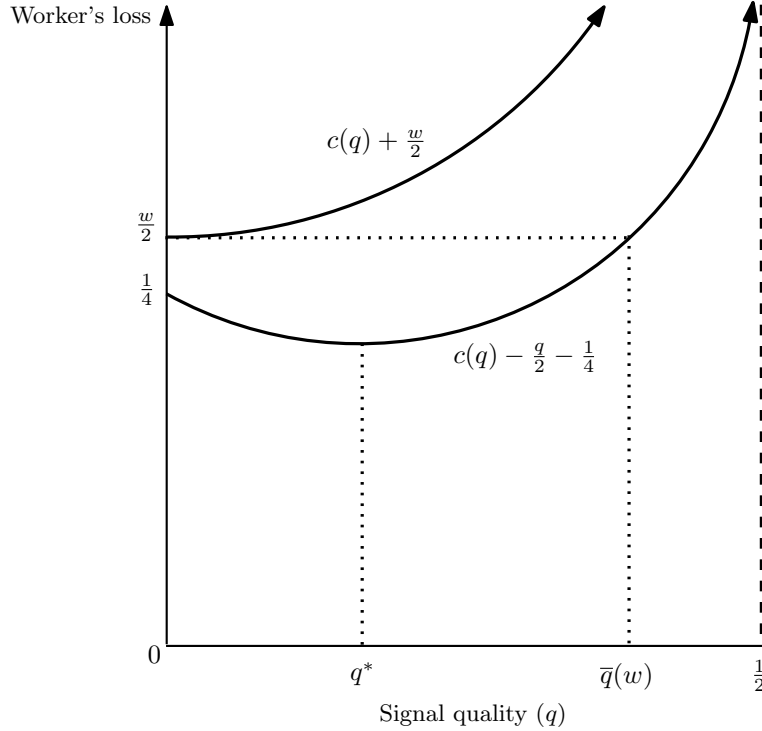
Figure 2.4: Suppose that $w \in \left(\frac{1}{2}, 1\right)$. In this case, the worker holds an ex ante preference for β, which clashes with the manager's ex ante preference for α. If the worker believes that the principal will implement the manager's recommendation, then, just as in the case depicted in Figure 2.3, the worker will acquire information so long as the appointed manager is not overly biased toward α (i.e., provided that $\frac{1}{2} - m \leq \bar{q}(w)$). Note that this type of worker can be induced to acquire better information than the type depicted in Figure 2.3, as this worker's threshold, $\bar{q}(w)$, is higher.

Case 2: $w \geq 2c(q^*) - q^* + \frac{1}{2}$. *In this case,*

$$\hat{q}(w, m) = \begin{cases} 0 & if\ 0 < m < \frac{1}{2} - \bar{q}(w), \\ \frac{1}{2} - m & if\ \frac{1}{2} - \bar{q}(w) \leq m < \frac{1}{2} - q^*, \\ q^* & if\ \frac{1}{2} - q^* \leq m \leq \frac{1}{2}. \end{cases}$$

Proof. The set of optimal signal quality levels for the worker is given by

$$\underset{q \in \left[0, \frac{1}{2}\right)}{\arg \min} \kappa(q; w, m),$$

which can be expressed as

$$\underset{q \in Q_m}{\arg \min} \kappa(q; w, m),$$
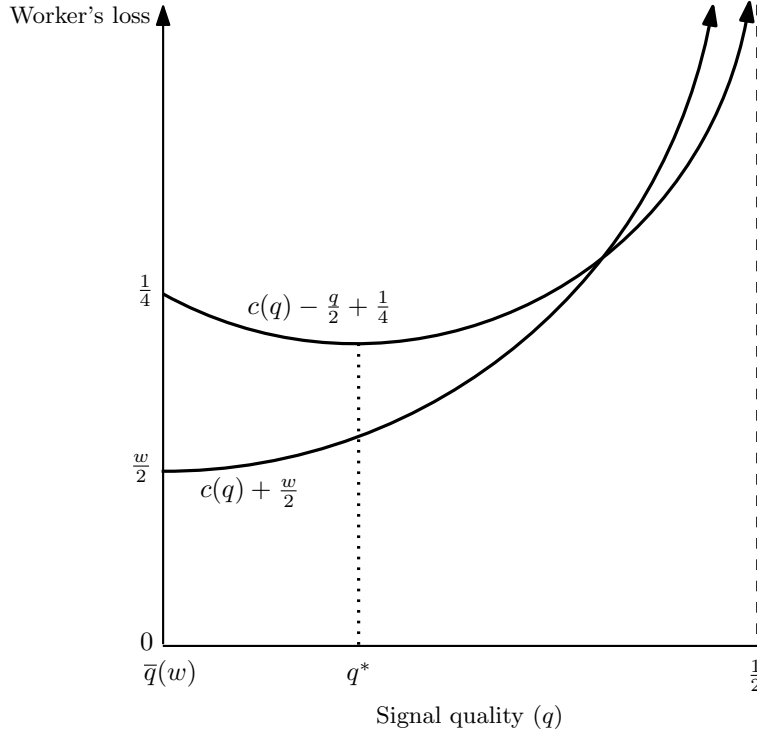
Figure 2.5: Suppose that $w \in \left(0, 2c(q^*) - q^* + \frac{1}{2}\right)$. In this case, the worker holds a fairly strong ex ante bias toward α. If the worker believes that the principal will implement the manager's recommendation, then he realizes that, since the manager favors α ex ante, α will be implemented even if the signal is uninformative. Thus, he does not find it worthwhile to investigate the alternative by acquiring information. (For this worker type, $\overline{q}(w) \equiv 0$.)

where

$$Q_m \equiv \underbrace{\left[\arg\min_{q \in \left[0, \frac{1}{2} - m\right)} c(q) + \frac{w}{2} \right]}_{= \{0\} \text{ if } m < \frac{1}{2}, \varnothing \text{ otherwise}} \cup \underbrace{\left[\arg\min_{q \in \left[\frac{1}{2} - m, \frac{1}{2}\right)} c(q) - \frac{q}{2} + \frac{1}{4} \right]}_{= \max\left\{\frac{1}{2} - m, q^*\right\}}$$

$$= \begin{cases} \left\{0, \frac{1}{2} - m\right\} & \text{if } 0 < m < \frac{1}{2} - q^*, \\ \{0, q^*\} & \text{if } \frac{1}{2} - q^* \le m < \frac{1}{2}, \\ \{q^*\} & \text{if } m = \frac{1}{2}. \end{cases}$$

Thus $\hat{q}\left(w, \frac{1}{2}\right) = q^*$ for all $w \in W$. From this point, suppose that $m < \frac{1}{2}$. Consider the two cases in the statement of the result.

Case 1: $w < 2c(q^*) - q^* + \frac{1}{2}$. In this case, for any $m < \frac{1}{2}$,

$$\underbrace{c(q^*) - \frac{q^*}{2} + \frac{1}{4}}_{=\min_{q \in \left[0, \frac{1}{2}\right)} c(q) - \frac{q}{2} + \frac{1}{4}} > \underbrace{\frac{w}{2}}_{=\kappa(0; w, m)},$$

so

$$\arg\min_{q \in Q_m} \kappa(q; w, m) = \{0\}.$$

Thus $\hat{q}(w, m) = 0$.

Case 2: $w \geq 2c(q^*) - q^* + \frac{1}{2}$. First suppose that $0 < m < \frac{1}{2} - q^*$. In this case, $Q_m = \left\{0, \frac{1}{2} - m\right\}$. Note that $\kappa(0; w, m) = \frac{w}{2}$ and $\kappa\left(\frac{1}{2} - m; w, m\right) = c\left(\frac{1}{2} - m\right) - \frac{1/2 - m}{2} + \frac{1}{4}$. Lemma B.2 implies that

$$\arg\min_{q \in Q_m} \kappa(q; w, m) = \begin{cases} \{0\} & \text{if } 0 < m < \frac{1}{2} - \bar{q}(w), \\ \left\{0, \frac{1}{2} - m\right\} & \text{if } m = \frac{1}{2} - \bar{q}(w), \\ \left\{\frac{1}{2} - m\right\} & \text{if } \frac{1}{2} - \bar{q}(w) < m < \frac{1}{2} - q^*. \end{cases}$$

Now suppose that $\frac{1}{2} - q^* \leq m < \frac{1}{2}$. In this case, $Q_m = \{0, q^*\}$. Note that

$$\underbrace{c(q^*) - \frac{q^*}{2} + \frac{1}{4}}_{=\kappa(q^*; w, m)} \leq \underbrace{\frac{w}{2}}_{=\kappa(0; w, m)},$$

where equality holds if and only if $w = 2c(q^*) - q^* + \frac{1}{2}$. Thus,

$$\arg\min_{q \in Q_m} \kappa(q; w, m) = \begin{cases} \{q^*\} & \text{if } w > 2c(q^*) - q^* + \frac{1}{2} \text{ and } \frac{1}{2} - q^* \leq m, \\ \{0, q^*\} & \text{if } w = 2c(q^*) - q^* + \frac{1}{2} \text{ and } \frac{1}{2} - q^* \leq m. \end{cases}$$

Now the assumption that the worker chooses the highest of optimal signal qualities implies that

$$\hat{q}(w, m) = \begin{cases} 0 & \text{if } 0 < m < \frac{1}{2} - \bar{q}(w), \\ \frac{1}{2} - m & \text{if } \frac{1}{2} - \bar{q}(w) \leq m < \frac{1}{2} - q^*, \\ q^* & \text{if } \frac{1}{2} - q^* \leq m < \frac{1}{2}. \end{cases}$$

$\square$

Lemma 2.3 is useful for characterizing outcomes of informative equilibria, in which the principal follows the worker's recommendation (and the worker anticipates this behavior). The result shows that a worker that is not overly biased toward α can be induced to acquire better information when the appointed manager is moderately biased toward α. The worker

acquires a signal that is just informative enough to offset the manager's bias. An overly
biased manager, though, will destroy the worker's incentives to acquire information: the
quality of signal required to influence the manager's recommendation is too costly.

Using the characterization of Lemma 2.3, it is possible to characterize payoffs in infor-
mative equilibria:

Lemma 2.4. *In any babbling equilibrium, the principal's expected payoff is $-\frac{1}{2}\cdot\min\{p, 1-p\}$.
In an informative equilibrium in which the appointed agents have types $w^* \in W$ and $m^* \in M$,
respectively, the principal's expected payoff is $-\frac{1}{2} \cdot \left(\frac{1}{2} - \hat{q}(w^*, m^*)\right)$. The principal's expected
payoff is at least as high under an informative equilibrium as under a babbling equilibrium.*

Proof. In a babbling equilibrium, the principal chooses her ex ante preferred option. In an
informative equilibrium, the principal rubberstamps the appointed manager's recommen-
dation. By Lemma 2.3, the appointed worker acquires a signal according to $\hat{q}$. Direct
computations yield the payoffs shown above. Finally, from Definition 2.1, it follows that, if a
worker acquires a signal of quality $\hat{q}$ in an informative equilibrium, then $\hat{q} \geq \left|\frac{1}{2} - p\right|$. Hence
$\frac{1}{2} - \hat{q} \leq \min\{p, 1-p\}$, from which it follows that $-\frac{1}{2} \cdot \left(\frac{1}{2} - \hat{q}\right) \geq -\frac{1}{2} \cdot \min\{p, 1-p\}$: the
principal's expected payoff is at least as high in an informative equilibrium as in a babbling
equilibrium.[13]   □

Examples of Informative Equilibria

Example 2.1. Suppose that $p = \frac{1}{2}$, $W = \{\frac{1}{2}\}$, $M = \left(0, \frac{1}{2}\right]$, and $c(q) = \frac{q^2}{1-2q}$. In this case,
$q^* = \frac{2-\sqrt{2}}{4} \approx 0.146$, and $\overline{q}(\frac{1}{2}) = \frac{1}{4}$. Note that the principal is ex ante indifferent between α
and β. Hence, she is willing to follow any recommendation that she believes to be consistent
with the worker's signal. By Lemma 2.3, the strategy profile with the following components
is an informative equilibrium:

- $w^* \equiv \frac{1}{2}$ (by default);

- $m^* \equiv \frac{1}{4}$;

- For all $(w, m) \in W \times M$,

$$\iota^*(w, m) \equiv \begin{cases} 0 & \text{if } 0 < m < \frac{1}{4}, \\ \frac{1}{2} - m & \text{if } \frac{1}{4} \leq m < \frac{\sqrt{2}}{4}, \\ \frac{2-\sqrt{2}}{4} & \text{otherwise};\end{cases}$$

---

[13]Intuitively, the principal always has the option to disregard the recommendation and choose her ex ante
preferred option. In the informative equilibrium, she chooses to follow the recommendation based upon her
belief about the signal's informativeness.

- For all $(w, m, s, q) \in W \times M \times \{a, b\} \times \left[0, \frac{1}{2}\right)$,

$$\rho^*(w, m, s, q) \equiv \begin{cases} \hat{\alpha} & \text{if } s = a \text{ or } q < \frac{1}{2} - m, \\ \hat{\beta} & \text{otherwise;} \end{cases}$$

- For all $(w, m, r) \in W \times M \times \left\{\hat{\alpha}, \hat{\beta}\right\}$,

$$\delta^*(w, m, r) \equiv \begin{cases} \alpha & \text{if } r = \hat{\alpha}, \\ \beta & \text{otherwise.} \end{cases}$$

In this equilibrium, the principal appoints the most biased manager (with preference parameter $\frac{1}{4}$) that the worker finds it worthwhile to convince. Under ι^*, given any appointed manager, the worker acquires either an uninformative signal or a signal that is just precise enough to influence the manager's recommendation. In particular, when the worker and manager follow ι^* and ρ^*, respectively, the manager's recommendation does not contradict an informative signal. Thus, the principal makes a decision that is consistent with the recommendation.

In the equilibrium described in Example 2.1, the principal and worker have aligned interests over the decision, but the principal deliberately appoints a manager whose ex ante preferences differ from hers. Such a manager's bias serves to strengthen the worker's incentive to acquire information. Given the worker's signal, the manager (at least weakly) prefers the same action as the principal, so the principal realizes an unambiguous welfare increase by appointing the biased manager. Example 2.2 shows how the principal can exploit workers' biases to further improve the quality of the decision.

Example 2.2. Suppose that $p = \frac{1}{2}$, $W = \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$, $M = \left(0, \frac{1}{2}\right]$, and $c(q) = \frac{q^2}{1-2q}$. Just as in Example 2.1, $q^* = \frac{2-\sqrt{2}}{4}$, and $\bar{q}(\frac{1}{2}) = \frac{1}{4}$. Furthermore, $2c(q^*) - q^* + \frac{1}{2} = \sqrt{2} - 1 \approx 0.414$, $\bar{q}(\frac{1}{4}) = 0$, and $\bar{q}(\frac{3}{4}) = \frac{1+\sqrt{17}}{16} \approx 0.320$. The strategy profile with the following components is an informative equilibrium:

- $w^* \equiv \frac{3}{4}$;

- $m^* \equiv \frac{1}{2} - \bar{q}(\frac{3}{4}) = \frac{7-\sqrt{17}}{16} \approx 0.180$.

- For all $m \in M$,

$$\iota^*\left(\frac{1}{4}, m\right) \equiv \begin{cases} 0 & \text{if } m < \frac{1}{2}, \\ \frac{2-\sqrt{2}}{4} & \text{otherwise,} \end{cases}$$

$$\iota^*\left(\frac{1}{2}, m\right) \equiv \begin{cases} 0 & \text{if } 0 < m < \frac{1}{4}, \\ \frac{1}{2} - m & \text{if } \frac{1}{4} \leq m < \frac{\sqrt{2}}{4}, \\ \frac{2-\sqrt{2}}{4} & \text{otherwise,} \end{cases}$$

$$\iota^*\left(\frac{3}{4}, m\right) \equiv \begin{cases} 0 & \text{if } 0 < m < \frac{7-\sqrt{17}}{16}, \\ \frac{1}{2} - m & \text{if } \frac{7-\sqrt{17}}{16} \leq m < \frac{\sqrt{2}}{4}, \\ \frac{2-\sqrt{2}}{4} & \text{otherwise;} \end{cases}$$

- For all $(w, m, s, q) \in W \times M \times \{a, b\} \times \left[0, \frac{1}{2}\right)$,

$$\rho^*(w, m, s, q) \equiv \begin{cases} \hat{\alpha} & \text{if } s = a \text{ or } q < \frac{1}{2} - m, \\ \hat{\beta} & \text{otherwise;} \end{cases}$$

- For all $(w, m, r) \in W \times M \times \left\{\hat{\alpha}, \hat{\beta}\right\}$,

$$\delta^*(w, m, r) \equiv \begin{cases} \alpha & \text{if } r = \hat{\alpha}, \\ \beta & \text{otherwise.} \end{cases}$$

The structure of this equiibrium is similar to that of the equilibrium presented in Example 2.1, but, in this case, the principal can exploit the bias of a worker that favors $\hat{\beta}$ ex ante. Such a worker will be willing to acquire a more precise (though costlier) signal to overcome a manager's bias. Thus, the principal achieves a higher expected utility in this case. Note, also, that the worker of type $\frac{1}{4}$ has no incentive to exert any effort (except in the knife-edge case in which the appointed manager is neutral), since, with an uninformative signal, he can ensure that his preferred choice of α is implemented. By Lemma 2.3, he prefers this outcome to any one in which his signal determines the decision (since $\frac{1}{4} < 2c(q^*) - q^* + \frac{1}{2}$).

Main Results

The results of this section generalize the main insights conveyed by the examples of the previous section: a manager with a moderate degree of bias can be helpful for the decision making process, particularly if paired with a worker that holds a clashing bias. Proposition 2.2 shows that, in principle, any pair of worker and manager in which information transmission is "efficient" (in the sense that it benefits the principal and is not too costly for the worker) can be appointed as part of an informative equilibrium.

Proposition 2.2. *Suppose that there exist $w^* \in W$ and $m^* \in M$ such that*

$$\left|p - \frac{1}{2}\right| \leq \frac{1}{2} - m^* \leq \bar{q}(w^*). \tag{2.4}$$

There exists an informative equilibrium in which w^* and m^* are appointed.

Proof. Let w^* and m^* satisfy (2.4). Set the remaining components of the strategy profile as follows:

- For all $(w, m) \in W \times M$,

$$
\iota^*(w, m) \equiv
\begin{cases}
0 & \text{if } w \neq w^* \text{ or } m \neq m^*, \\
\frac{1}{2} - m^* & \text{if } w = w^* \text{ and } m = m^* \text{ and } m^* < \frac{1}{2} - q^*, \\
q^* & \text{otherwise.}
\end{cases}
$$

- For all $(w, m, s, q) \in [(W \times M) \setminus \{(w^*, m^*)\}] \times \{a, b\} \times [0, \frac{1}{2})$,

$$
\rho^*(w, m, s, q) \equiv
\begin{cases}
\hat{\alpha} & \text{if } s = a, \\
\hat{\beta} & \text{otherwise.}
\end{cases}
$$

- For all $(s, q) \in \{a, b\} \times [0, \frac{1}{2})$,

$$
\rho^*(w^*, m^*, s, q) \equiv
\begin{cases}
\hat{\alpha} & \text{if } s = a \text{ or } q < \frac{1}{2} - m, \\
\hat{\beta} & \text{otherwise.}
\end{cases}
$$

- For all $(w, m, r) \in [(W \times M) \setminus \{(w^*, m^*)\}] \times \{\hat{\alpha}, \hat{\beta}\}$,

$$
\delta^*(w, m, r) \equiv
\begin{cases}
\alpha & \text{if } p \leq \frac{1}{2}, \\
\beta & \text{otherwise.}
\end{cases}
$$

- For all $r \in \{\hat{\alpha}, \hat{\beta}\}$,

$$
\delta^*(w^*, m^*, r) \equiv
\begin{cases}
\alpha & \text{if } r = \hat{\alpha}, \\
\beta & \text{otherwise.}
\end{cases}
$$

In words, the players "agree" beforehand on a worker-manager pair, (w^*, m^*), that satisfies the hypothesis. If this pair is appointed, play proceeds according to an informative equilibrium that involves the pair. Otherwise, play proceeds according to a babbling equilibrium. Because $\left| p - \frac{1}{2} \right| \leq \frac{1}{2} - m^*$ by assumption, Lemma 2.1 implies that, when the appointed worker is of type w^* and the appointed manager is of type m^*, the principal and manager agree on the optimal action given the signal. $\square$

The equilibria of Proposition 2.2 arguably require an implausibly high degree of coordination among players, who must agree upon a specific pair, (w^*, m^*). It seems more natural to look instead for equilibria in which the actors play according to informative equilibria whenever it makes sense to do so, as in Examples 2.1 and 2.2. Proposition 2.3 characterizes the best such equilibria from the principal's perspective.

Proposition 2.3. *Suppose that the set*

$$\left\{ m \in M : \exists w \in W \ s.t. \ \left| p - \frac{1}{2} \right| \leq \frac{1}{2} - m \leq \overline{q}(w) \right\}$$

is nonempty, and, furthermore, that it has a minimum. Let $\overline{m}$ be that minimum. There is an informative equilibrium in which $\overline{m}$ is the appointed manager. The principal's payoff in this equilibrium is at least as high as it is in any other equilibrium. Furthermore, if $\frac{1}{2} - \overline{m} > \max\left\{ q^*, \left| p - \frac{1}{2} \right| \right\}$, then any equilibrium in which a different manager is appointed will yield a strictly lower payoff to the principal.

Proof. Let $\overline{m}$ satisfy the given condition, and choose any $w^* \in W$ for which $\overline{q}(w^*) \geq \frac{1}{2} - \overline{m}$. By Proposition 2.2, an informative equilibrium exists in which w^* and $\overline{m}$ are appointed, and the principal's payoff in this equilibrium is no less than her payoff under any babbling equilibrium. However, it is possible to construct a more plausible equilibrium than the one from Proposition 2.2 that is similar to the ones from Examples 2.1 and 2.2. Set

- $m^* \equiv \overline{m}$.

- For all $(w, m) \in W \times M$,

$$\iota^*(w, m) \equiv \begin{cases} \hat{q}(w, m) & \text{if } \hat{q}(w, m) \geq \left| p - \frac{1}{2} \right|, \\ 0 & \text{otherwise.} \end{cases}$$

- For all $(w, m, s, q) \in W \times M \times \{a, b\} \times \left[0, \frac{1}{2} \right)$,

$$\rho^*(w, m, s, q) \equiv \begin{cases} \hat{\alpha} & \text{if } s = a \text{ or } q < \frac{1}{2} - m, \\ \hat{\beta} & \text{otherwise.} \end{cases}$$

- For all $(w, m, r) \in W \times M \times \left\{ \hat{\alpha}, \hat{\beta} \right\}$,

$$\delta^*(w, m, r) \equiv \begin{cases} \alpha & \text{if } \left[\iota^*(w, m) \geq p - \frac{1}{2} \text{ and } r = \hat{\alpha} \right] \text{ or } \iota^*(w, m) < \frac{1}{2} - p, \\ \beta & \text{otherwise.} \end{cases}$$

It is straightforward to verify that the strategy profile $(w^*, m^*, \iota^*, \rho^*, \delta^*)$ constitutes an informative equilibrium. Note, in particular, the following features:

- The principal appoints the most biased candidate manager type, $\overline{m}$, for which some available worker will acquire a sufficiently strong signal to influence the manager's preference. In fact,

$$\hat{q}(w^*, m^*) = \max_{\substack{w \in W, \\ m \in M}} \hat{q}(w, m). \tag{2.5}$$

To see (2.5), observe that, since $m^* \geq \frac{1}{2} - \overline{q}(w^*)$, $\hat{q}(w^*, m^*) = \max\left\{q^*, \frac{1}{2} - m^*\right\}$. By the definition of m^*, given any $m \in M$ with $m < m^*$ and any $w \in W$, $\frac{1}{2} - m > \overline{q}(w)$.[14] Thus $\hat{q}(w, m) = 0 < \hat{q}(w^*, m^*)$ for every (w, m) with $m < m^*$. Now take any $(w, m) \in W \times M$ with $m \geq m^*$. From the characterization of $\hat{q}$ in Lemma 2.3,

$$\hat{q}(w, m) \leq \max\left\{q^*, \frac{1}{2} - m\right\} \leq \max\left\{q^*, \frac{1}{2} - m^*\right\} = \hat{q}(w^*, m^*).$$

- Any appointed worker acquires an informative signal according to $\hat{q}(w, m)$ if and only if this signal quality is high enough to overcome the principal's bias. Otherwise the worker acquires a free and uninformative signal. To see that this behavior is optimal, recall (from Lemma 2.3) that $\hat{q}(w, m)$ gives the worker's preferred signal quality under the belief that the principal will rubberstamp the manager's recommendation. Any appointed manager, under ρ^*, sincerely reveals her preference to the principal given the signal (just as in Lemma 2.3), so the only case in which $\hat{q}(w, m)$ may not be optimal for the worker arises when the principal ignores the manager's recommendation. Given δ^*, this situation occurs only when the principal expects that the signal is too weak to overcome her bias and thus disregards the manager's recommendation—that is, when $\hat{q}(w, m) < \left|p - \frac{1}{2}\right|$. Note that it does not make sense for the worker to acquire a signal of higher quality than $\hat{q}(w, m)$, because this quality (if positive) is strong enough to influence the manager's recommendation, and any higher quality will not be detectable by the principal.[15]

- Any appointed manager observes the signal's realization and quality and sincerely reveals her preferred action to the principal. No manager has an incentive to deviate from this recommendation strategy. If the principal will ignore the recommendation based upon her inference of the signal's informativeness, deviation will have no effect. If the principal will follow the recommendation, then it is in the manager's best interest to report her preference sincerely.

---

[14]Note that $\frac{1}{2} - m > \frac{1}{2} - m^* \geq \left|p - \frac{1}{2}\right|$, so $\frac{1}{2} - m < \left|p - \frac{1}{2}\right|$ is impossible.

[15]It may seem as though the manager could serve to enforce the principal's decision rule. For example, suppose that

$$0 < \frac{1}{2} - m < \left|p - \frac{1}{2}\right| \leq \overline{q}(w).$$

Then it might appear that, as part of an equilibrium, a manager of type m could send a recommendation that corresponds to the signal if $q \geq \left|p - \frac{1}{2}\right|$ and send $\hat{\alpha}$ otherwise. Indeed, a signal of such quality is not prohibitively costly for the worker, and, given such a signal, the manager and principal agree on the optimal action. Note, however, that such a recommendation strategy is not sequentially rational for the manager: if the worker were to deviate and acquire a signal of quality $\frac{1}{2} - m$, then, given that the principal does not observe the signal's quality, the manager can achieve the implementation of her preferred action (given the signal) by still submitting a recommendation that corresponds to the signal. Furthermore, in the absence of direct observability of the signal by the principal, it seems implausible that the principal could write a contract that would commit the manager to such a decision rule.

- The principal anticipates the signal's quality and rubberstamps the manager's recommendation only if her expectation of the signal's quality is sufficiently high to overcome her bias. Otherwise, she ignores the recommendation and chooses her ex ante preferred action.

Combining Lemma 2.4 and (2.5) yields the result that any other equilibrium—whether babbling or informative—will give the principal a payoff that is no greater than her payoff under this equilibrium. Finally, suppose that $\frac{1}{2} - m^* > \max\left\{q^*, \left|p - \frac{1}{2}\right|\right\}$. Then, since $\frac{1}{2} - m^* \leq \bar{q}(w^*)$, $\hat{q}(w^*, m^*) = \frac{1}{2} - m^* > \left|p - \frac{1}{2}\right|$. By Lemma 2.4, the outcome of this informative equilibrium is strictly better for the principal than that of any babbling equilibrium. Consider an informative equilibrium in which a different manager, $m \in M \setminus \{m^*\}$, is appointed. If $m < m^*$, then, as noted in the explanation of (2.5), $\hat{q}(w, m) = 0 < \hat{q}(w^*, m^*)$ for every $w \in W$. If $m > m^*$, then, for any $w \in W$,

$$\hat{q}(w, m) \leq \max\left\{q^*, \frac{1}{2} - m\right\} < \frac{1}{2} - m^* = \hat{q}(w^*, m^*).$$

In either case, the principal's expected payoff—given by Lemma 2.4—is strictly lower in an equilibrium in which $m \neq m^*$ is appointed. $\qquad\square$

In a nutshell, Proposition 2.3 characterizes a "best" equilibrium, in which the most biased manager that does not destroy all potential workers' information acquisition incentives exists. It is important to note a key distinction between managers and workers in this environment. With managers, too much bias can be harmful. As the following results show, however, a worker that is more biased (toward β) is always weakly preferred to one that is less biased.

Proposition 2.4. *Let $w, w' \in W$ satisfy $w < w'$. Let $M_w \subseteq M$ be the set of manager types, m, such that an informative equilibrium in which w and m are appointed exists. Define $M_{w'} \subseteq M$ analogously. $M_w \subseteq M_{w'}$.*

Proof. Fix w and w', and consider any $m \in M_w$. Let $(w, m, \iota, \rho, \delta)$ be an informative equilibrium. Observe that $(w', m, \iota', \rho', \delta')$ is also an informative equilibrium, where

- For every $(\hat{w}, \hat{m}) \in W \times M$,

$$\iota'(\hat{w}, \hat{m}) = \begin{cases} \iota(\hat{w}, \hat{m}) & \text{if } \hat{w} \neq w' \text{ or } \hat{m} \neq m, \\ \hat{q}(w', m) & \text{otherwise.} \end{cases}$$

- For every $(\hat{w}, \hat{m}, s, q) \in W \times M \times \{a, b\} \times \left[0, \frac{1}{2}\right)$,

$$\rho'(\hat{w}, \hat{m}) = \begin{cases} \rho(\hat{w}, \hat{m}) & \text{if } \hat{w} \neq w' \text{ or } \hat{m} \neq m, \\ \hat{\alpha} & \text{if } \hat{w} = w' \text{ and } \hat{m} = m \text{ and } \left[s = a \text{ or } q < \frac{1}{2} - m\right], \\ \hat{\beta} & \text{otherwise.} \end{cases}$$

- For every $(\hat{w}, \hat{m}, r) \in W \times M \times \left\{ \hat{\alpha}, \hat{\beta} \right\}$,

$$
\delta'(\hat{w}, \hat{m}, r) = \begin{cases} \delta(\hat{w}, \hat{m}, r) & \text{if } \hat{w} \neq w' \text{ or } \hat{m} \neq m, \\ \alpha & \text{if } \hat{w} = w' \text{ and } \hat{m} = m \text{ and } r = \hat{\alpha}, \\ \beta & \text{otherwise.} \end{cases}
$$

In words, the strategy profile $(w', m, \iota', \rho', \delta')$ is identical to the original informative equilibrium $(w, m, \iota, \rho, \delta)$, except that, when the pair (w', m) is appointed, this new strategy profile prescribes the worker to acquire a signal of quality $\hat{q}(w', m)$ (which, by Lemmas 2.3 and B.3, is at least as high as in the original equilibrium), the manager to recommend her preferred choice to the principal, and the principal to rubberstamp the manager's recommendation. Since information acquisition is weakly better in this strategy profile than in the original informative equilibrium, the manager finds it optimal to base her recommendation on the signal, and the principal is happy to rubberstamp the manager's recommendation. Also, note that play off the equilibrium path remains unchanged; its optimality in the new strategy profile follows immediately from the assumption that the original strategy profile is an equilibrium. Since $(w', m, \iota', \rho', \delta')$ is an informative equilibrium, $m \in M_{w'}$. $\qquad \square$

Proposition 2.4 shows that the principal loses nothing from the appointment of a worker that is more biased toward β: if anything, a more biased worker will acquire better information than a less biased one, and all equilibrium outcomes that are implementable with a less biased worker remain implementable with a more biased one. The final result, Corollary 2.1, demonstrates that, provided that the pool of candidate manager types is sufficiently rich, the principal can achieve a strict gain by appointing a more biased worker.

Corollary 2.1. *Let $w, w' \in W$ satisfy $w < w'$. Let V_w be the supremum of the set of the principal's payoffs under all equilibria in which w is the appointed worker. Define $V_{w'}$ analogously. $V_w \leq V_{w'}$, and the inequality is strict if there exists a manager type, $m \in M$, such that $\max \left\{ q^*, \bar{q}(w) \right\} < \frac{1}{2} - m \leq \bar{q}(w')$.*

Proof. The fact that $V_w \leq V_{w'}$ follows immediately from Proposition 2.4. Suppose there exists a manager type, $m \in M$ that satisfies $\max \left\{ q^*, \bar{q}(w) \right\} < \frac{1}{2} - m \leq \bar{q}(w')$. Then, by Lemmas 2.3 and 2.4,

$$
\begin{aligned}
V_{w'} &\geq -\frac{1}{2} \cdot \left(\frac{1}{2} - \hat{q}(w', m) \right) \\
&= -\frac{1}{2} \cdot m \\
&> -\frac{1}{2} \cdot \left(\frac{1}{2} - \max \left\{ q^*, \bar{q}(w) \right\} \right) \\
&\geq V_w.
\end{aligned}
$$

$\qquad \square$

2.4 Discussion

This paper highlights an important benefit of discord in organizations: improved information acquisition incentives, which lead to better decision making. The results are similar to those obtained in Chapter 1, but the environment—particularly the role of the middle manager—is quite different here. Other works that look at the benefits of discord for information acquisition and revelation are due to Rotemberg and Saloner (1995); Shin (1998); Dewatripont and Tirole (1999); and Van den Steen (2010a).

The paper makes two stark assumptions that may be relaxed in future work. The first assumption is that information acquisition is of a symmetric nature: the "quality" of the signal that the worker chooses determines the signal's informativeness in either state of the world. In practice, though, information acquisition can be asymmetric: an actor that is ex ante biased toward a certain action may acquire a signal that is extremely informative in the state of the world in which that action is optimal but is much less informative in the other state of the world.[16] In this case, and in contrast to the results of this paper, the principal may not want to appoint an extremely biased worker. Another stark assumption is that all actors' types are common knowledge. Learning about types and abilities is a salient feature of the employment relationship. Incorporating these features into a dynamic version of the model will introduce some reputational incentives and is likely to produce additional important insights.

---

[16]Kamenica and Gentzkow (2011) characterize the optimal signal for an expert in such an environment.

Chapter 3

Efficient Electorates

3.1 Introduction

Consider the problem faced by a social planner that seeks to appoint an electorate in a social choice setting with pure common values, private noisy information about an unobservable payoff-relevant state of the world, and costless voting. An economic argument in favor of direct democracy in such a setting emphasizes the regime's efficiency in aggregating information: if each citizen votes for an alternative according to her private information—which, on average, is correct—then the probability that the welfare-maximizing outcome is implemented in a majority-rule election approaches one as the number of citizens is allowed to grow arbitrarily large.[1] This argument relies on at least two crucial, but implicit, assumptions.

The first assumption is that voters are sincere in expressing their preferences via ballots. As Austen-Smith and Banks (1996) show, sincere voting does not generally constitute an equilibrium of a voting game. As a behavioral assumption, though, especially for large electorates, sincere voting seems entirely plausible.[2] The second assumption, which is less plausible (and which I relax in this paper) is that the sources of voters' private information are sufficiently uncorrelated.[3] In most real-world settings, correlation is likely to emerge for at least two reasons. First, voters share information sources. Second, voters are likely to exchange information with each other prior to an election. In either case, redundancies exist among voters' information sets. As I show, such redundancies, under sincere and costless voting, can render direct democracy suboptimal.

---

[1] This intuition was first formalized by the Marquis de Condorcet in his celebrated "Jury Theorem" (Ladha, 1992). See Ladha (1992); Feddersen and Pesendorfer (1997); McLennan (1998); and Feddersen and Pesendorfer (1999) for examples of extensions.

[2] As Feddersen (2004) notes, however, some empirical studies on voting behavior lend support to strategic voting models. Some experimental studies (Ali, Goeree, Kartik, and Palfrey, 2008; Battaglini, Morton, and Palfrey, 2008) have yielded similar results. Degan and Merlo (2009) derive conditions under which the assumption of sincere voting is falsifiable; in particular, Degan and Merlo show that sincere voting cannot be rejected based on data from only one election.

[3] The original Condorcet Jury Theorem assumes independence of the private signals. Ladha (1992) extends the result to cases in which the signals are weakly correlated.

It is worth noting that direct democracies are rarely observed in practice. The most obvious (and arguably leading) explanation for this empirical regularity is that, even if private voting costs are negligible, the aggregate social cost of conducting an election is likely to be increasing in the electorate's size and could be astronomical for large electorates. In particular, if shrinking the size of the electorate achieves substantial cost savings while meeting acceptable standards for information aggregation, then direct democracy will not be socially optimal. As this paper illustrates, however, even if the social planner is unwilling to compromise on the issue of information aggregation, direct democracy can be suboptimal. Adding an incentive to economize on election costs—which increase in the size of the electorate—will only strengthen this result.

I also show that it is not always possible for the planner to appoint an electorate that will aggregate information perfectly. For various information allocation structures, I characterize the optimal electorate. As I illustrate with examples, the structure of the optimal electorate is unstable: slight perturbations to the information allocation can change the optimal electorate significantly.

Section 3.2 provides an overview of the existing literature. Section 3.3 develops a general framework that can also be used for future work on related problems. Section 3.4 imposes additional structure on the general framework (to allow a focus on the electorate formation problem) and presents a few results. Section 3.5 applies these results to a specific environment of voting in social networks and discusses a number of illustrative examples. Section 3.6 highlights some directions for future research and concludes.

3.2 Related Literature

Perhaps the earliest and most well known result on information aggregation in elections is the Condorcet Jury Theorem:

Theorem 3.1 (Jury Theorem, Condorcet (1785), as stated in Ladha (1992))**.** *Suppose that the votes of the members of an assembly are independent, and suppose that each member of the assembly votes for the "correct" alternative with a probability of $p > \frac{1}{2}$. Then the probability that the correct alternative is chosen by strictly more than half of the members of the assembly is strictly higher than p and approaches one as the size of the assembly approaches infinity.*

This result has been generalized (e.g., by Ladha (1992), who shows that the result continues to hold with correlated votes, as long as the degree of correlation is sufficiently low), but, as Austen-Smith and Banks (1996) observe, the original theorem and many of its generalizations assume that each agent votes sincerely (i.e., as though she is a dictator). Austen-Smith and Banks (1996) demonstrate that sincere voting does not generally constitute an equilibrium. McLennan (1998), however, shows that, in games in which the Condorcet Jury Theorem holds for sincere voting strategy profiles, there also exist equilibrium strategy profiles for which the result holds. Coughlan (2000) adds communication (in particular, a

nonbinding "straw vote") to the jury setting and shows that sincere voting strategies can, under fairly general conditions, be part of subgame perfect equilibria. The jury setting, however, does not allow abstention. In voting games with abstention, Feddersen and Pesendorfer (1996) identify the so-called "swing voter's curse," in which uninformed voters may prefer to abstain—or even vote against the preferences that are derived from their private information—given that their votes matter only when they are pivotal, and, in particular, when they oppose informed voters' votes.[4] Krishna and Morgan (2012) show that, when voting is costless, voluntary voting schemes outperform compulsory voting schemes in terms of expected aggregate welfare. This insight suggests that, when voters are not sophisticated enough to coordinate their strategies and make inferences about others' information, a social planner that knows something about the underlying information structure might want to restrict voting rights to members of a proper subset of the population.

This paper also contributes to the literature on committees.[5] The idea that voting rules that favor a proper subset of the citizenry can maximize social welfare is demonstrated by Chwe (1999). The key assumptions that drive Chwe's results are substantively different from the ones that I make, though: Chwe considers an environment with individuals that have different priors and that do not communicate with each other. The idea is that electoral rules that favor minorities by allowing them to decide the outcome can induce greater participation by minorities, hence improving information aggregation. In my environment, all individuals are ex ante identical except for their access to information sources. I treat the electoral rule as exogenous and look at how the planner should restrict participation in the election so as to effectively aggregate information.

Koriyama and Szentes (2009) study a committee formation problem in a pure common values environment with uncertainty and costly information acquisition. Koriyama and Szentes find that, while the optimal committee size is bounded, the inefficiencies generated by oversized committees are small. Gershkov and Szentes (2009) consider a similar environment and characterize a mechanism for eliciting and aggregating information from agents so that ex ante social welfare is maximized. Information acquisition is not part of the model considered in this paper, though. Thus, the results are more appropriate for settings in which agents are endowed with information (or receive it costlessly).

3.3   A General Model

There is a social planner; a finite set, $N = \{1, \ldots, n\}$, of citizens; and a collection, $\mathcal{E} \subseteq 2^N \setminus \{\varnothing\}$, of feasible electorates.[6] The social planner is responsible for choosing an electorate, $E \in \mathcal{E}$. Members of E are called *electors*.

---

[4]This phenomenon, which is analogous to the "winner's curse" in common-value auctions, is fairly robust. For example, see Feddersen and Pesendorfer (1999) and McMurray (2013).

[5]A useful survey piece on this area is due to Gerling, Grüner, Kiel, and Schulte (2005).

[6]The notion of feasibility is used to capture institutional constraints (e.g., regional representation requirements for legislatures).

The state of the world, ω, is a random element that takes values in the measurable space $(\Omega, \mathcal{F})$ and has distribution π. No one observes the state of the world, but a k-dimensional random vector,

$$\hat{\boldsymbol{\omega}} \equiv \begin{pmatrix} \hat{\omega}_1 \\ \vdots \\ \hat{\omega}_k \end{pmatrix},$$

of signals is realized. Each signal is a random element that takes values in the measurable space $(\hat{\Omega}, \mathcal{G})$. The signals are conditionally independent and identically distributed according to $\mu(\cdot|\omega)$.[7] Each citizen $j \in N$ observes at least one of the signals in $\hat{\boldsymbol{\omega}}$. The allocation of information is exogenous and can be summarized by an $n \times k$ 0-1 matrix

$$M \equiv \begin{bmatrix} m_{1,1} & \cdots & m_{1,k} \\ \vdots & \ddots & \vdots \\ m_{n,1} & \cdots & m_{n,k} \end{bmatrix},$$

where, for each $j \in N$ and $i \in K \equiv \{1, \ldots, k\}$,

$$m_{j,i} = \begin{cases} 1 & \text{if Citizen } j \text{ observes } \hat{\omega}_i, \\ 0 & \text{otherwise.} \end{cases}$$

I refer to M as the *information allocation matrix*. For a citizen $j \in N$, denote the citizen's set of observed signals by $K_j \equiv \{i \in K : m_{j,i} = 1\}$. For an electorate E, denote the set of indices of signals that are observed by at least one member of the electorate by $K_E \equiv \bigcup_{j \in E} K_j$. For any nonempty proper subset $K' \subset K$, let $\hat{\boldsymbol{\omega}}_{K'} \equiv \left((\hat{\omega}_i)_{i \in K'} \right)^\top$ be the $|K'|$-dimensional vector of signals with indices in K'.

Let X be a finite set of alternatives, and let ΔX represent the set of probability distributions on X. Each elector can either vote (i.e., choose an alternative in X) or abstain: let $B \equiv X \cup \{\text{abstain}\}$ represent this action space, to which I refer as the "ballot space." An element of B is called a "ballot." Given a tuple of electors' ballots (called a "ballot profile"), an alternative in X is chosen according to an electoral rule R:

Definition 3.1. An *electoral rule* is a function $R : \bigcup_{E \in \mathcal{E}} B^E \to \Delta X$ that satisfies the following conditions:

---

[7]To be precise, for all distinct $i_1, \ldots, i_m \in K \equiv \{1, \ldots, k\}$ and for all $A_1, \ldots, A_m \in \mathcal{G}$,

$$\Pr\left(\bigcap_{\ell=1}^m \{\hat{\omega}_{i_\ell} \in A_\ell\} \,\middle|\, \omega \right) = \prod_{\ell=1}^m \mu\left(A_\ell | \omega \right).$$

Symmetry: $R(b) = R(b \circ \rho)$ for all $E \in \mathcal{E}$, $b \in B^E$, and $\rho \in \Sigma(E)$, where $\Sigma(E)$ denotes the set of permutations on E.

Monotonicity: For all $x \in X, E \in \mathcal{E}$, and $b, b' \in B^E$,

$$R(b)(\{x\}) \geq R(b')(\{x\}) \text{ whenever}$$
$$\left(\frac{|\{j \in E : b(j) = x\}|}{|\{j \in E : b(j) \neq \text{abstain}\}|} > \frac{|\{j \in E : b'(j) = x\}|}{|\{j \in E : b'(j) \neq \text{abstain}\}|} \text{ and} \right.$$
$$\left. \left[\frac{|\{j \in E : b(j) = y\}|}{|\{j \in E : b(j) \neq \text{abstain}\}|} \leq \frac{|\{j \in E : b'(j) = y\}|}{|\{j \in E : b'(j) \neq \text{abstain}\}|} \text{ for all } y \in X \setminus \{x\} \right] \right).$$

Given a ballot profile, an electoral rule determines a probability distribution over alternatives. The fact that an electoral rule maps a ballot profile to a probability distribution over alternatives, rather than directly to an alternative, allows random tie-breaking procedures. The property of symmetry captures the permutability (and hence the anonymity) of the ballots. The monotonicity property says, in words, that a strict increase in the fraction of votes for exactly one alternative weakly increases the likelihood that that alternative is implemented. This condition is weak and very natural.

Citizen j has a preference relation $\succsim_j$ over $\Delta(X \times \Omega)$. The social planner has a preference relation $\succsim_P$ over $\mathcal{E} \times \Delta(X \times \Omega)$.

This discussion describes an extensive form game of incomplete information, the timing of which is shown in Figure 3.1.
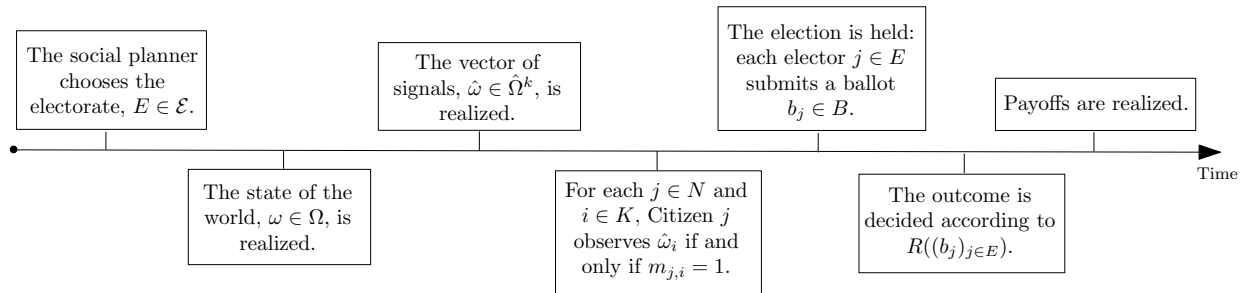
Figure 3.1: The timing of the electorate formation problem

Definition 3.2. A *ballot strategy* for citizen $j \in N$ is a function $\sigma_j : \hat{\Omega}^{K_j} \times \mathcal{E} \to \Delta B$.

A pure ballot strategy describes a citizen's ballot, if she is chosen as an elector, as a function of her observed signal and of the social planner's choice of electorate. Note that a citizen does not actually submit a ballot if she is not chosen as an elector.

Definition 3.3. A *ballot strategy profile* is an ordered n-tuple $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ of ballot strategies.

For any information allocation matrix M, a probability distribution on the product space $(X \times \Omega, 2^X \otimes \mathcal{F})$ is induced by an electorate E, an electoral rule R, and a ballot strategy profile $\boldsymbol{\sigma}$.[8] Let $\mathcal{L}(M, E, R, \boldsymbol{\sigma})$ denote this probability distribution.

Definition 3.4. Let E be a feasible electorate, R be an electoral rule, and $\boldsymbol{\sigma}$ be a ballot strategy profile. *E solves the planner's problem under M, R, and $\boldsymbol{\sigma}$* if

$$(E, \mathcal{L}(M, E, R, \boldsymbol{\sigma})) \succsim_{\mathrm{P}} (E', \mathcal{L}(M, E', R, \boldsymbol{\sigma}))$$

for all feasible electorates E'.

Definition 3.5. Let E be a feasible electorate, R be an electoral rule, $\boldsymbol{\sigma}$ be a ballot strategy profile, and j be a citizen. σ_j *solves j's problem under M, E, R, and $\boldsymbol{\sigma}_{-j}$* if

$$\mathcal{L}(M, E, R, \boldsymbol{\sigma}) \succsim_j \mathcal{L}(M, E, R, (\boldsymbol{\sigma}_{-j}, \nu_j))$$

for each ballot strategy ν_j for citizen j.

3.4 The Electorate Formation Problem

Based on the discussion of Section 3.3, the electorate formation problem can be framed as an extensive form Bayesian game, and appropriate solution concepts can be employed to characterize outcomes. This paper, however, focuses on the solution to the planner's electorate formation problem. The objective is to understand the effect of the allocation of information in determining this solution. Here I introduce some assumptions that help to simplify matters. Throughout this section, I use the following notation: given a set Y and an element $y \in Y$, the (degenerate) probability distribution that assigns probability one to y is denoted by δ_y.

Assumption 3.1. There are two states of the world: $(\Omega, \mathcal{F}) = (\{-1, 1\}, 2^{\{-1,1\}})$.

Assumption 3.2 (Uniform priors). Voters' prior beliefs are given by $\pi(\{-1\}) = \pi(\{1\}) = \frac{1}{2}$.

Assumption 3.3. There are two alternatives: $X = \{-1, 1\}$.

Assumption 3.4. The action of abstention is represented by 0: $B = \{-1, 0, 1\}$.

---

[8]In particular, the state of the world is realized according to π, which is a probability distribution on $(\Omega, \mathcal{F})$. Conditional on the realization of the state of the world, signals are realized according to the probability distribution μ. The conditional probability distribution over the space of signal vectors and the information allocation matrix together determine, for each citizen, a conditional probability distribution over sets of observed signals. These conditional probability distributions over sets of observed signals, together with the electorate and ballot strategy profile, determine a conditional probability distribution over ballot tuples. The conditional probability distribution over ballot tuples and the electoral rule jointly determine a conditional probability distribution on $(X, 2^X)$. The product of this conditional probability distribution and π constitutes a probability distribution on $(X \times \Omega, 2^X \otimes \mathcal{F})$.

Assumption 3.5 (Simple majority rule). The electoral rule is R_{SM}, which, in this environment, may be characterized by the condition

$$R_{\mathrm{SM}}\left((b(j))_{j \in E}\right) = \begin{cases} \delta_{\mathrm{sgn}(\sum_{j \in E} b(j))} & \text{if } \sum_{j \in E} b(j) \neq 0, \\ \left(\frac{1}{2}, \frac{1}{2}\right) & \text{otherwise.} \end{cases}$$

Assumption 3.6. There are two signals: $\left(\hat{\Omega}, \mathcal{G}\right) = \left(\{-1, 1\}, 2^{\{-1,1\}}\right)$. The conditional distribution, μ, can be summarized by the condition $\mu(\{s\}|s) = p$, where $p \in \left(\frac{1}{2}, 1\right)$, for each $s \in \{-1, 1\}$.

Assumption 3.7 (Pure common values with symmetry). For each citizen $j \in N$, the preference relation $\succsim_j$ on $\Delta(X \times \Omega)$ has a von Neumann-Morgenstern utility representation with the following utility index:

$$u(x, \omega) = \begin{cases} 1 & \text{if } x = \omega, \\ 0 & \text{otherwise.} \end{cases}$$

Given the commonality of the citizens' preferences, from now on I let $\succsim_{\mathrm{C}}$ represent the citizens' preference relation. Under Assumption 3.7, $\succsim_{\mathrm{C}}$ has a utility representation, $U : \Delta(X \times \Omega) \to \mathbb{R}$, given by

$$U(\eta) \equiv \eta(\{(-1, -1), (1, 1)\}).$$

Assumption 3.8 (Benevolent but parsimonious social planner). The social planner has lexicographic preferences over $\mathcal{E} \times \Delta(X \times \Omega)$: given $E, E' \in \mathcal{E}$ and $\eta, \eta' \in \Delta(X \times \Omega)$,

$$(E, \eta) \succsim_{\mathrm{P}} (E', \eta') \text{ if and only if } \eta \succ_{\mathrm{C}} \eta' \text{ or } (\eta \sim_{\mathrm{C}} \eta' \text{ and } |E| \leq |E'|).$$

In words, the social planner is concerned foremost with the citizens' welfare. If citizens are indifferent between two potential outcomes, the social planner prefers the outcome that involves the smaller electorate. The interpretation of this feature is that the social cost of appointing and polling an electorate is increasing in the size of the electorate, and, while the social planner is unwilling to compromise on the issue of citizens' welfare, if that issue becomes irrelevant (i.e., if the citizens are indifferent), then the social planner is concerned with minimizing social costs.

Assumption 3.9 (No substantive restrictions on electorates). $\mathcal{E} = 2^N \setminus \{\varnothing\}$: all nonempty subsets of N are feasible electorates.

Definition 3.6. The *trivial electorate* is N.

When the social planner chooses the trivial electorate, she institutes a direct democracy. When the social planner chooses a singleton as the electorate, she institutes a dictatorship. When the social planner chooses any other feasible electorate, she institutes a representative democracy.

Suppose that Elector $j \in E$ observes the signals $\hat{\boldsymbol{\omega}}_{K_j} = \hat{\mathbf{s}}_{K_j} \equiv (\hat{s}_i)_{i \in K_j}$. Then Elector j's updated beliefs regarding the state of the world are derived by Bayes' Rule and are summarized by the probability mass function

$$
f_{\omega|\hat{\boldsymbol{\omega}}_{K_j}}\left(s\,\big|\hat{\mathbf{s}}_{K_j}\right) = \begin{cases} \frac{1}{1+\left(\frac{p}{1-p}\right)^{\sum_{i \in K_j} \hat{s}_i}} & \text{if } s = -1, \\ \frac{1}{1+\left(\frac{1-p}{p}\right)^{\sum_{i \in K_j} \hat{s}_i}} & \text{if } s = 1, \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}
$$

A similar application of Bayes' Rule yields the probability mass function that describes the posterior distribution over Ω given a k-tuple of signals $\hat{\mathbf{s}}$:

$$
f_{\omega|\hat{\boldsymbol{\omega}}}\left(s\,|\hat{\mathbf{s}}\right) = \begin{cases} \frac{1}{1+\left(\frac{p}{1-p}\right)^{\sum_{i=1}^{k} \hat{s}_i}} & \text{if } s = -1, \\ \frac{1}{1+\left(\frac{1-p}{p}\right)^{\sum_{i=1}^{k} \hat{s}_i}} & \text{if } s = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}
$$

Another probability mass function that will be useful is the one over $\hat{\Omega}^k$:

$$
f_{\hat{\boldsymbol{\omega}}}\left(\hat{\mathbf{s}}\right) = \frac{1}{2}\left[p^{\frac{1}{2}\left(k - \sum_{i=1}^{k} \hat{s}_i\right)}(1-p)^{\frac{1}{2}\left(k + \sum_{i=1}^{k} \hat{s}_i\right)} + (1-p)^{\frac{1}{2}\left(k - \sum_{i=1}^{k} \hat{s}_i\right)}p^{\frac{1}{2}\left(k + \sum_{i=1}^{k} \hat{s}_i\right)}\right]. \tag{3.3}
$$

Definition 3.7. A citizen $j \in N$ is *naïve* if her ballot strategy, σ_j, satisfies the following conditions:

(i) $\sigma_j\left(\hat{\mathbf{s}}_{K_j}, E\right) = \delta_{-1}$ whenever $f_{\omega|\hat{\boldsymbol{\omega}}_{K_j}}\left(1\,\big|\hat{\mathbf{s}}_{K_j}\right) < \frac{1}{2}$;

(ii) $\sigma_j\left(\hat{\mathbf{s}}_{K_j}, E\right) = \delta_1$ whenever $f_{\omega|\hat{\boldsymbol{\omega}}_{K_j}}\left(1\,\big|\hat{\mathbf{s}}_{K_j}\right) > \frac{1}{2}$;

(iii) $\sigma_j\left(\hat{\mathbf{s}}_{K_j}, E\right) = \delta_0$ whenever $f_{\omega|\hat{\boldsymbol{\omega}}_{K_j}}\left(1\,\big|\hat{\mathbf{s}}_{K_j}\right) = \frac{1}{2}$.

A naïve citizen, if chosen as an elector, votes for the outcome that maximizes her expected utility based on her posterior beliefs (as described in (3.1)) about the state of the world. If both outcomes yield her the same expected utility, she abstains.[9] Hence, a naïve elector's ballot is the outcome of a counting problem: if the number of -1 signals that she observes is strictly greater than the number of 1 signals that she observes, she votes -1. If she observes an equal number of each signal, she abstains. If neither condition holds, she votes 1. A naïve citizen j's ballot strategy $\bar{\sigma}_j$ can be described succinctly as

$$
\bar{\sigma}_j\left(\hat{\mathbf{s}}_{K_j}, E\right) = \text{sgn}\left(\sum_{i \in K_j} \hat{s}_i\right)
$$

---

[9]In principle, if both outcomes yield a naïve citizen the same expected utility (based on her signal and on her beliefs about her neighbors' signals), she should be indifferent—and, in particular, willing to randomize—among all three ballots. I impose the condition that she abstains in such a case primarily for convenience. As it turns out, the main insights do not depend on this assumption.

for each electorate E.

From now on, unless stated explicitly otherwise, I will maintain naïveté as a behavioral assumption:

Assumption 3.10 (Behavioral assumption). Citizens are naïve.

By fixing citizens' behavior, Assumption 3.10 allows me to ignore the issue of citizens' strategic considerations and, in particular, to focus exclusively on the issue of primary interest: the effect of the allocation of information on the solution to the planner's problem. The pressing question can be posed as follows: "When citizens are naïve, which feasible electorate solves the planner's problem?"[10] To answer this question, it will be helpful to introduce some additional terminology.

First, let $\bar{\boldsymbol{\sigma}}$ denote the ballot strategy profile used by naïve citizens.

Definition 3.8. Let E be a feasible electorate and $\hat{\mathbf{s}}$ be a k-tuple of signals. The *election outcome under E and $\hat{\mathbf{s}}$*, denoted by $\mathcal{O}(E, \hat{\mathbf{s}})$, is the probability distribution in ΔX that is defined by

$$\mathcal{O}(E, \hat{\mathbf{s}}) \equiv R_{\text{SM}}\left(\left(\bar{\sigma}_j\left(\hat{\boldsymbol{s}}_{K_j}, E\right)\right)_{j \in E}\right)$$

$$= \begin{cases} \delta_{\text{sgn}\left(\sum_{j \in E} \text{sgn}\left(\sum_{i \in K_j} \hat{s}_i\right)\right)} & \text{if } \sum_{j \in E} \text{sgn}\left(\sum_{i \in K_j} \hat{s}_i\right) \neq 0, \\ \left(\frac{1}{2}, \frac{1}{2}\right) & \text{otherwise.} \end{cases}$$

In words, the election outcome under a feasible electorate E and n-tuple of signals $\hat{\mathbf{s}}$ is the probability distribution over alternatives that results when the realized signals are $\hat{\mathbf{s}}$ and an election takes place among members of E. Hence $\mathcal{O}(E, \hat{\mathbf{s}})(\{x\})$ gives the probability that alternative x is implemented following the election among members of E, given the signals $\hat{\mathbf{s}}$.

Definition 3.9. Let M be an information allocation matrix, E be a feasible electorate and $\hat{\mathbf{s}}$ be an n-tuple of signals. E is *efficient at $\hat{\mathbf{s}}$* if

$$\mathcal{O}(M, E, \hat{\mathbf{s}}) = \begin{cases} \delta_{-1} & \text{if } f_{\omega|\hat{\omega}}\left(-1 \,|\hat{\mathbf{s}}\right) > \frac{1}{2}, \\ \delta_1 & \text{if } f_{\omega|\hat{\omega}}\left(1 \,|\hat{\mathbf{s}}\right) > \frac{1}{2}. \end{cases} \tag{3.4}$$

That is, the electorate E is efficient at $\hat{\mathbf{s}}$ if, whenever one state of the world is strictly more likely than the other given the realized signal vector $\hat{\mathbf{s}}$, the alternative that corresponds to the more likely state of the world wins outright in the election involving the members of E.[11]

---

[10]An obvious critique of the formulation of this question is that naïveté need not constitute equilibrium play on the part of citizens. In a sense, I am imposing a form of bounded rationality. However, it will become clear shortly (in Remark 3.1) that, under the stated assumptions, naïveté is not entirely inconsistent with equilibrium concepts. It should be noted, though, that naïveté do not meet certain desirable equilibrium refinements, such as perfection criteria.

[11]Note that, when the signals in $\hat{\mathbf{s}}$ are evenly split between -1 and 1 (i.e., when $f_{\omega|\hat{\omega}}\left(1 \,|\hat{\mathbf{s}}\right) = \frac{1}{2}$), every electorate is efficient at $\hat{\mathbf{s}}$.

Definition 3.10. Let E be a feasible electorate. E is *efficient* if E is efficient at every $\hat{\mathbf{s}} \in \hat{\Omega}^n$.

In words, E is efficient if, at every k-tuple of signals under which the posterior distribution over states is not uniform, the alternative that corresponds to the more likely state of the world wins outright in the majority rule election involving the members of E. Note that the condition for efficiency can be stated succinctly as

$$\operatorname{sgn}\left(\sum_{j \in E} \operatorname{sgn}\left(\sum_{i \in K_j} \hat{s}_i\right)\right) = \operatorname{sgn}\left(\sum_{i=1}^k \hat{s}_i\right) \text{ whenever } \sum_{i=1}^k \hat{s}_i \neq 0.$$

Now it is possible to compare different electorates. Define a function $\mathcal{U} : \mathcal{E} \to \mathbb{R}$ as

$$\mathcal{U}(E) \equiv U\left(\mathcal{L}\left(E, R_{\mathrm{SM}}, \bar{\boldsymbol{\sigma}}\right)\right)$$
$$= \sum_{\hat{\mathbf{s}} \in \{-1,1\}^k} f_{\hat{\boldsymbol{\omega}}}\left(\hat{\mathbf{s}}\right) \cdot \left[f_{\omega|\hat{\boldsymbol{\omega}}}\left(-1 | \hat{\mathbf{s}}\right) \mathcal{O}\left(E, \hat{\mathbf{s}}\right)\left(\{-1\}\right) + f_{\omega|\hat{\boldsymbol{\omega}}}\left(1 | \hat{\mathbf{s}}\right) \mathcal{O}\left(E, \hat{\mathbf{s}}\right)\left(\{1\}\right)\right]. \quad (3.5)$$

In words, $\mathcal{U}(E)$ is the expected utility to each citizen when E is chosen as the electorate. Since U is a representation of $\succsim_C$, the solution to the planner's problem under simple majority rule and the naïve ballot strategy profile can be formulated as

$$\min_{E \in \mathcal{M}} |E|,$$

where

$$\mathcal{M} \equiv \arg\max_{E' \in \mathcal{E}} \mathcal{U}\left(E'\right).$$

Consider the bracketed expression in (3.5):

$$f_{\omega|\hat{\boldsymbol{\omega}}}\left(-1 | \hat{\mathbf{s}}\right) \mathcal{O}\left(E, \hat{\mathbf{s}}\right)\left(\{-1\}\right) + f_{\omega|\hat{\boldsymbol{\omega}}}\left(1 | \hat{\mathbf{s}}\right) \mathcal{O}\left(E, \hat{\mathbf{s}}\right)\left(\{1\}\right). \quad (3.6)$$

(3.6) is a convex combination between $f_{\omega|\hat{\boldsymbol{\omega}}}\left(-1 | \hat{\mathbf{s}}\right)$ and $f_{\omega|\hat{\boldsymbol{\omega}}}\left(1 | \hat{\mathbf{s}}\right)$. Note that (3.6) evaluates to $\max\left\{f_{\omega|\hat{\boldsymbol{\omega}}}\left(-1 | \hat{\mathbf{s}}\right), f_{\omega|\hat{\boldsymbol{\omega}}}\left(1 | \hat{\mathbf{s}}\right)\right\}$ if and only if E is efficient at $\hat{\mathbf{s}}$. In particular, the bracketed expression in (3.5) is maximized for all $\hat{\mathbf{s}} \in \hat{\Omega}^n$ if and only if E is efficient. Now, the observation that $f_{\hat{\boldsymbol{\omega}}}\left(\hat{\mathbf{s}}\right)$ does not depend on E establishes the following result:

Proposition 3.1. *If the set of efficient electorates is nonempty, it is also the set of maximizers of $\mathcal{U}(\cdot)$.*

Remark 3.1. It follows from Proposition 3.1 that, when at least one efficient electorate exists, the smallest efficient electorate solves the planner's problem under simple majority rule and the naïve ballot strategy profile. Suppose that the planner chooses an efficient electorate E, and that all citizens besides j are naïve. Then citizen j's problem can be framed as

$$\max_{\sigma_j} U\left(\mathcal{L}\left(E, R_{\mathrm{SM}}, \left(\bar{\boldsymbol{\sigma}}_{-j}, \sigma_j\right)\right)\right). \quad (3.7)$$

Let $\mathcal{O}_j\left(E, \hat{\mathbf{s}}, \sigma_j\right)$ denote the outcome of the simple majority rule election when M is the electorate, all citizens in $N \setminus \{j\}$ are naïve, and Citizen j uses the ballot strategy σ_j. Then the maximand in (3.7) can be rewritten as

$$\sum_{\hat{\mathbf{s}} \in \hat{\Omega}^n} f_{\hat{\boldsymbol{\omega}}}\left(\hat{\mathbf{s}}\right) \cdot \left[\underbrace{f_{\omega|\hat{\boldsymbol{\omega}}}\left(-1 \,\middle|\, \hat{\mathbf{s}}\right) \mathcal{O}_j\left(E, \hat{\mathbf{s}}, \sigma_j\right)\left(\{-1\}\right) + f_{\omega|\hat{\boldsymbol{\omega}}}\left(1 \,\middle|\, \hat{\mathbf{s}}\right) \mathcal{O}_j\left(E, \hat{\mathbf{s}}, \sigma_j\right)\left(\{1\}\right)}_{\text{Convex combination between } f_{\omega|\hat{\boldsymbol{\omega}}}\left(-1|\hat{\mathbf{s}}\right) \text{ and } f_{\omega|\hat{\boldsymbol{\omega}}}\left(1|\hat{\mathbf{s}}\right)} \right]. \qquad (3.8)$$

Since E is efficient and $f_{\omega|\hat{\boldsymbol{\omega}}}\left(\cdot \,\middle|\, \hat{\mathbf{s}}\right)$ does not depend on σ_j, the convex combination between $f_{\omega|\hat{\boldsymbol{\omega}}}\left(-1 \,\middle|\, \hat{\mathbf{s}}\right)$ and $f_{\omega|\hat{\boldsymbol{\omega}}}\left(1 \,\middle|\, \hat{\mathbf{s}}\right)$ in (3.8) is maximized, for each $\hat{\mathbf{s}} \in \hat{\Omega}^n$, when $\sigma_j = \bar{\sigma}_j$—that is, when Citizen j is naïve. Because $f_{\hat{\boldsymbol{\omega}}}\left(\hat{\mathbf{s}}\right)$ also does not depend on σ_j, (3.8) is maximized when Citizen j is naïve. Hence the naïve ballot strategy solves each citizen's problem under majority rule, an efficient electorate, and the naïve ballot strategy profile for all other citizens.

Proposition 3.1 shows that, in this setting, a smallest efficient electorate (if any efficient electorates exist) is socially optimal. Suppose, however, that the analyst seeks an electorate whose behavior provides full information about the distribution of signals. To be more precise, suppose that the goal is to appoint an electorate that, for any realization of signals, will implement the outcome (i.e. element of ΔX) that would prevail if, given the same realization of signals, the election were decided according to the posterior distribution that the realization determines. It is clear that this condition, which I call *full information aggregation*, is stronger than efficiency.[12] Formally, it is defined as follows.

Definition 3.11. Let E be a feasible electorate and $\hat{\mathbf{s}}$ be a k-tuple of signals. *E aggregates information at $\hat{\mathbf{s}}$* if $\mathcal{O}\left(E, \hat{\mathbf{s}}\right)\left(\{x\}\right) = f_{\omega|\hat{\boldsymbol{\omega}}}\left(x \,\middle|\, \hat{\mathbf{s}}\right)$ for each $x \in X$.

Definition 3.12. Let E be a feasible electorate. *E fully aggregates information* if E aggregates information at every $\hat{\mathbf{s}} \in \hat{\Omega}^k$.

The definitions of efficiency and full information aggregation provide obvious tests for each of the two properties: see whether the property holds at each k-tuple of signals. However, the behavioral assumptions allow convenient shortcuts for checking these properties. First, a couple of definitions.

Definition 3.13. A ballot strategy σ_j^{b} employed by citizen j is *monotonic* if, for every $E \in \mathcal{E}$,

(i) If $\sigma_j^{\mathrm{b}}\left(\hat{\mathbf{s}}_{K_j}, E\right) = \delta_{-1}$, then $\sigma_j^{\mathrm{b}}\left(\hat{\mathbf{s}}'_{K_j}, E\right) = \delta_{-1}$ for every $\hat{\mathbf{s}}' \in \hat{\Omega}^k$ that satisfies

$$\sum_{i \in K_j} \hat{s}'_i \le \sum_{i \in K_j} \hat{s}_i;$$

---

[12]Note that, when k is odd, efficiency and full information aggregation are equivalent.

(ii) If $\sigma_j^{\mathrm{b}}\left(\hat{\mathbf{s}}_{K_j}, E\right) = \delta_1$, then $\sigma_j^{\mathrm{b}}\left(\hat{\mathbf{s}}'_{K_j}, E\right) = \delta_1$ for every $\hat{\mathbf{s}}' \in \hat{\Omega}^k$ that satisfies

$$\sum_{i \in K_j} \hat{s}'_i \geq \sum_{i \in K_j} \hat{s}_i.$$

In words, under a monotonic ballot strategy, if a citizen chooses -1 for some tuple of observed signals, she also chooses -1 when she observes a tuple of signals in which there are more -1 signals. (The obvious analogue holds for 1 signals.)

Definition 3.14. A ballot strategy σ_j^{b} employed by citizen j is *symmetric* if, for every $E \in \mathcal{E}$ and $x \in X$, $\sigma_j^{\mathrm{b}}\left(-\hat{\mathbf{s}}_{K_j}, E\right) = \delta_{-x}$ whenever $\sigma_j^{\mathrm{b}}\left(\hat{\mathbf{s}}_{K_j}, E\right) = \delta_x$.

In words, under a symmetric ballot strategy, if a citizen chooses -1 for some tuple of observed signals, she chooses 1 when all components in the tuple are "flipped." (Again, the obvious analogue holds as well.)

Remark 3.2. Naïve citizens use monotonic and symmetric ballot strategies.

Proposition 3.2. *Let E be a feasible electorate. E is efficient if and only if E is efficient at $\hat{\mathbf{s}}$ whenever $\sum_{i \in K} \hat{s}_i \in \{1, 2\}$.*

The intuition behind the sufficiency of this condition for efficiency is as follows. Note that, depending on whether k is odd or even, either 1 or 2 is the smallest value of $\sum_{i \in K} \hat{s}_i$ for which $f_{\omega|\hat{\omega}}\left(1 \mid \hat{\mathbf{s}}\right) > \frac{1}{2}$. If the electorate E chooses 1 whenever a signal k-tuple $\hat{\mathbf{s}}$ with $\sum_{i \in K} \hat{s}_i \in \{1, 2\}$ is realized, then, by the monotonicity of the naïve ballot strategy, E also chooses 1 whenever the posterior distribution becomes more tilted (i.e., when $\sum_{i \in K} \hat{s}_i > 2$).

By the symmetry of the naïve ballot strategy, the electorate chooses -1 when $\sum_{i \in K} \hat{s}_i \in \{-1, -2\}$. By a similar argument to the one above, the electorate chooses -1 when the posterior distribution becomes even more tilted in favor of -1 (i.e., when $\sum_{i \in K} \hat{s}_i < -2$).

The only remaining case to consider is when k is even and the signals are evenly split (i.e., when $\sum_{i \in K} \hat{s}_i = 0$), but efficiency imposes no restrictions in this case.

Proof of Proposition 3.2. Necessity is obvious. Let $c \in \{1, 2\}$ satisfy $c \equiv k \pmod 2$, and suppose that E is efficient at $\hat{\mathbf{s}}$ whenever $\sum_{i \in K} \hat{s}_i = c$. Then

$$\left|\left\{j \in E : \bar{\sigma}_j\left(\hat{\mathbf{s}}_{K_j}, E\right) = \delta_{-1}\right\}\right| < \left|\left\{j \in E : \bar{\sigma}_j\left(\hat{\mathbf{s}}_{K_j}, E\right) = \delta_1\right\}\right| \quad \text{whenever} \quad \sum_{i \in K} \hat{s}_i = c. \quad (3.9)$$

Consider a k-tuple $\hat{\mathbf{s}}$ of signals that satisfies $\sum_{i \in K} \hat{s}_i > c$. Flipping $\frac{1}{2}\left[\sum_{i \in K} \hat{s}_i - c\right]$ of the ones in $\hat{\mathbf{s}}$ produces a new k-tuple $\hat{\mathbf{s}}'$ that satisfies $\sum_{i \in K} \hat{s}'_i = c$. More precisely, let $K' \subset \{i \in K : \hat{s}_k = 1\}$ satisfy $|K'| = \frac{k+c}{2}$, and, for each $i \in K$, define

$$\hat{s}'_i = \begin{cases} 1 & \text{if } i \in K', \\ -1 & \text{otherwise.} \end{cases}$$

Observe that $\sum_{i \in K} \hat{s}'_i = c$. Hence, by the monotonicity of the strategies in $\bar{\boldsymbol{\sigma}}$ and by (3.9),

$$\left|\left\{j \in E : \bar{\sigma}_j\left(\hat{\mathbf{s}}_{K_j}, E\right) = \delta_1\right\}\right| \geq \left|\left\{j \in E : \bar{\sigma}_j\left(\hat{\mathbf{s}}'_{K_j}, E\right) = \delta_1\right\}\right|$$
$$> \left|\left\{j \in E : \bar{\sigma}_j\left(\hat{\mathbf{s}}'_{K_j}, E\right) = \delta_{-1}\right\}\right|$$
$$\geq \left|\left\{j \in E : \bar{\sigma}_j\left(\hat{\mathbf{s}}_{K_j}, E\right) = \delta_{-1}\right\}\right|.$$

Hence E is efficient at $\hat{\mathbf{s}}$ whenever $\sum_{i \in K} \hat{s}_i \geq c$.

By the symmetry of the ballot strategies in $\bar{\boldsymbol{\sigma}}$, (3.9) implies that

$$\left|\left\{j \in E : \bar{\sigma}_j\left(\hat{\mathbf{s}}_{K_j}, E\right) = \delta_{-1}\right\}\right| > \left|\left\{j \in E : \bar{\sigma}_j\left(\hat{\mathbf{s}}_{K_j}, E\right) = \delta_1\right\}\right| \quad \text{whenever} \quad \sum_{i \in K} \hat{s}_i = -c. \tag{3.10}$$

Now, applying the monotonicity of the strategies in $\bar{\boldsymbol{\sigma}}$ and (3.10) yields the result that E is efficient at $\hat{\mathbf{s}}$ whenever $\sum_{i \in K} \hat{s}_i \leq -c$.

The only remaining case to consider is that in which $-c < \sum_{i \in K} \hat{s}_i < c$. This situation arises if and only if k is even and $\sum_{i \in K} \hat{s}_i = 0$, which means that the signals are evenly split; thus any outcome of the election is consistent with efficiency. $\qquad \square$

Corollary 3.1 follows from Proposition 3.2 and Definition 3.12.

Corollary 3.1. *Let E be a feasible electorate. E fully aggregates information if and only if the following conditions hold:*

(i) $\mathcal{O}\left(E, \hat{\mathbf{s}}\right) = \delta_1$ whenever $\sum_{i \in K} \hat{s}_i \in \{1, 2\}$;

(ii) $\mathcal{O}\left(E, \hat{\mathbf{s}}\right) = \left(\frac{1}{2}, \frac{1}{2}\right)$ whenever $\sum_{i \in K} \hat{s}_i = 0$.

3.5 Electorates in Social Networks

In this section—which consists mostly of illustrative examples—I consider the special case in which $n = k$ (so that M is square), the principal diagonal of M contains all ones, and M is symmetric. This case can represent the scenario in which each citizen $j \in N$ receives a private signal $\hat{s}_j$ and then, before the election, truthfully communicates her private signal to all of her direct neighbors in a social network that is described by an undirected graph in which the nodes are citizens and edges join two citizens that share their signals with each other.[13] I use the term "social network" (or just "network") to describe this type of graph. Given a citizen $j \in N$, let N_j represent j's neighbors in the network: that is, $N_j \equiv \{i \in N \setminus \{j\} : m_{i,j} = 1\}$.

Example 3.1. Consider the social network shown in Figure 3.2a. It is clear that, in this net-

---

[13]The adjacency matrix of the undirected graph is given by $M - I_n$, where I_n is the $n \times n$ identity matrix.

(a) A star network with ten citizens

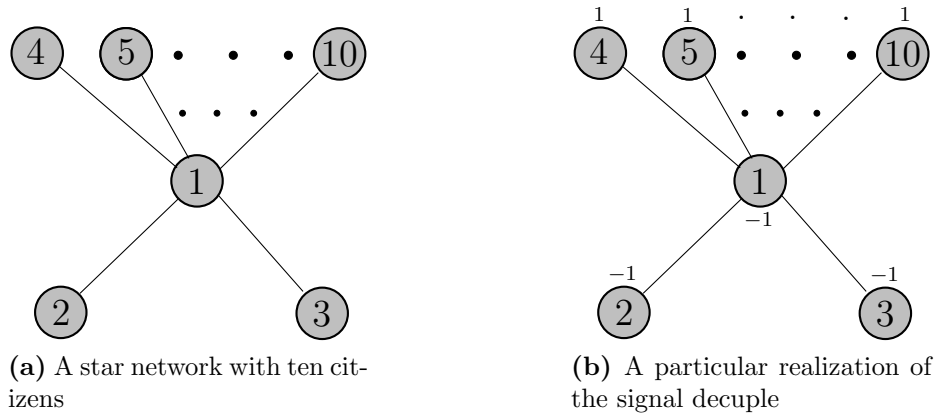(b) A particular realization of the signal decuple

Figure 3.2: Direct democracy can be inefficient

work, the dictatorship of Citizen 1 fully aggregates information. It is less obvious, however, that direct democracy is not efficient. To see that this is the case, suppose that the signals are as shown in Figure 3.2b. Then, when the social planner chooses the trivial electorate and all citizens are naïve, the ballot profile is as shown in Table 3.1. Hence -1 wins the

Citizen	1	2	3	4	5	6	7	8	9	10
Ballot	1	-1	-1	0	0	0	0	0	0	0

Table 3.1: The ballot profile for Figure 3.2b

election. However, given these specific realizations of the signals, the conditional probability that the state of the world is 1 is given by

$$\Pr\left(\omega = 1 \,|\, \hat{\omega}_1 = -1 \wedge \ldots \wedge \hat{\omega}_3 = -1 \wedge \hat{\omega}_4 = 1 \wedge \ldots \wedge \hat{\omega}_{10} = 1\right)$$
$$= f_{\omega|\hat{\omega}}\left(1\,|(-1,-1,-1,1,\ldots,1)\right)$$
$$> \frac{1}{2}.$$

So, in this example, the trivial electorate is not efficient, but the dictatorship of Citizen 1 is efficient. By Proposition 3.1, $\mathcal{U}(N) < \mathcal{U}(\{1\})$. In particular, this example illustrates that, in this setting, it is possible for a dictatorship to yield strictly higher ex ante welfare than a direct democracy. Intuitively, when there is one very well connected citizen, her signal can bias the outcome if everyone votes.

Remark 3.3. The assumption of naïve voting plays an important role in the inefficiency of direct democracy in Example 3.1. Since the priors are uniform and the electoral rule is simple majority, the trivial electorate fully aggregates information in any network if each

citizen votes informatively (i.e., casts a ballot based only upon her private signal). Hence, with strategic citizens, direct democracy is efficient in any network in this environment.

The star network in Example 3.1 belongs to a general class of networks in which there always exist dictatorships that fully aggregate information.

Definition 3.15. A network is a *generalized star* if there exists $j \in N$ for which $N_j = N \backslash \{j\}$.

In words, a generalized star is any network in which at least one citizen is connected to all other citizens. This class of networks includes stars and complete graphs. By Definitions 3.7 and 3.12, in any generalized star, a dictatorship by a citizen with the largest neighbor set fully aggregates information (and, therefore, is efficient).

Definition 3.16. Let E be an electorate and j be a citizen. *E observes j's signal* if $E \cap (\{j\} \cup N_j) \neq \varnothing$.

In words, the electorate E observes j's signal if j or one of j's neighbors is an elector. In all generalized stars, a citizen with the largest neighbor set observes each citizen's signal. As the following example illustrates, however, a dictator can achieve efficiency even if she does not observe all signals.

Example 3.2. Consider the four-citizen line, shown in Figure 3.3. A somewhat salient
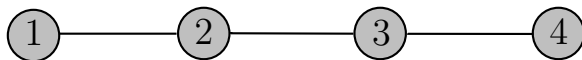

Figure 3.3: The four-citizen line

efficient electorate is $\{1, 4\}$. A less obvious one is $\{2\}$ (or $\{3\}$). Note, however, that $4 \notin N_2$ (and $1 \notin N_3$). It is also worth noting that, while $\{1, 4\}$ fully aggregates information, neither $\{2\}$ nor $\{3\}$ does.

Examples 3.1 and 3.2 together illustrate the basic tension that is at play in the electorate formation problem: on one hand, the social planner wants the electorate to observe as many signals as possible (which tends to increase the size of the electorate), but, on the other hand, the social planner wants to minimize the number of redundant signals observed (which tends to decrease the size of the electorate).

The following example illustrates an application of Proposition 3.2. It also shows that efficient electorates do not always exist.

Example 3.3. Consider the five-citizen line, shown in Figure 3.4. According to Proposition 3.2, to find efficient electorates, it suffices to consider the signal quintuples in which exactly three citizens observe the signal 1. I consider three cases that show that an efficient electorate does not exist in this social network.
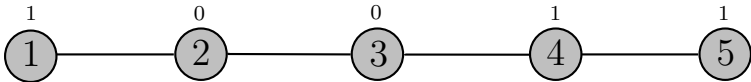
$$
\boxed{1} \quad\text{—}\quad \boxed{2} \quad\text{—}\quad \boxed{3} \quad\text{—}\quad \boxed{4} \quad\text{—}\quad \boxed{5}
$$

Figure 3.4: The five-citizen line

$$
\overset{1}{\boxed{1}} \quad\text{—}\quad \overset{0}{\boxed{2}} \quad\text{—}\quad \overset{0}{\boxed{3}} \quad\text{—}\quad \overset{1}{\boxed{4}} \quad\text{—}\quad \overset{1}{\boxed{5}}
$$

Figure 3.5: Case 1 of Example 3.3

Citizen	1	2	3	4	5
Ballot	0	-1	-1	1	1

Table 3.2: The ballot profile in Case 1 of Example 3.3

Case 1: Consider the realization of signals that is depicted in Figure 3.5. At this quintuple of signals, when citizens are naïve, the ballot profile is depicted in Table 3.2. Hence the following electorates are efficient at this signal quintuple:

- $\{4\}$
- $\{5\}$
- $\{4, 5\}$
- $\{1, 4\}$
- $\{1, 5\}$
- $\{1, 4, 5\}$
- $\{2, 4, 5\}$
- $\{3, 4, 5\}$
- $\{1, 2, 4, 5\}$
- $\{1, 3, 4, 5\}$

Case 2: At the signal quintuple that is depicted in Figure 3.6, the ballot profile is shown in Table 3.3. Of the electorates that are efficient at the signal quintuple described in

$$
\overset{1}{\boxed{1}} \quad\text{—}\quad \overset{1}{\boxed{2}} \quad\text{—}\quad \overset{0}{\boxed{3}} \quad\text{—}\quad \overset{0}{\boxed{4}} \quad\text{—}\quad \overset{1}{\boxed{5}}
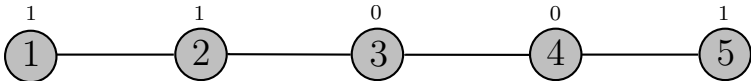$$

Figure 3.6: Case 2 of Example 3.3

Case 1, only $\{1, 5\}$ and $\{1, 2, 4, 5\}$ are also efficient at this signal quintuple.

Citizen	1	2	3	4	5
Ballot	1	1	−1	−1	0

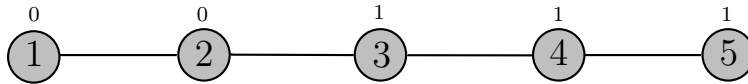**Table 3.3:** The ballot profile in Case 2 of Example 3.3

Figure 3.7: Case 3 of Example 3.3

Case 3: Now consider the signal quintuple that is depicted in Figure 3.7. Here, the ballot profile is as shown in Table 3.4. Note that neither $\{1,5\}$ nor $\{1,2,4,5\}$ is efficient at

Citizen	1	2	3	4	5
Ballot	−1	−1	1	1	1

Table 3.4: The ballot profile in Case 3 of Example 3.3

this signal quintuple.

So, for the social network described by the five-citizen line, there is no efficient electorate. It is straightforward, though tedious, to check that the $\mathcal{U}$-maximizing electorates in this network are $\{1,2,4,5\}$ and $\{1,2,3,4,5\}$. Recall that $\succsim_P$, by Assumption 3.8, is lexicographic. The planner's unique most preferred electorate, therefore, is $\{1,2,4,5\}$; in this example, the planner strictly prefers representative democracy to direct democracy.

Remark 3.4. Example 3.3 demonstrates that the allocation, and not merely the distribution, of signals has an effect on outcomes in this environment. For example, in each of the three cases analyzed in Example 3.3, the signal quintuple contains three ones and two zeros. However, as shown in Tables 3.2, 3.3, and 3.4, the election outcomes are not identical across the three cases. In other words, the allocation of signals is relevant, even if the distribution is fixed.

Example 3.4. It may be surprising that at least one electorate that fully aggregates information (and at least two additional efficient electorates) exist for the four-citizen line, yet no efficient electorate exists for the five-citizen line. It turns out that the appropriate generalization, in this context, of the four-citizen line to n citizens is not the n-citizen line, but rather the n-citizen comb, shown in Figure 3.8. (Note that the four-citizen comb is isomorphic to the four-citizen line.)

In the n-citizen comb, the electorate $E_n \equiv \left\{2j : j \in \left\{1, \ldots, \frac{n}{2}\right\}\right\}$ that comprises all citizens with even-numbered indices fully aggregates information. In fact, in the n-citizen comb,
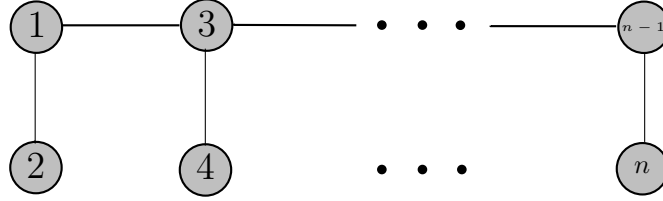
Figure 3.8: The n-citizen comb

the following relationship holds for all signal n-tuples $\hat{\mathbf{s}}$:

$$\left|\{i \in E_n : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_n\right) = \delta_1\}\right| - \left|\{i \in E_n : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_n\right) = \delta_{-1}\}\right| = \frac{1}{2}\sum_{j \in N}\hat{s}_j. \quad (3.11)$$

In words, (3.11) says that the difference between the number of 1 votes and the number of -1 votes by the electorate E_n is half the difference in the number of 1 signals and -1 signals in the signal n-tuple. This fact can be proved by induction. As the base case, consider $n = 2$. In this case, $E_2 = \{2\}$. If both citizens receive the same signal, then Elector 2 votes according to that signal, and (3.11) holds. If the two citizens receive different signals, Elector 2 abstains, and, once again, (3.11) holds.

Now suppose that there exists $m \in \mathbb{N}$ for which (3.11) holds when $n = 2m$. Consider the case in which $n = 2(m + 1)$. Let $N' \equiv \{1, \dots, 2m\}$. By the induction hypothesis,

$$\left|\left\{i \in E_{2m} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i'}, E_{2m}\right) = \delta_1\right\}\right| - \left|\left\{i \in E_{2m} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i'}, E_{2m}\right) = \delta_{-1}\right\}\right| = \frac{1}{2}\sum_{j \in N'}\hat{s}_j. \quad (3.12)$$

Because $N_i = N_i'$ for $i \in E_{2m}$, (3.12) can be rewritten as

$$\left|\{i \in E_{2m} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_{2m}\right) = \delta_1\}\right| - \left|\{i \in E_{2m} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_{2m}\right) = \delta_{-1}\}\right| = \frac{1}{2}\sum_{j \in N'}\hat{s}_j. \quad (3.13)$$

There are four possibilities:

Case 1: Suppose that $(\hat{s}_{2m+1}, \hat{s}_{2m+2}) = (-1, -1)$. In this case, Citizen $2m + 2$ votes for -1. Observe that

$$\left|\left\{i \in E_{2(m+1)} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_{2(m+1)}\right) = \delta_{-1}\right\}\right| = \left|\{i \in E_{2m} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_{2m}\right) = \delta_{-1}\}\right| + 1,$$

$$\left|\left\{i \in E_{2(m+1)} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_{2(m+1)}\right) = \delta_1\right\}\right| = \left|\{i \in E_{2m} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_{2m}\right) = \delta_1\}\right|,$$

$$\frac{1}{2}\sum_{j \in N}\hat{s}_j = \frac{1}{2}\sum_{j \in N'}\hat{s}_j + 1.$$

Hence, by (3.13),

$$\left|\left\{i \in E_{2(m+1)} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_{2(m+1)}\right) = \delta_1\right\}\right|$$
$$- \left|\left\{i \in E_{2(m+1)} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_{2(m+1)}\right) = \delta_{-1}\right\}\right| = \frac{1}{2}\sum_{j \in N}\hat{s}_j.$$

That is, (3.11) holds for $n = 2(m+1)$.

Case 2: Suppose that $(\hat{s}_{2m+1}, \hat{s}_{2m+2}) = (-1, 1)$. In this case, Citizen $2m + 2$ abstains, and so

$$\left|\left\{i \in E_{2(m+1)} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_{2(m+1)}\right) = \delta_{-1}\right\}\right| = \left|\left\{i \in E_{2m} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_{2m}\right) = \delta_{-1}\right\}\right|,$$
$$\left|\left\{i \in E_{2(m+1)} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_{2(m+1)}\right) = \delta_1\right\}\right| = \left|\left\{i \in E_{2m} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_{2m}\right) = \delta_1\right\}\right|,$$
$$\frac{1}{2}\sum_{j \in N}\hat{s}_j = \frac{1}{2}\sum_{j \in N'}\hat{s}_j.$$

By (3.13),

$$\left|\left\{i \in E_{2(m+1)} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_{2(m+1)}\right) = \delta_1\right\}\right|$$
$$- \left|\left\{i \in E_{2(m+1)} : \bar{\sigma}_i\left(\hat{s}_i, \hat{\mathbf{s}}_{N_i}, E_{2(m+1)}\right) = \delta_{-1}\right\}\right| = \frac{1}{2}\sum_{j \in N}\hat{s}_j.$$

So, once again, (3.11) holds for $n = 2(m+1)$.

Case 3: Suppose that $(\hat{s}_{2m+1}, \hat{s}_{2m+2}) = (1, -1)$. This case is identical to Case 2.

Case 4: Suppose that $(\hat{s}_{2m+1}, \hat{s}_{2m+2}) = (1, 1)$. This case is symmetric to Case 1.

By induction, (3.11) holds in general for the n-citizen comb. The fact that E_n fully aggregates information in the n-citizen comb is an immediate consequence of (3.11).

An intuition for why E_n fully aggregates information in the n-citizen comb is as follows. Consider two citizens, $j_1, j_2 \in N$, such that $N_{j_1} = \{j_2\}$. If j_1 is naïve, then j_1's ballot, which is based on $\hat{\omega}_{j_1}$ and $\hat{\omega}_{j_2}$, provides enough information regarding those two signals to properly update the posterior distribution on Ω. In particular, j_1 votes for either alternative if and only if she and j_2 have agreeing signals. Otherwise she abstains.

In particular, if n is even and the citizenry can be partitioned as

$$\left\{\{j_{2k-1}, j_{2k}\} : k \in \left\{1, \ldots, \frac{n}{2}\right\}\right\}$$

with $N_{j_{2k-1}} = \{j_{2k}\}$ for each $k \in \left\{1, \ldots, \frac{n}{2}\right\}$, then the electorate $\left\{j_{2k-1} : k \in \left\{1, \ldots, \frac{n}{2}\right\}\right\}$ fully aggregates information (and, therefore, is efficient as well). The class of networks in which such a partition is achievable includes combs, and, in particular, the four-citizen line. The class includes networks like the one shown in Figure 3.9, which can be derived from the comb by adding links among the set of non-electors. The five-citizen line, however, does not belong to this class.
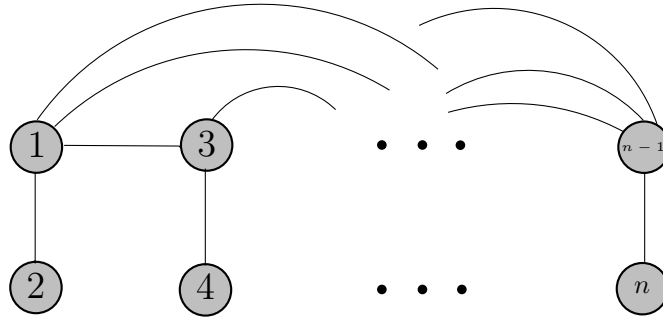
Figure 3.9: The analysis of the comb applies to this network as well.

Example 3.5. Example 3.4 illustrates why, in the four-citizen line, full information aggregation is possible. As shown in Example 3.2, though, there are efficient dictatorships (which do not fully aggregate information) in the four-citizen line. This result can be generalized to networks with n citizens, but, again, the n-citizen line is not the appropriate generalization. (Indeed, as Example 3.3 shows, an efficient electorate does not exist for the five-citizen line.) The appropriate generalization is a different class of social networks, which is characterized in Definition 3.17.

Definition 3.17. A social network of n citizens is a *quasistar* if there exists some $j \in N$ for which $|N_j| \geq n - 2$.

In words, a quasistar is any network in which there is at least one citizen who has at least $n - 2$ neighbors, and hence observes at least $n - 1$ signals. All generalized stars are quasistars, as is the four-citizen line. The five-citizen line, however, is not a quasistar.

Proposition 3.3. *In any quasistar in which n is even, the dictatorship of any citizen j for which $|N_j| \geq n - 2$ is efficient.*

Intuitively, the reason why Proposition 3.3 holds is as follows. When n is even and the dictator observes $n - 1$ or more signals, the "worst" cases are the following:

(i) The signals are evenly split and the dictator chooses the alternative that she thinks is strictly more likely based on the $n - 1$ signals that she observes. An example of this scenario, for $n = 6$, is illustrated in Figure 3.10a: the dictator (citizen 1) will vote for -1, even though both outcomes are equally likely, given the full sextuple of signals. Note, however, that efficiency places no restrictions on the election outcome in such a case.

(ii) The signals are unbalanced, and the signal that the dictator does not observe corresponds to the more likely outcome. An example is illustrated in Figure 3.10b. In this case, because n is even, the difference in the number of 0 signals and the number of 1 signals is strictly positive and even; in particular, it is at least 2. Since there is only

one signal that the dictator does not observe, there is an imbalance (in the correct direction) of at least 1 among the signals that the dictator observes. So the dictator implements the outcome that is consistent with efficiency.
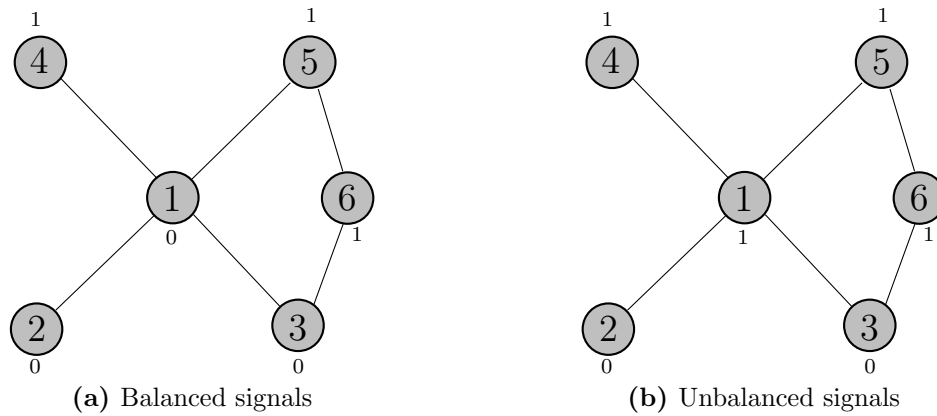

(a) Balanced signals   **(b)** Unbalanced signals

Figure 3.10: The two "worst" cases for Example 3.5

Proof of Proposition 3.3. Begin by noting that, if $|N_j| > n-2$, then it must be the case that $|N_j| = n-1$, and, by the analysis of Example 3.1, $\{j\}$ fully aggregates information; hence $\{j\}$ is efficient. (In this case, efficiency holds regardless of whether n is even or odd.)

Suppose that n is even and, without loss of generality, that $N_1 = \{2, \dots, n-1\}$ (i.e. citizen 1 observes all signals except that of citizen n—Figures 3.10a and 3.10b both show such a network with six citizens). According to the claim, $\{1\}$ is an efficient dictatorship. Using Proposition 3.2, to prove that $\{1\}$ is efficient, it suffices to consider the cases in which the n-tuple of signals $\hat{\mathbf{s}}$ satisfies $\sum_{j \in N} \hat{s}_j = 2$. Efficiency requires that Citizen 1 vote for alternative 1. There are two cases to consider.

Case 1: Suppose that $\hat{s}_n = -1$. Then $\sum_{j \in N_1} \hat{s}_j = 3$, so Citizen 1 votes for 1, as required.

Case 2: Suppose that $\hat{s}_n = 1$. In this case, $\sum_{j \in N_1} \hat{s}_j = 1$. Again, Citizen 1 votes for 1.

$\square$

To see the importance of the assumption (in the statement of Proposition 3.3) that n is even, consider the quasistar shown in Figure 3.11. In this network, n is odd. For this particular realization of signals, Citizen 1 abstains, whereas an efficient electorate should select 1.

A natural question to ask is whether any monotonic relationships hold between the "amount of information sharing" in a network and the size of the smallest electorate that satisfies efficiency or full information aggregation. Clearly information sharing in a network
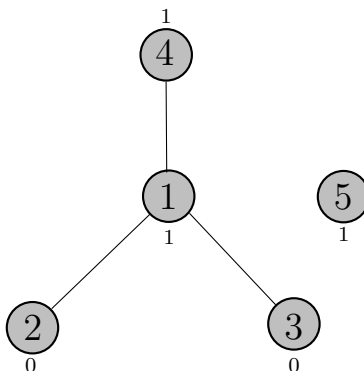
Figure 3.11: A quasistar with five citizens

is through edges, so the partial order of set containment on the edge sets can allow comparisons of information sharing between certain pairs of networks.[14] In terms of the information allocation matrices, the network that M represents allows more information sharing than the network that M' represents if and only if both matrices have the same dimensionality and $m_{i,j} = 1$ whenever $m'_{i,j} = 1$. This relationship can be written more succinctly with the partial ordering convention for vectors in Euclidean spaces: $M \geq M'$. Consider the following two extreme examples:

(i) $M = I_n$ (i.e. a network without communication, in which the trivial electorate is the only one that fully aggregates information)

(ii) $M'_{i,j} = 1$ if and only if $i = 1$ or $j = 1$ or $i = j$ (i.e. a star with Citizen 1 at the center, in which the dictatorship of Citizen 1 fully aggregates information)

From these examples, one may surmise that, given two networks M and M' in which electorates exist that fully aggregate information, if $M \geq M'$, then the smallest electorate that fully aggregates information in M is no larger than the smallest electorate that fully aggregates information in M'. An intuition for such a conjecture might be as follows: given that more information is being exchanged by individuals in M, it should be "easier" to fully aggregate information with an electorate of a given size in M'. The following two examples show, however, that this intuition is mistaken.[15]

Example 3.6. The four-citizen polygon (i.e. square), shown in the Figures 3.12a, 3.12b, and 3.12c, is a quasistar. The figures show three realizations of the signal quadruple. The associated ballot profiles are shown, respectively, in Tables 3.5, 3.6, and 3.7.   By Proposition 3.2, to check the efficiency of a given electorate, it is sufficient to consider signal quadruples that contain exactly three ones. Any such quadruple can be described by a rotation of Figure 3.12a. It is thus straightforward to verify that any dictatorship is efficient in

---

[14]Another measure, which I have not yet considered, is clustering.

[15]The problem with this argument is that it ignores the fact that adding links in the network can introduce redundancies in the electorate's information.
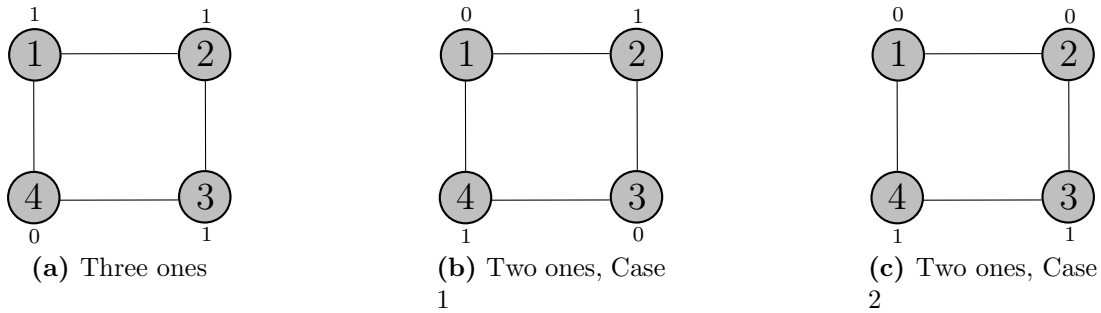
(a) Three ones

(b) Two ones, Case 1

(c) Two ones, Case 2

Figure 3.12: Three signal quadruples in the four-citizen polygon

Citizen	1	2	3	4
Ballot	1	1	1	1

Table 3.5: The ballot profile associated with Figure 3.12a

Citizen	1	2	3	4
Ballot	1	-1	1	-1

Table 3.6: The ballot profile associated with Figure 3.12b

Citizen	1	2	3	4
Ballot	-1	-1	1	1

Table 3.7: The ballot profile associated with Figure 3.12c

the square. To check full information aggregation, one also needs to consider signal quadruples that contain exactly two ones; any such quadruple is a rotation of either Figure 3.12b or Figure 3.12c. By considering all three of these signal configurations and their rotations, it is not difficult to see that the only electorate that fully aggregates information in this network is the trivial electorate, $\{1, 2, 3, 4\}$. Now, recall from Example 3.2 that the electorate $\{1, 4\}$ fully aggregates information for the four-citizen line. The four-citizen polygon allows more information exchange than the four-citizen line, yet the unique electorate that fully aggregates information in the four-citizen polygon is strictly larger than an electorate that fully aggregates information in the four-citizen line.

Example 3.7. Recall Example 3.3, which illustrates that no efficient electorate exists in the five-citizen line, and hence that full information aggregation is impossible in the five-citizen

line. On the other hand, the trivial electorate fully aggregates information in the five-citizen society with no communication, which allows strictly less information exchange than the five-citizen line.

The fact that an efficient electorate exists in the five-citizen society with no links and the fact that no efficient electorate exists for the five-citizen line, together with Proposition 3.1, imply that the maximal value of $\mathcal{U}(\cdot)$ that is achievable in the five-citizen society with no links is strictly greater than the maximal value of $\mathcal{U}(\cdot)$ that is achievable in the five-citizen line. In this case, a strict increase in the amount of information exchange leads to a strict decrease in achievable ex ante welfare! Once again, note that the behavioral assumptions are crucial here: with strategic citizens, direct democracy is always efficient, since every citizen can choose to vote informatively by disposing of the information that she receives from others.

Example 3.8. It is natural to ask whether efficiency is achievable in lines of more than five citizens. Consider the n-citizen line, shown in Figure 3.13. As it turns out, direct democracy
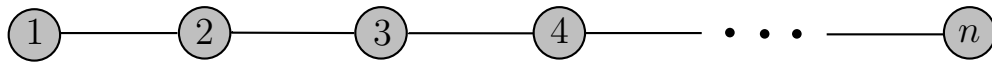
Figure 3.13: The n-citizen line

is efficient when $n = 6$. When $n \in \{7, 9, 10\}$, efficiency is unachievable, but direct democracy is optimal. Perhaps surprisingly, when $n = 8$, the electorate $\{1, 2, 4, 5, 7, 8\}$ is efficient.

Example 3.9. Example 3.6 and similar analyses of the polygons for $n \in \{1, 2, 3\}$ suggest that the trivial electorate fully aggregates information in a polygon (or circle): a network of the form shown in Figure 3.14. Indeed, in the cases of $n = 5$ and $n = 7$, the trivial electorate

Figure 3.14: The n-citizen circle

is efficient. For $n = 9$, though, efficiency is not achievable, although the trivial electorate is optimal. In fact, as Figure 3.15 shows, the trivial electorate need not be efficient for large n, whether odd or even. The ballot profiles for the signal tuples depicted in Figures 3.15a
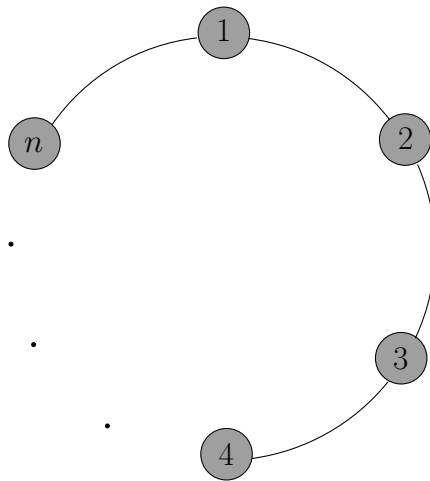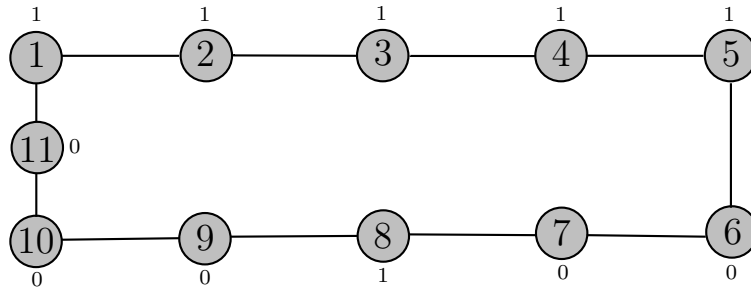

(a) Odd Case
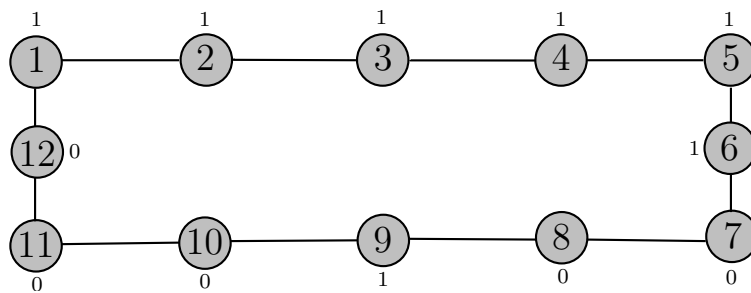
(b) Even Case

Figure 3.15: The trivial electorate may not be efficient

and 3.15b are given, respectively, in Tables 3.8 and 3.9. In both cases, efficiency requires that

Citizen	1	2	3	4	5	6	7	8	9	10	11
Ballot	1	1	1	1	1	-1	-1	-1	-1	-1	-1

Table 3.8: The ballot profile associated with Figure 3.15a

Citizen	1	2	3	4	5	6	7	8	9	10	11	12
Ballot	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1

Table 3.9: The ballot profile associated with Figure 3.15b

alternative 1 win outright. In the signal tuple shown in Figure 3.15a, however, alternative -1

wins outright when the trivial electorate is chosen. In the signal tuple shown in Figure 3.15b, polling the trivial electorate results in a tie.

For some small circles, efficiency can be achieved with nontrivial electorates. For example, the case of $n = 6$, polling every other citizen (i.e., either all odd-indexed citizens or all even-indexed citizens) is efficient. In the case of $n = 8$, the electorate $\{1, 2, 4, 5, 7\}$ is efficient, as are its rotations.

Example 3.10. In certain environments—for example, one of electoral competition with political parties—the social network may be partitioned into two or more sets of highly clustered nodes, with relatively few links that cross sets. An extreme case would be a social network that consists of two or more connected components. In these settings, at least two salient hypotheses arise. The first is in the spirit of the so-called "independence of irrelevant alternatives" axiom from choice theory: given a society and a party in the society, the optimal electorate for the full society either contains no members from the party or contains an optimal electorate (and no additional members) from that party. To be precise, let (N, A) and (N', A') be two networks with $N \subset N'$, $A \subset A'$, and $(j, j') \notin A'$ whenever $j \in N$ and $j' \in N' \setminus N$. The hypothesis is that, if E' is an optimal electorate for (N', A') and $E' \cap N \neq \emptyset$, then $E' \cap N$ is an optimal electorate for (N, A).

Unfortunately, this hypothesis turns out to be false, as illustrated by Figure 3.16. In the
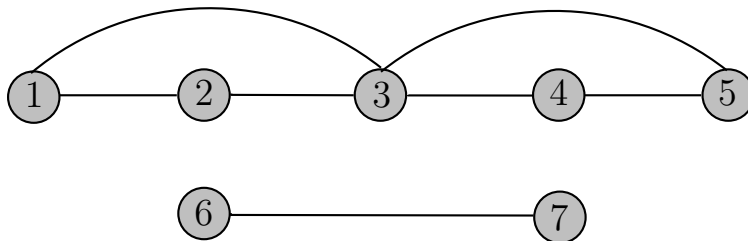
Figure 3.16: The hypothesis of Example 3.10 fails.

network shown in the figure, there are six optimal electorates. Each optimal electorate is composed of five citizens: 3, 6, 7, and exactly two from the set $\{1, 2, 4, 5\}$. For the top component, the unique efficient and optimal electorate is $\{3\}$. For the bottom component, any electorate is efficient. The intuition behind the failure of this conjecture is that, since ballots are weighted equally, a planner may want to include extra members from larger components of the network to account for differences in information quality across components.

Example 3.11. Continue with the environment of Example 3.10. Another hypothesis—motivated by the so-called "expansion consistency" axiom from choice theory—is that, given the optimal electorate for a party in a society, every optimal electorate of the full society includes either no members or all members of any optimal electorate for the party. Formally, let (N, A) and (N', A') be two networks with $N \subset N'$, $A \subset A'$, and $(j, j') \notin A'$ whenever $j \in N$ and $j' \in N' \setminus N$. The hypothesis is that, if E is an optimal electorate for (N, A),

then, for every optimal electorate E' for (N', A'), we have either $E \subseteq E'$ or $E \cap E' = \emptyset$. This hypothesis fails as well, for a similar reason as the hypothesis of Example 3.10. See Figure 3.17. In the figure, the unique efficient electorate for the top component is the trivial
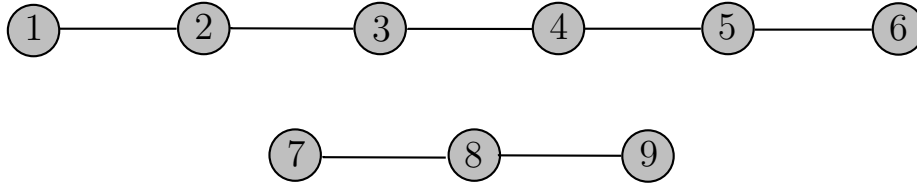

Figure 3.17: The hypothesis of Example 3.11 fails.

electorate. In the bottom component, any electorate that includes Citizen 8 is efficient. However, the two optimal electorates for the entire network are $\{1, 2, 3, 5, 6, 7, 8, 9\}$ and $\{1, 2, 4, 5, 6, 7, 8, 9\}$.

Example 3.12. While the behavioral assumption of naïveté appears reasonable, it includes one feature that is particularly tenuous and arbitrary: electors whose posterior beliefs are evenly distributed between the two alternatives will abstain. A natural question to ask is whether the basic results so far are very sensitive to this assumption. What happens if, for example, electors whose posterior beliefs are uniform follow their private signals rather than abstaining? Example 3.9, in which each citizen observes exactly three signals (and, therefore, never has uniform posterior beliefs), shows that efficiency and information aggregation may fail in a direct democracy even under a weaker (or otherwise modified) form of the naïveté assumption. To see that the main result of Example 3.1 (that dictatorship can be strictly more efficient than direct democracy) continues to hold under the above modification of the naïveté assumption, consider the network and the tuple of signals shown in Figure 3.18. The ballot profile is shown in Table 3.10. Here, the trivial electorate selects alternative -1,

Citizen	1	2	3	4	5	6	7	8	9	10	11
Ballot	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1

Table 3.10: The ballot profile associated with Figure 3.18 (under alternate naïveté assumption)

despite the preponderance of ones in the tuple of signals. Evidently, direct democracy is not efficient in this example. Since the network is a generalized star, Citizen 1, acting as a dictator, can fully aggregate information.

3.6 Discussion

In this final section, I present several conclusions that can be drawn from the analysis of Section 3.5, and I conclude with some open questions that remain.
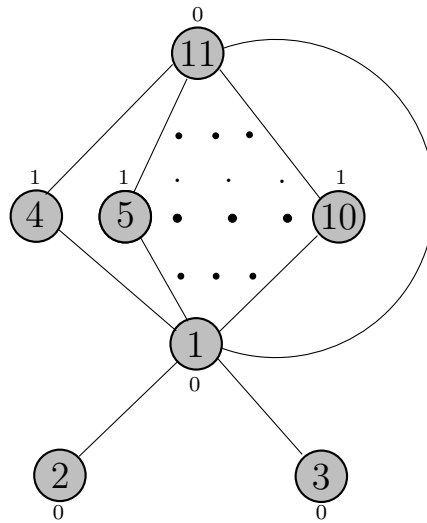
Figure 3.18: Direct democracy may not be efficient even under alternate behavioral assumptions

Some Facts

The examples of Section 3.5 illustrate the instability of the solution to the planner's problem. As a result of this instability, a complete characterization of the set of networks in which efficient electorates—and of the subset in which electorates that fully aggregate information—exist is elusive. At first glance, it might seem that, in the network setting of Section 3.5, an optimal—if not efficient—electorate will be formed by a maximal independent set in the graph.[16] As the examples illustrate, though, optimal electorates are typically not independent. While they often are dominating, though, neither independence nor dominance is necessary for efficiency, let alone for optimality.

From the examples, it is possible to derive some necessary conditions (which are not sufficient, but are easier to check than those in Proposition 3.2 and Corollary 3.1) for electorates to be efficient or to fully aggregate information. Note that, if the electorate M does not observe the signal of some citizen j, then j's signal has no effect on the election outcome. With this observation in mind, the following facts become clear:

Fact 3.1. Suppose that n is odd. The electorate M is efficient only if it observes each citizen's signal.

If n is odd and there is at least one citizen whose signal the electorate does not observe, the electorate chooses the same outcome regardless of the values of the unobserved signals. Obviously, this situation can lead to the worse (in expected terms) alternative being implemented when the signals that the electorate observes are evenly split (or close to being evenly split).

---

[16]A *dominating set* in a graph is a set of nodes such that every node not in the set is a neighbor of some node in the set. An *independent set* in a graph is a set of nodes of which no two members are neighbors. A *maximal independent set* is one that is dominating and independent.

Fact 3.2. Suppose that n is even. The electorate M is efficient only if there is at most one citizen's signal that it does not observe.

To see what can go wrong when the condition in Fact 3.2 is violated, note that, in this case, there are at least two signals that will not affect the electorate's decision. When the signals that the electorate observes are evenly split (or close to being evenly split), the unobserved signals can swing the posterior distribution in either direction, but the electorate's ballots may be tied or even tilted in favor of the worse alternative.

Fact 3.3. The electorate M achieves full information aggregation only if it observes each citizen's signal.

For odd n, Fact 3.3 follows immediately from Fact 3.1. For even n, it follows by noting that, if the full n-tuple of signals is evenly split between zeros and ones, an electorate that fully aggregates information should cast an equal number of ballots for each alternative. However, the event in which the full n-tuple of signals is evenly split is not discernible by an electorate that does not observe each citizen's signal.

Fact 3.4 follows from the observation that, when each elector has an odd number of neighbors, no elector's ballot strategy allows abstention. In particular, if there is an even number of citizens, then full information aggregation requires a tie vote whenever the signals are evenly split, which means that the size of the electorate must be even.

Fact 3.4. Let (N, K) be a network in which n is even, and let M be an electorate for which $|N_i|$ is odd for each $i \in M$. M fully aggregates information only if $|M|$ is even.

The following conjecture strengthens Fact 3.3.

Conjecture 3.1. The electorate M achieves full information aggregation only if there exists $r \in \mathbb{N}$ such that, for each $j \in N$, $|M \cap (\{j\} \cup N_j)| = r$.

In words, Conjecture 3.1 hypothesizes that an electorate achieves full information aggregation only if it observes all citizens' signals and each signal is "equally represented" in the electorate. Note that, as Example 3.9 illustrates, the converse of Conjecture 3.1 does not hold.

A final point worth noting is that the behavior of an electorate depends on the network structure only to the extent that the network structure determines the sources of information for each elector. Hence, given two networks (N, K) and (N, K') and an electorate $M \in 2^N \setminus \{\varnothing\}$, if the neighbor set of each $i \in M$ is the same between the two networks, then both $\mathcal{O}(M, \cdot)$ and $\mathcal{U}(M)$ will be identical across the two networks.

Future Work

The two major open questions can be framed as follows:

- *For which information allocations does an efficient electorate exist?* To answer this question, it suffices to characterize the following set for each $n, k \in \mathbb{N}$:

$$\left\{M \in \{0,1\}^{n \times k} : \exists E \subseteq \{1,\ldots,n\} \; \left(\forall \hat{\mathbf{s}} \in \{-1,1\}^k\right. \right.$$
$$\left. \left. \left[\operatorname{sgn}\left(\mathbf{sgn}\left(M_E \hat{\mathbf{s}}\right) \cdot \boldsymbol{\iota}_{|E|}\right) = \operatorname{sgn}\left(\hat{\mathbf{s}} \cdot \boldsymbol{\iota}_n\right) \vee \hat{\mathbf{s}} \cdot \boldsymbol{\iota}_n = 0\right]\right)\right\},$$

where $\mathbf{sgn}\left(\cdot\right)$ is the componentwise sign function, M_E is the $|E| \times k$ 0-1 matrix formed by taking only the rows of M with indices in E, and, for any $m \in \mathbb{N}$, $\boldsymbol{\iota}_m$ is the m-dimensional vector of all ones.

- *How does the solution to the planner's problem depend on the allocation of information?* That is, given $n, k \in \mathbb{N}$ and $M \in \{0,1\}^{n \times k}$, characterize the sets of minimum cardinality from the collection

$$\arg\max_{E \subseteq \{1,\ldots,n\}} \sum_{\hat{\mathbf{s}} \in \{-1,1\}^k} f_{\hat{\boldsymbol{\omega}}}\left(\hat{\mathbf{s}}\right) g\left(M, E, \hat{\mathbf{s}}\right),$$

where

$$g\left(M, E, \hat{\mathbf{s}}\right) \equiv \begin{cases} f_{\omega|\hat{\omega}}\left(-1 \,|\, \hat{\mathbf{s}}\right) & \text{if } \mathbf{sgn}\left(M_E \hat{\mathbf{s}}\right) \cdot \boldsymbol{\iota}_{|E|} < 0, \\ f_{\omega|\hat{\omega}}\left(1 \,|\, \hat{\mathbf{s}}\right) & \text{if } \mathbf{sgn}\left(M_E \hat{\mathbf{s}}\right) \cdot \boldsymbol{\iota}_{|E|} > 0, \\ \frac{1}{2}\left[f_{\omega|\hat{\omega}}\left(-1 \,|\, \hat{\mathbf{s}}\right) + f_{\omega|\hat{\omega}}\left(1 \,|\, \hat{\mathbf{s}}\right)\right] & \text{otherwise.} \end{cases}$$

Although the trivial electorate is often suboptimal in this setting, it generally seems to outperform most other electorates. Some less ambitious questions than the ones posed above would be the following:

- What is a tight lower bound on the efficiency of direct democracy?

- What circumstances are worst for the efficiency of direct democracy?

- When is direct democracy optimal?

- When is dictatorship optimal?

One extension that may also prove to be fruitful in obtaining general characterization results is to allow richer signal spaces, as the instability of the solution to the planner's problem may be driven—or at least exacerbated—by the binary signal structure and the tiebreaking assumption that it necessitates.

Another avenue for future research is to look at more general electoral rules. In particular, electoral rules that give greater weight to better informed citizens are likely to allow better outcomes. Indeed, there is no compelling reason to restrict attention to electoral rules that give each elector equal influence, let alone to restrict attention to simple majority rule. The optimal design of a general mechanism that comprises both an electorate and an electoral rule is certainly a more ambitious problem than the one studied here. Given the increased flexibility that it provides the mechanism designer, though, it may yield stronger results.

Appendix A

Supplement to Chapter 1

A.1 Proofs of Main Results

Proposition 1.1. *In the simple case, when the principal retains decision making authority, the worker's effort level, e_r^*, is positive and is characterized by $q'(e_r^*) = 2c'(e_r^*)$. The principal rubberstamps the worker's signal and achieves an expected payoff of $U_r^* \equiv \frac{q(e_r^*)}{2} - \frac{1}{4}$.*

Proof. By Bayes's Rule, the posterior beliefs of both the principal and the worker can be described as

$$\Pr(\{\omega = A\}|\{s = a\}) = \Pr(\{\omega = B\}|\{s = b\}) = \frac{1}{2} + q(e).$$

The principal (who is unbiased) rubberstamps the signal. Therefore the worker's optimal effort level, e_r^*, satisfies

$$e_r^* \in \arg\max_{e \geq 0} - [\lambda_W \cdot (1 - \pi_W) + (1 - \lambda_W) \cdot \pi_W] \cdot \left[\frac{1}{2} - q(e)\right] - c(e)$$

$$= \arg\max_{e \geq 0} \frac{q(e)}{2} - c(e) - \frac{1}{4}.$$

The characterization of the maximizer follows from Lemma 1.3. Setting e_r^* to be the maximizer and observing that, once the worker's effort level, e, is sunk, the worker's expected utility (gross of information acquisition costs) is the same as the principal's (specifically, $\frac{q(e)}{2} - \frac{1}{4}$), yields the result. $\square$

Lemma 1.2. *The worker's signal informs the decision of a manager of type π_M if and only if the signal's quality is high enough to overcome the manager's skepticism: $q(e) \geq \left|\pi_M - \frac{1}{2}\right|$.*

Proof. Suppose that the signal's quality is $q(e)$ and that its realization is a. Then the manager chooses α if and only if her posterior belief places at least as much weight on A as on B. This condition can be written as

$$\left(1 + \frac{\pi_M}{1 - \pi_M} \cdot \frac{\frac{1}{2} + q(e)}{\frac{1}{2} - q(e)}\right)^{-1} \leq \left(1 + \frac{1 - \pi_M}{\pi_M} \cdot \frac{\frac{1}{2} - q(e)}{\frac{1}{2} + q(e)}\right)^{-1},$$

which is equivalent to

$$q(e) \geq \frac{1}{2} - \pi_M. \tag{A.1}$$

Similarly, when the signal's quality is $q(e)$ and its realization is b, the manager chooses β if and only if

$$\left(1 + \frac{\pi_M}{1 - \pi_M} \cdot \frac{\frac{1}{2} - q(e)}{\frac{1}{2} + q(e)}\right)^{-1} \geq \left(1 + \frac{1 - \pi_M}{\pi_M} \cdot \frac{\frac{1}{2} + q(e)}{\frac{1}{2} - q(e)}\right)^{-1},$$

which is equivalent to

$$q(e) \geq \pi_M - \frac{1}{2}. \tag{A.2}$$

Thus the signal informs the manager's decision if and only if both (A.1) and (A.2) hold simultaneously (i.e., iff $q(e) \geq \left|\pi_M - \frac{1}{2}\right|$). □

Proposition 1.2. *In the simple case, when the principal appoints a manager of type π_M, the worker's effort level, $e_d^*(\pi_M)$, is determined by the manager's skepticism. In particular,*

$$e_d^*(\pi_M) = \begin{cases} e_r^* & \text{if } 0 \leq \left|\pi_M - \frac{1}{2}\right| \leq q(e_r^*), \\ q^{-1}\left(\left|\pi_M - \frac{1}{2}\right|\right) & \text{if } q(e_r^*) < \left|\pi_M - \frac{1}{2}\right| \leq \phi, \\ 0 & \text{if } \phi < \left|\pi_M - \frac{1}{2}\right| < \frac{1}{2}. \end{cases}$$

If $\left|\pi_M - \frac{1}{2}\right| \leq \phi$, the manager will rubberstamp the worker's signal. Otherwise, the manager will implement the action toward which she is biased. The principal's expected utility, $U_d^*(\pi_M)$, is

$$U_d^*(\pi_M) = \frac{q\left(e_d^*(\pi_M)\right)}{2} - \frac{1}{4} = \begin{cases} \frac{q(e_r^*)}{2} - \frac{1}{4} & \text{if } 0 \leq \left|\pi_M - \frac{1}{2}\right| \leq q(e_r^*), \\ \frac{\left|\pi_M - \frac{1}{2}\right|}{2} - \frac{1}{4} & \text{if } q(e_r^*) < \left|\pi_M - \frac{1}{2}\right| \leq \phi, \\ -\frac{1}{4} & \text{if } \phi < \left|\pi_M - \frac{1}{2}\right| < \frac{1}{2}. \end{cases}$$

Proof. This result is a special case of Proposition 1.3. □

Corollary 1.1. *In the simple case, when the principal can choose among candidate managers of different types, she prefers to appoint a manager of type $\pi_M \in \left[\frac{1}{2} - \phi, \frac{1}{2} - q(e_r^*)\right) \cup \left(\frac{1}{2} + q(e_r^*), \frac{1}{2} + \phi\right]$ than to retain decision making authority. Such a manager will rubberstamp the worker's signal, which would have informed the principal's choice, too, if she had retained control of the decision. In equilibrium, the principal appoints a manager of type $\pi_M \in \left\{\frac{1}{2} - \phi, \frac{1}{2} + \phi\right\}$.*

Proof. This result follows from a comparison of the expected utility levels in Propositions 1.1 and 1.2 and from Lemma 1.2. □

Corollary 1.2. *Consider a pool of workers that have $\pi_W = \lambda_W = \frac{1}{2}$ but vary in their returns to effort and costs of effort (i.e., in the attributes $q(\cdot)$ and $c(\cdot)$). The unbiased principal's preference ranking (from most preferred to least preferred) among these workers coincides with the workers' ordering (from highest to lowest) by tolerance for skepticism.*

Proof. Consider two workers, W_1 and W_2, from the pool, and suppose that their tolerances for skepticism, respectively, are ϕ_1 and ϕ_2, where $\phi_1 \geq \phi_2$. By Proposition 1.2, when the principal appoints W_i, the principal realizes a maximal expected payoff of $\frac{\phi_i}{2} - \frac{1}{4}$ by appointing a manager of skepticism ϕ_i. Thus her maximal expected payoff under W_1 is at least as high as under W_2: she weakly prefers to appoint W_1 than to appoint W_2. Note that, when $\phi_1 > \phi_2$, the principal's preference for W_1 is strict. The result follows. $\square$

Corollary 1.3. *Consider a pool of workers that have $\pi_W = \lambda_W = \frac{1}{2}$ but vary in their returns to effort and costs of effort (i.e., in the attributes $q(\cdot)$ and $c(\cdot)$). Define the following partial order among the workers:*

$$W_1 \succeq W_2 \overset{\Delta}{\Longleftrightarrow} \left(c_1 \circ q_1^{-1}\right)''(\cdot) \leq \left(c_2 \circ q_2^{-1}\right)''(\cdot),$$

where W_1 and W_2 are two workers from the pool with attributes of $q_1(\cdot)$ and $c_1(\cdot)$ (for W_1) and $q_2(\cdot)$ and $c_2(\cdot)$ (for W_2), respectively. If members of a subset of the pool can be ranked according to $\succeq$, then the unbiased principal's preference ranking, restricted to that subset, coincides with $\succeq$.

Proof. Consider two workers from the pool. Suppose that the first worker, W_1, has attributes $q_1(\cdot)$ and $c_1(\cdot)$, and that the second one, W_2, has attributes $q_2(\cdot)$ and $c_2(\cdot)$. For each $i \in \{1, 2\}$, denote the twice differentiable function $x \mapsto 2c_i\left(q_i^{-1}(x)\right)$ by $h_i(\cdot)$. By Lemma A.1, each $h_i(\cdot)$ has a unique positive fixed point, ϕ_i. By Definition 1.8, ϕ_i represents worker i's tolerance for skepticism. Observe that

$$h_i(0) = 0, \tag{A.3}$$

$$h_i'(0) = \frac{2c_i'\left(q_i^{-1}(0)\right)}{q_i'\left(q_i^{-1}(0)\right)}$$

$$= 0, \tag{A.4}$$

$$h_i''(\cdot) = \underbrace{\frac{2c_i''\left(q_i^{-1}(t)\right)}{q_i'\left(q_i^{-1}(t)\right)^2}}_{>0} - \underbrace{\frac{2q_i''\left(q_i^{-1}(t)\right) \cdot c_i'\left(q_i^{-1}(t)\right)}{q_i'\left(q_i^{-1}(t)\right)^3}}_{<0}$$

$$> 0. \tag{A.5}$$

Denote $h_1(\cdot) - h_2(\cdot)$ by $\Delta(\cdot)$. By (A.3), $\Delta(0) = 0$. By (A.4), $\Delta'(0) = 0$.

Suppose that $W_1 \succeq W_2$. Then $\Delta''(\cdot) \leq 0$, and, because $\Delta'(0) = 0$, $\Delta'(\cdot) \leq 0$ throughout its domain. This fact, combined with the fact that $\Delta(0) = 0$, leads to the conclusion that $\Delta(\cdot) \leq 0$. Thus

$$\phi_1 = h_1(\phi_1)$$

$$= h_1(\phi_1) - h_2(\phi_1) + h_2(\phi_1)$$
$$= \Delta(\phi_1) + h_2(\phi_1)$$
$$\leq h_2(\phi_1).$$

By Lemma A.2, $\phi_1 \geq \phi_2$. Corollary 1.2, therefore, implies that the principal has a weak preference for W_1.

Now suppose that $W_1 \succ W_2$. The inequalities used in the argument above remain valid but are now strict. In particular, $\Delta''(\cdot) < 0$, and so $\Delta'(t) < 0$ for $t > 0$, from which it follows that $\Delta(t) < 0$ for $t > 0$. Thus $\phi_1 < h_2(\phi_1)$, from which it follows that $\phi_1 > \phi_2$. By Corollary 1.2, the principal has a strict preference for W_2. The result follows. $\square$

Lemma 1.3. *Let $t_W \equiv (\pi_W, \lambda_W, q(\cdot), c(\cdot))$ be a worker type. The function $e \mapsto -c(e) - \frac{\xi(t_W)}{2} + \xi(t_W)q(e)$, defined on $[0, \infty)$, has a unique maximizer, denoted by $\hat{e}(t_W)$, that is positive and is characterized by the condition $c'(\hat{e}(t_W)) = \xi(t_W)q'(\hat{e}(t_W))$.*

Proof. Because $q'(0) > 0$, $c'(0) = 0$, and $\xi(t_W) > 0$, the objective function is increasing at $e = 0$. Thus, if a maximizer exists, it must be positive. Because the objective function is strictly concave and twice differentiable, if a solution exists to the first-order condition $c'(\hat{e}(t_W)) = \xi(t_W)q'(\hat{e}(t_W))$, it will be both the unique solution and the unique maximizer. Since $\lim_{e \to \infty} q'(e) = 0$ and $c'(\cdot) > 0$, $\xi(t_W)q'(e) - c'(e) < 0$ for values of e that are sufficiently large. On the other hand, $\xi(t_W)q'(0) - c'(0) > 0$. Because both $q'(\cdot)$ and $c'(\cdot)$ are continuous, the Intermediate Value Theorem implies that a solution exists to the first-order condition. $\square$

Lemma 1.4. *Consider a worker of type $t_W \equiv (\pi_W, \lambda_W, q(\cdot), c(\cdot)) \in \mathcal{T}_W$, and suppose that $\phi_\alpha(t_W) = 0$. Then the worker is biased toward β and prefers the outcome in which he shirks and β is implemented unconditionally to any outcome in which he informs the decision. Similarly, if $\phi_\beta(t_W) = 0$, the worker is biased toward α and prefers the outcome in which he shirks and α is implemented unconditionally to any outcome in which he informs the decision.*

Proof. Suppose that $\phi_\alpha(t_W) = 0$. By Definitions A.1 and A.2, the function $g\left(\cdot; t_W, -\frac{\tau(t_W)}{2\xi(t_W)}\right)$ has no fixed point. Then, by Lemma A.4, both of the following relationships hold:

$$-\frac{\tau(t_W)}{2\xi(t_W)} > 0, \tag{A.6}$$

$$q(\hat{e}(t_W)) < -\frac{\tau(t_W)}{2\xi(t_W)} + \frac{c(\hat{e}(t_W))}{\xi(t_W)}. \tag{A.7}$$

From (A.6), it follows that $\tau(t_W) < 0$, meaning that the worker is biased toward β. From (A.7), it follows that

$$-c(\hat{e}(t_W)) - \frac{\xi(t_W)}{2} + \xi(t_W)q(\hat{e}(t_W)) < -\frac{\xi(t_W)}{2} - \frac{\tau(t_W)}{2}. \tag{A.8}$$

That is, the worker obtains a higher payoff (the right hand side of (A.8)) from the outcome in which he shirks and β is implemented unconditionally than when he exerts his optimal effort level under a policy of rubberstamping. The result for the $\phi_\alpha(t_W)$ case follows. The proof for the case in which $\phi_\beta(t_W) = 0$ follows an analogous argument. $\qquad\square$

Proposition 1.3. *Suppose that the principal's type is $(\pi, \lambda) \in (0, 1)^2$, and that she appoints a manager of type $\pi_M \in \mathcal{T}_M$ and a worker of type $t_W \equiv (\pi_W, \lambda_W, q(\cdot), c(\cdot)) \in \mathcal{T}_W$. The principal's appointments can produce any of three types of outcomes, which are determined by a mutually exclusive and exhaustive set of conditions:*

Skepticism does not affect worker's incentives. *Suppose that all of the following conditions hold:*

- $0 \leq \left|\pi_M - \frac{1}{2}\right| \leq q\left(\hat{e}(t_W)\right)$;
- $\pi_M - \frac{1}{2} \leq \phi_\beta(t_W)$;
- $\frac{1}{2} - \pi_M \leq \phi_\alpha(t_W)$.

Then the manager rubberstamps the worker's signal. The worker's effort level and the principal's expected utility, respectively, are

$$e_d^*(\pi_M\,;t_W) = \hat{e}(t_W),$$

$$U_d^*(t_W, \pi_M\,;\pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q\left(\hat{e}(t_W)\right)\right).$$

Skepticism strengthens worker's incentives. *Suppose that either of the following conditions holds:*

- $q(\hat{e}(t_W)) < \pi_M - \frac{1}{2} \leq \phi_\beta(t_W)$;
- $q(\hat{e}(t_W)) < \frac{1}{2} - \pi_M \leq \phi_\alpha(t_W)$.

Then the manager rubberstamps the worker's signal, and

$$e_d^*(\pi_M\,;t_W) = q^{-1}\left(\left|\pi_M - \frac{1}{2}\right|\right),$$

$$U_d^*(t_W, \pi_M\,;\pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \left|\pi_M - \frac{1}{2}\right|\right).$$

Skepticism destroys worker's incentives. *Suppose that either of the following conditions holds:*

- $\pi_M - \frac{1}{2} > \phi_\beta(t_W)$;
- $\frac{1}{2} - \pi_M > \phi_\alpha(t_W)$.

Then the manager chooses her preferred action, and

$$e_d^*(\pi_M \,; t_W) = 0,$$

$$U_d^*(t_W, \pi_M \,; \pi, \lambda) = \begin{cases} \pi\lambda - \lambda & \text{if } \pi_M > \frac{1}{2}, \\ \pi\lambda - \pi & \text{if } \pi_M < \frac{1}{2}. \end{cases}$$

Proof. First, I show that the conditions are mutually exclusive and exhaustive, as claimed. The first set of conditions holds if and only if both of the following two conditions hold:

$$\pi_M - \frac{1}{2} \leq \min\left\{q(\hat{e}(t_W)), \phi_\beta(t_W)\right\}, \tag{A.9}$$

$$\frac{1}{2} - \pi_M \leq \min\left\{q(\hat{e}(t_W)), \phi_\alpha(t_W)\right\}. \tag{A.10}$$

Since $q(\hat{e}(t_W)) > 0$ by Lemma 1.3 and $\phi_\eta(t_W) \geq 0$ for $\eta \in \{\alpha, \beta\}$, the right hand sides of both (A.9) and (A.10) are nonnegative. Hence, at least one of the two conditions must hold. Suppose that (A.9) fails. Then it must be the case either that

$$q(\hat{e}(t_W)) < \pi_M - \frac{1}{2} \leq \phi_\beta(t_W), \tag{A.11}$$

which is the second case from the statement of the result, or that

$$\pi_M - \frac{1}{2} > \phi_\beta(t_W), \tag{A.12}$$

which is the third case from the statement of the result. It is clear that (A.9), (A.11), and (A.12) are mutually exclusive. Similarly, if (A.10) fails, then either of two (mutually exclusive) analogues of (A.11) and (A.12) will hold. These analogues correspond, respectively, to the second and third cases from the statement of the result. Thus, the conditions are mutually exclusive and exhaustive.

Suppose that the manager is unbiased: $\pi_M = \frac{1}{2}$. Then it is clear that the conditions from the first case are met. Furthermore, by (1.4), the worker's expected utility, as a function of his effort level, is the function from Lemma 1.3. Thus the worker's effort level, $e_d^*(\pi_M \,; t_W)$, is $\hat{e}(t_W)$. A direct computation shows that the principal's expected utility, $U_d^*(t_W, \pi_M \,; \pi, \lambda)$, is as stated in the result.

Suppose that the manager is skeptical regarding β: $\pi_M > \frac{1}{2}$. Then the worker's objective function can be written as

$$U_W(e \,; \pi_M, t_W) \equiv \begin{cases} -c(e) - \frac{\xi(t_W)}{2} + \frac{\tau(t_W)}{2} & \text{if } e < q^{-1}(\left|\pi_M - \frac{1}{2}\right|), \\ -c(e) - \frac{\xi(t_W)}{2} + \xi(t_W)q(e) & \text{otherwise.} \end{cases} \tag{A.13}$$

Observe that

$$\arg\max_{e \geq 0} U_W(e \,; \pi_M, t_W) = \arg\max_{e \in M} U_W(e \,; \pi_M, t_W),$$

where

$$M \equiv$$

$$\underbrace{\left[\underset{e \in \left[0, q^{-1}\left(\pi_M - \frac{1}{2}\right)\right)}{\arg\max} -c(e) - \frac{\xi(t_W)}{2} + \frac{\tau(t_W)}{2} \right]}_{=\{0\}} \cup \left[\underset{e \in \left[q^{-1}\left(\pi_M - \frac{1}{2}\right), \infty\right)}{\arg\max} -c(e) - \frac{\xi(t_W)}{2} + \xi(t_W)q(e) \right].$$

Let $M_1 \equiv \arg\max_{e \in \left[q^{-1}\left(\pi_M - \frac{1}{2}\right), \infty\right)} -c(e) - \frac{\xi(t_W)}{2} + \xi(t_W)q(e)$. Since $U_W(0\,;\pi_M, t_W) = -\frac{\xi(t_W)}{2} + \frac{\tau(t_W)}{2}$,

$$\underset{e \geq 0}{\arg\max}\, U_W(e\,;\pi_M, t_W) = \begin{cases} \{0\} & \text{if } \max_{e \in \left[q^{-1}\left(\pi_M - \frac{1}{2}\right), \infty\right)} -c(e) + \xi(t_W)q(e) < \frac{\tau(t_W)}{2}, \\ M_1 & \text{if } \max_{e \in \left[q^{-1}\left(\pi_M - \frac{1}{2}\right), \infty\right)} -c(e) + \xi(t_W)q(e) > \frac{\tau(t_W)}{2}, \\ \{0\} \cup M_1 & \text{otherwise,} \end{cases}$$

and

$$e_d^*(\pi_M\,;t_W) = \begin{cases} 0 & \text{if } \max_{e \in \left[q^{-1}\left(\pi_M - \frac{1}{2}\right), \infty\right)} -c(e) + \xi(t_W)q(e) < \frac{\tau(t_W)}{2}, \\ \max M_1 & \text{otherwise.} \end{cases} \tag{A.14}$$

Since $-c(e) - \frac{\xi(t_W)}{2} + \xi(t_W)q(e)$ is increasing for $e \in [0, \hat{e}(t_W))$ and decreasing for $e \in (\hat{e}(t_W), \infty)$,

$$\max M_1 = \begin{cases} \hat{e}(t_W) & \text{if } \pi_M - \frac{1}{2} \leq q(\hat{e}(t_W)), \\ q^{-1}\left(\pi_M - \frac{1}{2}\right) & \text{otherwise,} \end{cases} \tag{A.15}$$

and

$$\underset{e \in \left[q^{-1}\left(\pi_M - \frac{1}{2}\right), \infty\right)}{\max} -c(e) + \xi(t_W)q(e) =$$

$$\begin{cases} -c(\hat{e}(t_W)) + \xi(t_W)q(\hat{e}(t_W)) & \text{if } \pi_M - \frac{1}{2} \leq q(\hat{e}(t_W)), \\ -c\left(q^{-1}\left(\pi_M - \frac{1}{2}\right)\right) + \xi(t_W) \cdot \left(\pi_M - \frac{1}{2}\right) & \text{otherwise.} \end{cases} \tag{A.16}$$

Combining (A.14), (A.15), and (A.16) yields

$$e_d^*(\pi_M\,;t_W) =$$

$$\begin{cases} \hat{e}(t_W) & \text{if } -c(\hat{e}(t_W)) + \xi(t_W)q(\hat{e}(t_W)) \geq \frac{\tau(t_W)}{2} \text{ and } \pi_M - \frac{1}{2} \leq q(\hat{e}(t_W)), \\ q^{-1}\left(\pi_M - \frac{1}{2}\right) & \text{if } -c\left(q^{-1}\left(\pi_M - \frac{1}{2}\right)\right) + \xi(t_W) \cdot \left(\pi_M - \frac{1}{2}\right) \geq \frac{\tau(t_W)}{2} \\ & \quad \text{and } \pi_M - \frac{1}{2} > q(\hat{e}(t_W)), \\ 0 & \text{otherwise.} \end{cases} \tag{A.17}$$

Suppose that the first case holds: $0 < \pi_M - \frac{1}{2} \leq \min\{q(\hat{e}(t_W)), \phi_\beta(t_W)\}$. Then $\pi_M - \frac{1}{2} \leq q(\hat{e}(t_W))$ and $\phi_\beta(t_W) > 0$. By Lemma A.6, $\phi_\beta(t_W) \geq q(\hat{e}(t_W))$. If $\phi_\beta(t_W) > q(\hat{e}(t_W))$, then, by Lemma A.5, $\tau(t_W) \leq 0$ or $q(\hat{e}(t_W)) > \frac{\tau(t_W)}{2\xi(t_W)} + \frac{c(\hat{e}(t_W))}{\xi(t_W)}$. In either case, $-c(\hat{e}(t_W)) + \xi(t_W)q(\hat{e}(t_W)) \geq \frac{\tau(t_W)}{2}$. If $\phi_\beta(t_W) = q(\hat{e}(t_W))$, then, by the definition of $\phi_\beta(t_W)$, $q(\hat{e}(t_W)) = \frac{\tau(t_W)}{2\xi(t_W)} + \frac{c(\hat{e}(t_W))}{\xi(t_W)}$. Equivalently, $-c(\hat{e}(t_W)) + \xi(t_W)q(\hat{e}(t_W)) = \frac{\tau(t_W)}{2}$. By (A.17), $e_d^*(\pi_M\,;t_W) = \hat{e}(t_W)$. Since the worker's effort level exceeds the manager's skepticism, the manager rubberstamps the worker's signal. A direct computation confirms that the principal's expected utility is $(2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q\left(\hat{e}(t_W)\right)\right)$.

Now suppose that the second case holds: $q(\hat{e}(t_W)) < \pi_M - \frac{1}{2} \leq \phi_\beta(t_W)$. Then, by Lemma A.7,

$$\frac{\tau(t_W)}{2\xi(t_W)} + \frac{c\left(q^{-1}\left(\pi_M - \frac{1}{2}\right)\right)}{\xi(t_W)} \leq \pi_M - \frac{1}{2},$$

which implies that

$$-c\left(q^{-1}\left(\pi_M - \frac{1}{2}\right)\right) + \xi(t_W) \cdot \left(\pi_M - \frac{1}{2}\right) \geq \frac{\tau(t_W)}{2}.$$

By (A.17), $e_d^* = q^{-1}\left(\pi_M - \frac{1}{2}\right)$: the worker exerts just enough effort to influence the manager. The principal's expected utility can be computed directly as $(2\pi\lambda - \pi - \lambda) \cdot (1 - \pi_M)$.

Finally, suppose that the third case holds: $\pi_M - \frac{1}{2} > \phi_\beta(t_W)$. By Lemma A.7,

$$\frac{\tau(t_W)}{2\xi(t_W)} + \frac{c\left(q^{-1}\left(\pi_M - \frac{1}{2}\right)\right)}{\xi(t_W)} > \pi_M - \frac{1}{2},$$

which implies that

$$-c\left(q^{-1}\left(\pi_M - \frac{1}{2}\right)\right) + \xi(t_W) \cdot \left(\pi_M - \frac{1}{2}\right) < \frac{\tau(t_W)}{2}. \tag{A.18}$$

There are two subcases to consider:

Subcase 1: $\phi_\beta(t_W) = 0$. In this case, Lemma A.7 implies that $q(\hat{e}(t_W)) < \frac{\tau(t_W)}{2\xi(t_W)} + \frac{c(\hat{e}(t_W))}{\xi(t_W)}$; hence

$$-c(\hat{e}(t_W)) + \xi(t_W)q(\hat{e}(t_W)) < \frac{\tau(t_W)}{2}. \tag{A.19}$$

By (A.17), (A.18), and (A.19), $e_d^*(\pi_M\,;t_W) = 0$.

Subcase 2: $\phi_\beta(t_W) > 0$. In this case, Lemma A.4 shows that $\phi_\beta(t_W) \geq q(\hat{e}(t_W))$, and so

$$\pi_M - \frac{1}{2} > \phi_\beta(t_W) \geq q(\hat{e}(t_W)). \tag{A.20}$$

By (A.17), (A.18), and (A.20), $e_d^*(\pi_M\,;t_W) = 0$.

Since the worker exerts no effort, the manager chooses her ex ante preferred action, α. The principal's expected utility, therefore, is $-\lambda \cdot (1 - \pi) = \pi\lambda - \lambda$.

The proof for the case in which $\pi_M < \frac{1}{2}$ follows an analogous argument. $\square$

Lemma 1.5. *For any principal type $(\pi, \lambda) \in (0,1)^2$, given $q \in \left(\frac{|\pi-\lambda|}{2\pi+2\lambda-4\pi\lambda}, \frac{1}{2} \right)$, the principal prefers the decision to be informed by a signal of quality q than for either action to be implemented unconditionally. On the other hand, if $q \in \left[0, \frac{|\pi-\lambda|}{2\pi+2\lambda-4\pi\lambda} \right)$, the principal prefers the unconditional implementation of her ex ante preferred action.*

Proof. By the proof of Lemma 1.1, a principal of type (π, λ) has aligned preferences with a manager of type $\frac{\pi-\pi\lambda}{\pi+\lambda-2\pi\lambda}$. This manager's skepticism is $\left| \frac{\pi-\pi\lambda}{\pi+\lambda-2\pi\lambda} - \frac{1}{2} \right| = \frac{|\pi-\lambda|}{2\pi+2\lambda-4\pi\lambda}$. The result now follows from Lemma 1.2. $\square$

Corollary 1.4. *Suppose that the principal's type is $(\pi, \lambda) \in (0,1)^2$, and that she appoints a worker of type $t_W \equiv (\pi_W, \lambda_W, q(\cdot), c(\cdot)) \in \mathcal{T}_W$ but retains decision making authority. The principal's appointment of the worker can produce any of three types of outcomes, which are determined by a mutually exclusive and exhaustive set of conditions:*

Informational standard does not affect worker's incentives. *Suppose that all of the following conditions hold:*

- $0 \leq \frac{|\pi-\lambda|}{2\pi+2\lambda-4\pi\lambda} \leq q\left(\hat{e}(t_W)\right)$;
- $\frac{\pi-\lambda}{2\pi+2\lambda-4\pi\lambda} \leq \phi_\beta(t_W)$;
- $\frac{\lambda-\pi}{2\pi+2\lambda-4\pi\lambda} \leq \phi_\alpha(t_W)$.

Then the principal rubberstamps the worker's signal. The worker's effort level and the principal's expected utility, respectively, are

$$e_r^*(\pi, \lambda \, ; t_W) = \hat{e}(t_W),$$

$$U_r^*(t_W \, ; \pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q\left(\hat{e}(t_W)\right) \right).$$

Informational standard strengthens worker's incentives. *Suppose that either of the following conditions holds:*

- $q(\hat{e}(t_W)) < \frac{\pi-\lambda}{2\pi+2\lambda-4\pi\lambda} \leq \phi_\beta(t_W)$;
- $q(\hat{e}(t_W)) < \frac{\lambda-\pi}{2\pi+2\lambda-4\pi\lambda} \leq \phi_\alpha(t_W)$.

Then the principal rubberstamps the worker's signal, and

$$e_r^*(\pi, \lambda \, ; t_W) = q^{-1}\left(\frac{|\pi-\lambda|}{2\pi+2\lambda-4\pi\lambda} \right),$$

$$U_r^*(t_W \, ; \pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \frac{|\pi-\lambda|}{2\pi+2\lambda-4\pi\lambda} \right).$$

Informational standard destroys worker's incentives. *Suppose that either of the following conditions holds:*

- $\frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda} > \phi_\beta(t_W);$
- $\frac{\lambda - \pi}{2\pi + 2\lambda - 4\pi\lambda} > \phi_\alpha(t_W).$

Then the principal chooses her preferred action, and

$$e_r^*(\pi, \lambda\,; t_W) = 0,$$

$$U_r^*(t_W\,; \pi, \lambda) = \begin{cases} \pi\lambda - \lambda & \text{if } \pi > \lambda, \\ \pi\lambda - \pi & \text{if } \pi < \lambda. \end{cases}$$

Proof. The proof of this result is analogous to that of Proposition 1.3. $\qquad\square$

Corollary 1.5. *Suppose that*

$$\max_{t_W = (\pi_W, \lambda_W, q(\cdot), c(\cdot)) \in \mathcal{T}_W} q(\hat{e}(t_W))$$

exists and that there exists a self-driven candidate worker, of type $t_W^* = (\pi_W^*, \lambda_W^*, q^*(\cdot), c^*(\cdot))$, such that

$$t_W^* \in \arg\max_{t_W = (\pi_W, \lambda_W, q(\cdot), c(\cdot)) \in \mathcal{T}_W} q(\hat{e}(t_W)).$$

For any candidate manager π_M^*, (t_W^*, π_M^*) is a solution to (1.6).

Proof. Let (t_W^*, π_M^*) satisfy the above conditions. Given that π_M^* can be any candidate manager from $\mathcal{T}_M$, it suffices to establish the following two facts:

(i) $U_d^*(t_W^*, \pi_M\,; \pi, \lambda) = U_d^*(t_W^*, \pi_M'\,; \pi, \lambda)$ for every $\pi_M, \pi_M' \in \mathcal{T}_M$.

(ii) $U_d^*(t_W^*, \pi_M\,; \pi, \lambda) \geq U_d^*(t_W, \pi_M\,; \pi, \lambda)$ for every $t_W \in \mathcal{T}_W$ and $\pi_M \in \mathcal{T}_M$.

(i) follows immediately from the assumption that t_W^* is self-driven. To establish (ii), let $t_W = (\pi_W, \lambda_W, q(\cdot), c(\cdot))$ and π_M be given. By Proposition 1.3, $U_d^*(t_W, \pi_M\,; \pi, \lambda)$ can take one of three forms. First, it is possible that the manager of type π_M has no effect on the incentives of the worker of type t_W, in which case

$$\begin{aligned} U_d^*(t_W, \pi_M\,; \pi, \lambda) &= (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q(\hat{e}(t_W)) \right) \\ &\leq (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q^*(\hat{e}(t_W^*)) \right) \ \text{ since } q(\hat{e}(t_W)) \leq q^*(\hat{e}(t_W^*)) \\ &= U_d^*(t_W^*, \pi_M\,; \pi, \lambda). \end{aligned}$$

Second, it is possible that the manager strengthens the worker's incentives, in which case

$$U_d^* (t_W, \pi_M ; \pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \left| \pi_M - \frac{1}{2} \right| \right)$$

$$\leq (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q^*(\hat{e}(t_W^*)) \right) \quad \text{since } t_W^* \text{ is self-driven}$$

$$= U_d^* (t_W^*, \pi_M ; \pi, \lambda).$$

The only remaining possibility is that the manager destroys the worker's incentives, in which case

$$U_d^* (t_W, \pi_M ; \pi, \lambda) \leq \pi\lambda - \min\{\pi, \lambda\}$$

$$\leq (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q^*(\hat{e}(t_W^*)) \right) \tag{A.21}$$

$$= U_d^* (t_W^*, \pi_M ; \pi, \lambda),$$

where (A.21) follows from the fact that, since the worker is self-driven, the quality of signal that he obtains exceeds the principal's informational standard:

$$q^*(\hat{e}(t_W^*)) \geq \sup_{\pi_M' \in \mathcal{T}_M} \left| \pi_M' - \frac{1}{2} \right| \geq \left| \frac{\pi - \pi\lambda}{\pi + \lambda - 2\pi\lambda} - \frac{1}{2} \right| = \frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda}.$$

Since (ii) is satisfied in all three cases, the result holds. $\square$

Corollary 1.6. *If all candidate workers are underqualified, the principal should retain control of the decision. In particular, let*

$$t_W^* \in \mathcal{T}_W,$$

$$\pi_M^* \equiv \frac{\pi\lambda - \pi}{2\pi\lambda - \pi - \lambda}.$$

The pair (t_W^*, π_M^*) is a solution to (1.6).

Proof. Suppose that all candidate workers are underqualified. Proposition 1.3 implies that $e_d^*(\pi_M ; t_W) = 0$ for every $\pi_M \in \mathcal{T}_M$ and $t_W \in \mathcal{T}_W$. Thus, the principal should appoint a manager that favors the alternative toward which she herself is biased. (Her choice of worker will not matter, since no worker exerts effort.) Thus, it is optimal for the principal to retain control of the decision. $\square$

Corollary 1.7. *Suppose that the pool of candidate managers is sufficient to strengthen incentives. Let $\mathcal{T}^*$ denote the set of pairs of candidate workers and managers in which the candidate manager would strengthen the candidate worker's incentives if both were appointed:*

$$\left\{ (t_W, \pi_M) \in \mathcal{T}_W \times \mathcal{T}_M : q(\hat{e}(t_W)) < \pi_M - \frac{1}{2} \leq \phi_\beta(t_W) \text{ or } q(\hat{e}(t_W)) < \frac{1}{2} - \pi_M \leq \phi_\alpha(t_W) \right\}.$$

If

$$\max_{(t_W,\pi_M)\in\mathcal{T}^*}\left|\pi_M-\frac{1}{2}\right|$$

exists and is no less than the principal's informational standard, then any pair

$$(t_W^*,\pi_M^*)\in\arg\max_{(t_W,\pi_M)\in\mathcal{T}^*}\left|\pi_M-\frac{1}{2}\right|$$

is a solution to (1.6).

Proof. Let $t_W^* = (\pi_W^*,\lambda_W^*,q^*(\cdot),c^*(\cdot))$ and π_M^* satisfy the stated assumptions. Proposition 1.3 implies that $U_d^*\left(t_W^*,\pi_M^*\,;\pi,\lambda\right) = (2\pi\lambda-\pi-\lambda)\cdot\left(\frac{1}{2}-\left|\pi_M^*-\frac{1}{2}\right|\right)$. Consider any other appointment $(t_W,\pi_M)\in\mathcal{T}_W\times\mathcal{T}_M$. By Proposition 1.3, $U_d^*\left(t_W,\pi_M\,;\pi,\lambda\right)$ can take one of three forms. If the manager of type π_M has no effect on the incentives of the worker of type t_W, then, because $\mathcal{T}_M$ is sufficient to strengthen incentives, there exists $\pi_M'\in\mathcal{T}_M$ such that $(t_W,\pi_M')\in\mathcal{T}^*$. Hence

$$U_d^*\left(t_W,\pi_M\,;\pi,\lambda\right) = (2\pi\lambda-\pi-\lambda)\cdot\left(\frac{1}{2}-q(\hat{e}(t_W))\right)$$

$$< (2\pi\lambda-\pi-\lambda)\cdot\left(\frac{1}{2}-\left|\pi_M'-\frac{1}{2}\right|\right)$$

$$\leq (2\pi\lambda-\pi-\lambda)\cdot\left(\frac{1}{2}-\left|\pi_M^*-\frac{1}{2}\right|\right)$$

$$= U_d^*\left(t_W^*,\pi_M^*\,;\pi,\lambda\right).$$

If the manager of type π_M strengthens the worker's incentives, then $(t_W,\pi_M)\in\mathcal{T}^*$, and so

$$U_d^*\left(t_W,\pi_M\,;\pi,\lambda\right) = (2\pi\lambda-\pi-\lambda)\cdot\left(\frac{1}{2}-\left|\pi_M'-\frac{1}{2}\right|\right)$$

$$\leq (2\pi\lambda-\pi-\lambda)\cdot\left(\frac{1}{2}-\left|\pi_M^*-\frac{1}{2}\right|\right)$$

$$= U_d^*\left(t_W^*,\pi_M^*\,;\pi,\lambda\right).$$

Finally, if the manager destroys the worker's incentives, then, since

$$\left|\pi_M^*-\frac{1}{2}\right|\geq\frac{|\pi-\lambda|}{2\pi+2\lambda-4\pi\lambda}$$

by assumption,

$$U_d^*\left(t_W,\pi_M\,;\pi,\lambda\right)\leq\pi\lambda-\min\{\pi,\lambda\}$$

$$\leq (2\pi\lambda-\pi-\lambda)\cdot\left(\frac{1}{2}-\left|\pi_M^*-\frac{1}{2}\right|\right)$$

$$= U_d^*\left(t_W^*,\pi_M^*\,;\pi,\lambda\right).$$

The result follows. $\square$

Corollary 1.8. *Suppose that the pool of candidate managers is sufficient to strengthen incentives and that there exist $t_W \in \mathcal{T}_W$ and $\pi_M \in \mathcal{T}_M$ such that either of the following conditions holds:*

- $\frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda} < \pi_M - \frac{1}{2} \le \phi_\beta(t_W)$.

- $\frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda} < \frac{1}{2} - \pi_M \le \phi_\alpha(t_W)$.

It is suboptimal for the principal to retain control of the decision.

Proof. Consider any candidate worker type $t'_W = (\pi'_W, \lambda'_W, q'(\cdot), c'(\cdot)) \in \mathcal{T}_W$. By Corollary 1.4, the principal's expected utility from appointing this candidate worker and retaining control of the decision, $U_r^*(t'_W \,;\, \pi, \lambda)$, takes one of three forms. If the principal's informational standard does not affect the worker's incentives, then

$$U_r^*(t'_W \,;\, \pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q'(\hat{e}(t'_W)) \right).$$

Because $\mathcal{T}_M$ is sufficient to strengthen incentives, there exists a candidate manager type, $\pi'_M \in \mathcal{T}_M$, such that

$$
\begin{aligned}
U_d^*(t'_W, \pi'_M \,;\, \pi, \lambda) &= (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \left| \pi'_M - \frac{1}{2} \right| \right) \\
&> (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q'(\hat{e}(t'_W)) \right) \\
&= U_r^*(t'_W \,;\, \pi, \lambda).
\end{aligned}
$$

If the principal's informational standard strengthens the worker's incentives, then

$$U_r^*(t'_W \,;\, \pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda} \right).$$

On the other hand, if the principal's informational standard destroys the worker's incentives,

$$U_r^*(t'_W \,;\, \pi, \lambda) = \pi\lambda - \min\{\pi, \lambda\}.$$

In either case, by assumption and by Proposition 1.3,

$$
\begin{aligned}
U_d^*(t_W, \pi_M \,;\, \pi, \lambda) &= (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \left| \pi_M - \frac{1}{2} \right| \right) \\
&> U_r^*(t'_W \,;\, \pi, \lambda).
\end{aligned}
$$

The result follows. $\qquad\square$

Corollary 1.9. *Suppose that $\mathcal{T}_M$ is a dense subset of $(0,1)$. Consider the worker and manager pair (t_W, π_M), where $\tau(t_W) \cdot \left(\pi_M - \frac{1}{2} \right) > 0$ and $\phi(t_W) > \frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda}$. There exists $\pi'_M \in \mathcal{T}_M$ such that $\tau(t_W) \cdot \left(\pi'_M - \frac{1}{2} \right) < 0$ and $U_d^*(t_W, \pi_M \,;\, \pi, \lambda) < U_d^*(t_W, \pi'_M \,;\, \pi, \lambda)$.*

Proof. Let $t_W = (\pi_W, \lambda_W, q(\cdot), c(\cdot))$. First, note that $\tau(t_W) \cdot \left(\pi_M - \frac{1}{2}\right) > 0$ implies both that $\tau(t_W) \neq 0$ and that $\pi_M \neq \frac{1}{2}$. Suppose that $\tau(t_W) > 0$, so that $\pi_M > \frac{1}{2}$. By Proposition 1.3,

$$U_d^*(t_W, \pi_M ; \pi, \lambda) = \begin{cases} (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q(\hat{e}(t_W))\right) & \text{if } 0 < \pi_M - \frac{1}{2} \leq q(\hat{e}(t_W)), \\ (2\pi\lambda - \pi - \lambda) \cdot (1 - \pi_M) & \text{if } q(\hat{e}(t_W)) < \pi_M - \frac{1}{2} \leq \phi_\beta(t_W), \\ \pi\lambda - \lambda & \text{if } \phi_\beta(t_W) < \pi_M - \frac{1}{2} < \frac{1}{2}. \end{cases}$$

Suppose that $\pi_M - \frac{1}{2} \leq \phi_\beta(t_W)$. By Lemmas A.6 and A.8, $q(\hat{e}(t_W)) \leq \phi_\beta(t_W) < \phi_\alpha(t_W)$. Because $\mathcal{T}_M$ is dense in $(0,1)$, there exists $\pi'_M \in \mathcal{T}_M \cap \left[\frac{1}{2} - \phi_\alpha(t_W), \frac{1}{2} - \phi_\beta(t_W)\right)$. Note that $q(\hat{e}(t_W)) \leq \phi_\beta(t_W) < \frac{1}{2} - \pi'_M \leq \phi_\alpha(t_W)$. Thus $\pi'_M < \frac{1}{2}$, and

$$U_d^*(t_W, \pi_M ; \pi, \lambda) \leq (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \phi_\beta(t_W)\right)$$
$$< (2\pi\lambda - \pi - \lambda) \cdot \pi'_M$$
$$= U_d^*(t_W, \pi'_M ; \pi, \lambda).$$

Now suppose that $\pi_M - \frac{1}{2} > \phi_\beta(t_W)$. By Lemma A.8, $\phi_\alpha(t_W) = \phi(t_W) > \frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda}$. Furthermore, since $\mathcal{T}_M$ is dense in $(0,1)$, there exists $\pi'_M \in \mathcal{T}_M \cap \left[\frac{1}{2} - \phi_\alpha(t_W), \frac{1}{2} - \frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda}\right)$. Note that $\max\left\{q(\hat{e}(t_W)), \frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda}\right\} < \frac{1}{2} - \pi'_M \leq \phi_\alpha(t_W)$. Once again, $\pi'_M < \frac{1}{2}$, and

$$U_d^*(t_W, \pi_M ; \pi, \lambda) = \pi\lambda - \lambda$$
$$< (2\pi\lambda - \pi - \lambda) \cdot \pi'_M$$
$$= U_d^*(t_W, \pi'_M ; \pi, \lambda).$$

Thus the result holds whenever $\tau(t_W) > 0$. An analogous argument establishes the result for the case in which $\tau(t_W) < 0$. □

Lemma 1.6. *Consider a candidate manager whose skepticism is no greater than the principal's informational standard, i.e.,*

$$\left|\pi_M - \frac{1}{2}\right| \leq \frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda}.$$

Then, for any candidate worker type, t_W, $U_d^*(t_W, \pi_M ; \pi, \lambda) \leq U_r^*(t_W ; \pi, \lambda)$.

Proof. Suppose that $\pi \geq \lambda$. Then $\left|\pi_M - \frac{1}{2}\right| \leq \frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda}$, so both $\pi_M - \frac{1}{2} \leq \frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda}$ and $\frac{1}{2} - \pi_M \leq \frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda}$. Consider the following cases:

Case 1: Suppose that both of the following conditions hold:

- $0 \leq \frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda} \leq q(\hat{e}(t_W))$;
- $\frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda} \leq \phi_\beta(t_W)$.

Then both of the following conditions hold as well:

- $0 \leq \left|\pi_M - \frac{1}{2}\right| \leq q(\hat{e}(t_W))$;
- $\pi_M - \frac{1}{2} \leq \phi_\beta(t_W)$.

Suppose that $\frac{1}{2} - \pi_M > \phi_\alpha(t_W)$. Then the worker shirks, and the manager chooses β. In particular, by Proposition 1.3 and Corollary 1.4,

$$U_d^*(t_W, \pi_M \,; \pi, \lambda) = \pi\lambda - \pi$$
$$\leq \pi\lambda - \frac{\pi + \lambda}{2}$$
$$= (2\pi\lambda - \pi - \lambda) \cdot \frac{1}{2}$$
$$< (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q(\hat{e}(t_W))\right)$$
$$= U_r^*(t_W \,; \pi, \lambda).$$

On the other hand, if $\frac{1}{2} - \pi_M \leq \phi_\alpha(t_W)$, then

$$U_d^*(t_W, \pi_M \,; \pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q(\hat{e}(t_W))\right)$$
$$= U_r^*(t_W \,; \pi, \lambda).$$

Case 2: Suppose that $q(\hat{e}(t_W)) < \frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda} \leq \phi_\beta(t_W)$. If $0 \leq \pi_M - \frac{1}{2} \leq q(\hat{e}(t_W))$, then

$$U_d^*(t_W, \pi_M \,; \pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q(\hat{e}(t_W))\right)$$
$$< (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda}\right)$$
$$= U_r^*(t_W \,; \pi, \lambda).$$

Similarly, if $q(\hat{e}(t_W)) < \pi_M - \frac{1}{2} \leq \frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda} \leq \phi_\beta(t_W)$, then

$$U_d^*(t_W, \pi_M \,; \pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \left(\pi_M - \frac{1}{2}\right)\right)$$
$$\leq (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda}\right)$$
$$= U_r^*(t_W \,; \pi, \lambda).$$

If $0 < \frac{1}{2} - \pi_M \leq q(\hat{e}(t_W))$, then

$$U_d^*(t_W, \pi_M \,; \pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q(\hat{e}(t_W))\right)$$

$$< (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda} \right)$$

$$= U_r^*(t_W \,;\, \pi, \lambda).$$

On the other hand, if $q(\hat{e}(t_W)) < \frac{1}{2} - \pi_M \leq \phi_\alpha(t_W)$, then

$$U_d^*(t_W, \pi_M \,;\, \pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \left(\frac{1}{2} - \pi_M \right) \right)$$

$$\leq (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda} \right)$$

$$= U_r^*(t_W \,;\, \pi, \lambda).$$

Finally, if $\frac{1}{2} - \pi_M > \phi_\alpha(t_W)$, then

$$U_d^*(t_W, \pi_M \,;\, \pi, \lambda) = \pi\lambda - \pi$$

$$\leq (2\pi\lambda - \pi - \lambda) \cdot \frac{1}{2}$$

$$< (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda} \right)$$

$$= U_r^*(t_W \,;\, \pi, \lambda).$$

Case 3: Suppose that $\frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda} > \phi_\beta(t_W)$. Then $U_r^*(t_W \,;\, \pi, \lambda) = \pi\lambda - \lambda$. If $\pi_M - \frac{1}{2} > \phi_\beta(t_W)$, then $U_d^*(t_W, \pi_M \,;\, \pi, \lambda) = U_r^*(t_W \,;\, \pi, \lambda)$. Similarly, if $\frac{1}{2} - \pi_M > \phi_\alpha(t_W)$, then

$$U_d^*(t_W, \pi_M \,;\, \pi, \lambda) = \pi\lambda - \pi$$

$$\leq U_r^*(t_W \,;\, \pi, \lambda).$$

In all other cases, the manager of type π_M rubberstamps a signal of quality $q \in \left[0, \frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda} \right)$. Hence

$$U_d^*(t_W, \pi_M \,;\, \pi, \lambda) = (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q \right)$$

$$< (2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - \frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda} \right)$$

$$= \pi\lambda - \lambda$$

$$= U_r^*(t_W \,;\, \pi, \lambda).$$

Due to the symmetry of the environment, the analysis for the cases in which $\pi < \lambda$ is analogous. $\qquad\square$

Lemma 1.7. *Consider a candidate manager whose skepticism is greater than the principal's informational standard, i.e.,*

$$\left| \pi_M - \frac{1}{2} \right| > \frac{|\pi - \lambda|}{2\pi + 2\lambda - 4\pi\lambda}.$$

Let t_W and t'_W be two worker types. Suppose that $U_d^*(t_W, \pi_M\,;\pi, \lambda) \geq U_d^*(t'_W, \pi_M\,;\pi, \lambda)$. Then $U_M^*(t_W\,;\pi_M) \geq U_M^*(t'_W\,;\pi_M)$.

Proof. Suppose that $\pi \geq \lambda$. Let $t_W = (\pi_W, \lambda_W, q_1(\cdot), c_1(\cdot))$ and $t'_W \equiv (\pi'_W, \lambda'_W, q_2(\cdot), c_2(\cdot))$ be given. Also, let $q \equiv q_1(\hat{e}(t_W))$ and $q' \equiv q_2(\hat{e}(t'_W))$.

Case 1: $\boldsymbol{\pi_M > \frac{1}{2}}$. Observe that each of $U_d^*(t_W, \pi_M\,;\pi, \lambda)$, $U_M^*(t_W, \pi_M)$, $U_d^*(t'_W, \pi_M\,;\pi, \lambda)$, and $U_M^*(t'_W, \pi_M)$ can take any of three forms.[1] In particular, the skepticism of a manager of type π_M either has no effect on, strengthens, or destroys the incentives of a worker of type t_W. Label these subcases S_1, S_2, and S_3, respectively. Similarly, the manager's skepticism either has no effect on, strengthens, or destroys the incentives of a worker of type t'_W. Label these subcases S'_1, S'_2, and S'_3, respectively. Tables A.1 and A.2 show the principal's and manager's payoffs under the different subcases. Consider

Quantity

		$U_d^*(t_W, \pi_M\,;\pi, \lambda)$	$U_M^*(t_W\,;\pi_M)$
	S_1	$(2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q\right)$	$-\frac{1}{2} \cdot \left(\frac{1}{2} - q\right)$
Subcase	S_2	$(2\pi\lambda - \pi - \lambda) \cdot (1 - \pi_M)$	$-\frac{1}{2} \cdot (1 - \pi_M)$
	S_3	$\pi\lambda - \lambda$	$-\frac{1}{2} \cdot (1 - \pi_M)$

Table A.1: Comparisons of $U_d^*(t_W, \pi_M\,;\pi, \lambda)$ and $U_M^*(t_W\,;\pi_M)$ under subcases S_1, S_2, and S_3 when $\pi_M < \frac{1}{2}$

the following combinations of subcases:

$\boldsymbol{S_1}$ **and** $\boldsymbol{S'_1}$**:** Suppose that $U_d^*(t_W, \pi_M\,;\pi, \lambda) \geq U_d^*(t'_W, \pi_M\,;\pi, \lambda)$. Then $q \geq q'$, which implies that $U_M^*(t_W\,;\pi_M) \geq U_M^*(t'_W\,;\pi_M)$.

$\boldsymbol{S_1}$ **and** $\boldsymbol{S'_2}$**:** Suppose that $U_d^*(t_W, \pi_M\,;\pi, \lambda) \geq U_d^*(t'_W, \pi_M\,;\pi, \lambda)$. Then $q \geq \pi_M - \frac{1}{2}$, which implies that $U_M^*(t_W\,;\pi_M) \geq U_M^*(t'_W\,;\pi_M)$.

$\boldsymbol{S_1}$ **and** $\boldsymbol{S'_3}$**:** Suppose that $U_d^*(t_W, \pi_M\,;\pi, \lambda) \geq U_d^*(t'_W, \pi_M\,;\pi, \lambda)$. Since S_1 holds, $\pi_M - \frac{1}{2} \leq q$, and so $\frac{1}{2} - q \leq 1 - \pi_M$. Hence $U_M^*(t_W\,;\pi_M) \geq U_M^*(t'_W\,;\pi_M)$.

$\boldsymbol{S_2}$ **and** $\boldsymbol{S'_1}$**:** Suppose that $U_d^*(t_W, \pi_M\,;\pi, \lambda) \geq U_d^*(t'_W, \pi_M\,;\pi, \lambda)$. Then $\pi_M - \frac{1}{2} \geq q'$, which implies that $U_M^*(t_W\,;\pi_M) \geq U_M^*(t'_W\,;\pi_M)$.

---

[1]See Proposition 1.3.

Quantity

		$U_d^*(t_W', \pi_M\,;\pi,\lambda)$	$U_M^*(t_W'\,;\pi_M)$
	S_1'	$(2\pi\lambda - \pi - \lambda)\cdot\left(\frac{1}{2} - q'\right)$	$-\frac{1}{2}\cdot\left(\frac{1}{2} - q'\right)$
Subcase	S_2'	$(2\pi\lambda - \pi - \lambda)\cdot(1 - \pi_M)$	$-\frac{1}{2}\cdot(1 - \pi_M)$
	S_3'	$\pi\lambda - \lambda$	$-\frac{1}{2}\cdot(1 - \pi_M)$

Table A.2: Comparisons of $U_d^*(t_W', \pi_M\,;\pi,\lambda)$ and $U_M^*(t_W'\,;\pi_M)$ under subcases S_1', S_2', and S_3' when $\pi_M > \frac{1}{2}$

S_2 **and** S_2'**:** $U_M^*(t_W\,;\pi_M) = U_M^*(t_W'\,;\pi_M)$, so the implication holds vacuously.

S_2 **and** S_3'**:** $U_M^*(t_W\,;\pi_M) = U_M^*(t_W'\,;\pi_M)$, so the implication holds vacuously.

S_3 **and** S_1'**:** Suppose that $U_d^*(t_W, \pi_M\,;\pi,\lambda) \geq U_d^*(t_W', \pi_M\,;\pi,\lambda)$. Since the manager's skepticism exceeds the principal's informational standard,

$$q' \leq \frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda} < \pi_M - \frac{1}{2}.$$

Therefore,

$$1 - \pi_M < \frac{1}{2} - q',$$

and so $U_M^*(t_W\,;\pi_M) > U_M^*(t_W'\,;\pi_M)$.

S_3 **and** S_2'**:** $U_M^*(t_W\,;\pi_M) = U_M^*(t_W'\,;\pi_M)$, so the implication holds vacuously.

S_3 **and** S_3'**:** $U_M^*(t_W\,;\pi_M) = U_M^*(t_W'\,;\pi_M)$, so the implication holds vacuously.

Case 2: $\pi_M < \frac{1}{2}$**.** In this case, the payoffs can be written as shown in Tables A.3 and A.4.

Quantity

		$U_d^*(t_W, \pi_M\,;\pi,\lambda)$	$U_M^*(t_W\,;\pi_M)$
	S_1	$(2\pi\lambda - \pi - \lambda)\cdot\left(\frac{1}{2} - q\right)$	$-\frac{1}{2}\cdot\left(\frac{1}{2} - q\right)$
Subcase	S_2	$(2\pi\lambda - \pi - \lambda)\cdot\pi_M$	$-\frac{1}{2}\cdot\pi_M$
	S_3	$\pi\lambda - \pi$	$-\frac{1}{2}\cdot\pi_M$

Table A.3: Comparisons of $U_d^*(t_W, \pi_M\,;\pi,\lambda)$ and $U_M^*(t_W\,;\pi_M)$ under subcases S_1, S_2, and S_3 when $\pi_M < \frac{1}{2}$

Just as in the analysis of the previous case, each subcase can be considered individually:

Quantity

		$U_d^*(t_W', \pi_M\,;\pi,\lambda)$	$U_M^*(t_W'\,;\pi_M)$
	S_1'	$(2\pi\lambda - \pi - \lambda) \cdot \left(\frac{1}{2} - q'\right)$	$-\frac{1}{2} \cdot \left(\frac{1}{2} - q'\right)$
Subcase	S_2'	$(2\pi\lambda - \pi - \lambda) \cdot \pi_M$	$-\frac{1}{2} \cdot \pi_M$
	S_3'	$\pi\lambda - \pi$	$-\frac{1}{2} \cdot \pi_M$

Table A.4: Comparisons of $U_d^*(t_W', \pi_M\,;\pi,\lambda)$ and $U_M^*(t_W'\,;\pi_M)$ under subcases S_1', S_2', and S_3' when $\pi_M < \frac{1}{2}$

S_1 **and** S_1'**:** Suppose that $U_d^*(t_W, \pi_M\,;\pi,\lambda) \geq U_d^*(t_W', \pi_M\,;\pi,\lambda)$. Then $q \geq q'$, which implies that $U_M^*(t_W\,;\pi_M) \geq U_M^*(t_W'\,;\pi_M)$.

S_1 **and** S_2'**:** Suppose that $U_d^*(t_W, \pi_M\,;\pi,\lambda) \geq U_d^*(t_W', \pi_M\,;\pi,\lambda)$. Then $q \geq \frac{1}{2} - \pi_M$, which implies that $U_M^*(t_W\,;\pi_M) \geq U_M^*(t_W'\,;\pi_M)$.

S_1 **and** S_3'**:** Note that $U_M^*(t_W\,;\pi_M) \geq U_M^*(t_W'\,;\pi_M)$ is equivalent to $q \geq \frac{1}{2} - \pi_M$, which is implied by S_1.

S_2 **and** S_1'**:** Suppose that $U_d^*(t_W, \pi_M\,;\pi,\lambda) \geq U_d^*(t_W', \pi_M\,;\pi,\lambda)$. Then $\pi_M \leq \frac{1}{2} - q'$, which implies that $U_M^*(t_W\,;\pi_M) \geq U_M^*(t_W'\,;\pi_M)$.

S_2 **and** S_2'**:** $U_M^*(t_W\,;\pi_M) = U_M^*(t_W'\,;\pi_M)$, so the implication holds vacuously.

S_2 **and** S_3'**:** $U_M^*(t_W\,;\pi_M) = U_M^*(t_W'\,;\pi_M)$, so the implication holds vacuously.

S_3 **and** S_1'**:** Here, $U_d^*(t_W, \pi_M\,;\pi,\lambda) \geq U_d^*(t_W', \pi_M\,;\pi,\lambda)$ is impossible, as it implies that

$$q' \leq -\frac{\pi - \lambda}{2\pi + 2\lambda - 4\pi\lambda},$$

which holds only if $q' = 0$. Since $\pi_M < \frac{1}{2}$, though, $\frac{1}{2} - \pi_M > q'$, which cannot occur under S_1'. So the implication holds vacuously.

S_3 **and** S_2'**:** $U_M^*(t_W\,;\pi_M) = U_M^*(t_W'\,;\pi_M)$, so the implication holds vacuously.

S_3 **and** S_3'**:** $U_M^*(t_W\,;\pi_M) = U_M^*(t_W'\,;\pi_M)$, so the implication holds vacuously.

By the symmetry of the environment, the analysis of the case in which $\pi < \lambda$ is analogous. $\quad\square$

A.2 Technical Results

Lemma 1.1. *Let $p, \ell \in (0,1)$. For any $\ell' \in (0,1)$, there exists a unique $p' \in (0,1)$ such that a manager with attributes $\pi_M = p$ and $\lambda_M = \ell$ and a manager with attributes $\pi_M = p'$ and $\lambda_M = \ell'$ have aligned preferences.*

Proof. Let $p, \ell, \ell' \in (0,1)$ be given, and set

$$p' \equiv \left[1 + \frac{\ell \cdot (1-p) \cdot (1-\ell')}{p\ell'(1-\ell)}\right]^{-1} = \frac{p\ell' - p\ell\ell'}{p\ell' - \ell\ell' - p\ell + \ell}.$$

Observe that

$$\frac{\ell \cdot (1-p)}{p \cdot (1-\ell)} = \frac{\ell' \cdot (1-p')}{p' \cdot (1-\ell')}.$$

Thus, for every $q \in \left[0, \frac{1}{2}\right)$,

$$-\ell \cdot \left(1 + \frac{p}{1-p} \cdot \frac{\frac{1}{2}+q}{\frac{1}{2}-q}\right)^{-1} \lesseqqgtr -(1-\ell) \cdot \left(1 + \frac{1-p}{p} \cdot \frac{\frac{1}{2}-q}{\frac{1}{2}+q}\right)^{-1} \quad \text{if and only if}$$

$$-\ell' \cdot \left(1 + \frac{p'}{1-p'} \cdot \frac{\frac{1}{2}+q}{\frac{1}{2}-q}\right)^{-1} \lesseqqgtr -(1-\ell') \cdot \left(1 + \frac{1-p'}{p'} \cdot \frac{\frac{1}{2}-q}{\frac{1}{2}+q}\right)^{-1} \quad \text{(A.22)}$$

and

$$-\ell \cdot \left(1 + \frac{p}{1-p} \cdot \frac{\frac{1}{2}-q}{\frac{1}{2}+q}\right)^{-1} \lesseqqgtr -(1-\ell) \cdot \left(1 + \frac{1-p}{p} \cdot \frac{\frac{1}{2}+q}{\frac{1}{2}-q}\right)^{-1} \quad \text{if and only if}$$

$$-\ell' \cdot \left(1 + \frac{p'}{1-p'} \cdot \frac{\frac{1}{2}-q}{\frac{1}{2}+q}\right)^{-1} \lesseqqgtr -(1-\ell') \cdot \left(1 + \frac{1-p'}{p'} \cdot \frac{\frac{1}{2}+q}{\frac{1}{2}-q}\right)^{-1}. \quad \text{(A.23)}$$

The conditions (A.22) and (A.23), together, are equivalent to the statement that a manager with attributes $\pi_M = p, \lambda_M = \ell$ and a manager with attributes $\pi_M = p', \lambda_M = \ell'$ have aligned preferences. Furthermore, from these conditions, it is evident that p' is unique. $\quad\square$

Lemma A.1. *The function $x \mapsto 2c\left(q^{-1}(x)\right)$, defined on $\left[0, \frac{1}{2}\right)$, has exactly two fixed points. One is 0, and the other one, denoted by ϕ, belongs to the interval $\left(q(e_r^*), q\left(c^{-1}\left(\frac{1}{4}\right)\right)\right)$.*

Proof. This result is a special case of Lemma A.4. $\quad\square$

Lemma A.2. *For all $t \in (0, \phi)$, $t > 2c\left(q^{-1}(t)\right)$. For all $t \in \left(\phi, \frac{1}{2}\right)$, $t < 2c\left(q^{-1}(t)\right)$.*

Proof. Suppose that there exists $t \in (0, \phi)$ such that $t \leq 2c\left(q^{-1}(t)\right)$. Lemma A.1 immediately rules out equality, so it follows that $t < 2c\left(q^{-1}(t)\right)$. Let h denote the function $t \mapsto 2c\left(q^{-1}(t)\right) - t$, defined on $\left[0, \frac{1}{2}\right)$. By Lemma A.1, $q(e_r^*) < \phi$. Furthermore, $h(q(e_r^*)) = 2c(e_r^*) - q(e_r^*) < 0$ by Lemma A.3. By assumption, $h(t) > 0$. Since $h(\cdot)$ is continuous, the Intermediate Value Theorem implies that there exists $t' \in (\min\{q(e_r^*), t\}, \max\{q(e_r^*), t\}) \subset (0, \phi)$ such that $h(t') = 0$, which contradicts Lemma A.1. So $t > 2c\left(q^{-1}(t)\right)$ for all $t \in (0, \phi)$.

A symmetric argument, in which $q\left(c^{-1}\left(\frac{1}{4}\right)\right)$ replaces $q(e_r^*)$, establishes the fact that $t < 2c\left(q^{-1}(t)\right)$ for all $t \in \left(\phi, \frac{1}{2}\right)$. $\quad\square$

Lemma A.3. $\xi(t_W)q(\hat{e}(t_W)) > c(\hat{e}(t_W))$.

Proof. Suppose that $\xi(t_W)q(\hat{e}(t_W)) \leq c(\hat{e}(t_W))$. Then the value of the worker's objective function at 0 (i.e., $-\frac{\xi(t_W)}{2}$) is at least as large as at $\hat{e}(t_W)$, contradicting the statement of Lemma 1.3. $\qquad\square$

Lemma A.4. *Fix a worker type $t_W \equiv (\pi_W, \lambda_W, q(\cdot), c(\cdot))$ and a constant $\gamma \in \mathbb{R}$. Consider the function*

$$g(\cdot\,; t_W, \gamma) : \left[0, \frac{1}{2}\right) \to \mathbb{R}$$

$$x \mapsto \gamma + \frac{c(q^{-1}(x))}{\xi(t_W)}.$$

If $\gamma > 0$ and $q(\hat{e}(t_W)) < g(q(\hat{e}(t_W)); t_W, \gamma)$, then $g(\cdot\,; t_W, \gamma)$ has no fixed point. Otherwise, the following inequalities hold:

$$\gamma < \frac{1}{2}, \tag{A.24}$$

$$\hat{e}(t_W) < c^{-1}\left(\frac{\xi(t_W) \cdot [1 - 2\gamma]}{2}\right). \tag{A.25}$$

Hence, the interval

$$I(t_W, \gamma) \equiv \left(q(\hat{e}(t_W)), q\left(c^{-1}\left(\frac{\xi(t_W) \cdot [1 - 2\gamma]}{2}\right)\right)\right)$$

is well defined and nonempty. In this case, $g(\cdot\,; t_W, \gamma)$ has either one or two fixed points. In particular:

(i) *If $\gamma < 0$, then $g(\cdot\,; t_W, \gamma)$ has a unique fixed point, which belongs to $I(t_W, \gamma)$.*

(ii) *If $\gamma = 0$, then $g(\cdot\,; t_W, \gamma)$ has exactly two fixed points: one is 0, and the other belongs to $I(t_W, \gamma)$.*

(iii) *If $\gamma > 0$ and $q(\hat{e}(t_W)) = g(q(\hat{e}(t_W)); t_W, \gamma)$, then $q(\hat{e}(t_W))$ is the only fixed point of $g(\cdot\,; t_W, \gamma)$.*

(iv) *If $\gamma > 0$ and $q(\hat{e}(t_W)) > g(q(\hat{e}(t_W)); t_W, \gamma)$, then $g(\cdot\,; t_W, \gamma)$ has exactly two fixed points, which belong, respectively, to the intervals $(0, q(\hat{e}(t_W)))$ and $I(t_W, \gamma)$.*

Proof. Let $t_W = (\pi_W, \lambda_W, q(\cdot), c(\cdot))$ and γ be given. Define the following function, which is twice differentiable and strictly convex:

$$h : \left[0, \frac{1}{2}\right) \to \mathbb{R}$$

$$x \mapsto g(x\,; t_W, \gamma) - x.$$

Suppose $\gamma > 0$ and $q(\hat{e}(t_W)) < g(q(\hat{e}(t_W)) ; t_W, \gamma)$. Because $h(\cdot)$ is differentiable throughout the interior of its domain and strictly convex throughout its domain, any solution to the first-order condition $h'(x) = 0$ will be unique and, furthermore, will characterize the minimizer of $h(\cdot)$. Note that $q(\hat{e}(t_W))$ is a solution to the first-order condition, and that $h(q(\hat{e}(t_W))) > 0$. Thus $h(\cdot)$ is positive throughout its domain, which implies that $g(\cdot ; t_W, \gamma)$ has no fixed point.

Suppose that $\gamma \leq 0$. In this case, (A.24) is obvious. (A.25) follows by observing that

$$
\begin{aligned}
\hat{e}(t_W) &= c^{-1}(c(\hat{e}(t_W))) \\
&< c^{-1}(\xi(t_W) q(\hat{e}(t_W))) \text{ by Lemma A.3} \\
&< c^{-1}\left(\frac{\xi(t_W)}{2}\right) \\
&\leq c^{-1}\left(\frac{\xi(t_W) \cdot [1 - 2\gamma]}{2}\right).
\end{aligned}
$$

Suppose that $q(\hat{e}(t_W)) \geq g(q(\hat{e}(t_W)) ; t_W, \gamma)$. Then

$$
\begin{aligned}
\gamma &< \gamma + \frac{c(\hat{e}(t_W))}{\xi(t_W)} \\
&= g(q(\hat{e}(t_W)) ; t_W, \gamma) \\
&\leq q(\hat{e}(t_W)) \\
&< \frac{1}{2}
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{e}(t_W) &= q^{-1}(g^{-1}(g(q(\hat{e}(t_W)) ; t_W, \gamma) ; t_W, \gamma)) \\
&\leq q^{-1}\left(g^{-1}\left(q(\hat{e}(t_W)) ; t_W, \gamma\right)\right) \\
&< q^{-1}\left(g^{-1}\left(\frac{1}{2} ; t_W, \gamma\right)\right) \\
&= c^{-1}\left(\frac{\xi(t_W) \cdot [1 - 2\gamma]}{2}\right),
\end{aligned}
$$

so both (A.24) and (A.25) hold. Hence, either of the conditions $\gamma \leq 0$ or $q(\hat{e}(t_W)) \geq g(q(\hat{e}(t_W)) ; t_W, \gamma)$ is sufficient for $I(t_W, \gamma)$ to be well defined and nonempty. Now I prove statements (i) through (iv):

(i) Suppose that $\gamma < 0$. Because $h(0) = \gamma < 0$ and $h'(x) \leq 0$ for $x \in [0, q(\hat{e}(t_W))]$, $h(q(\hat{e}(t_W))) < 0$, and $g(\cdot ; t_W, \gamma)$ has no fixed point in $[0, q(\hat{e}(t_W))]$. Similarly, because

$$
h\left(q\left(c^{-1}\left(\frac{\xi(t_W) \cdot [1 - 2\gamma]}{2}\right)\right)\right) = \frac{1}{2} - q\left(c^{-1}\left(\frac{\xi(t_W) \cdot [1 - 2\gamma]}{2}\right)\right) > 0
$$

and $h'(x) > 0$ for $x \in \left(q(\hat{e}(t_W)), \frac{1}{2} \right)$, $g(\cdot)$ has no fixed point in $\left[q \left(c^{-1} \left(\frac{\xi(t_W) \cdot [1 - 2\gamma]}{2} \right) \right), \frac{1}{2} \right)$. On the other hand, since $h \left(q(\hat{e}(t_W)) \right) < 0 < h \left(q \left(c^{-1} \left(\frac{\xi(t_W) \cdot [1 - 2\gamma]}{2} \right) \right) \right)$, the Intermediate Value Theorem implies that $g(\cdot)$ has at least one fixed point in $I(t_W, \gamma)$. Suppose that there are at least two fixed points—that is, suppose that there exist $x_1, x_2 \in I(t_W, \gamma)$ such that $x_1 < x_2$ and $h(x_1) = 0 = h(x_2)$. By Rolle's Theorem, there exists $x_0 \in (x_1, x_2)$ such that $h'(x_0) = 0$, which contradicts the fact that $h'(x) > 0$ for every $x \in \left(q(\hat{e}(t_W)), \frac{1}{2} \right)$. Thus $g(\cdot; t_W, \gamma)$ has exactly one fixed point, which belongs to $I(t_W, \gamma)$.

(ii) Suppose that $\gamma = 0$. Since $h(0) = 0$, 0 is a fixed point of $g(\cdot; t_W, 0)$. Since $h'(x) < 0$ for $x \in (0, q(\hat{e}(t_W)))$, $h(q(\hat{e}(t_W))) < 0$, and $g(\cdot; t_W, 0)$ has no fixed point in $(0, q(\hat{e}(t_W))]$. By the same arguments as in the proof of (i), $g(\cdot; t_W, 0)$ also has no fixed point in $\left[q \left(c^{-1} \left(\frac{\xi(t_W)}{2} \right) \right), \frac{1}{2} \right)$, but it has exactly one fixed point in $I(t_W, \gamma)$.

(iii) Suppose that $\gamma > 0$ and $q(\hat{e}(t_W)) = g(q(\hat{e}(t_W)); t_W, \gamma)$. Clearly, $q(\hat{e}(t_W))$ is a fixed point of $g(\cdot; t_W, \gamma)$. Because $h'(q(\hat{e}(t_W))) = 0$ and $h''(\cdot) > 0$, $h(\cdot)$ achieves its minimum at $q(\hat{e}(t_W))$. Since $h(q(\hat{e}(t_W))) = 0$, $h(x) > 0$ for $x \in \left[0, \frac{1}{2} \right) \setminus \{q(\hat{e}(t_W))\}$. It follows that $q(\hat{e}(t_W))$ is the unique fixed point of $g(\cdot; t_W, \gamma)$.

(iv) Suppose that $\gamma > 0$ and $q(\hat{e}(t_W)) > g(q(\hat{e}(t_W)); t_W, \gamma)$. Since $h(0) > 0$, $h(q(\hat{e}(t_W))) < 0$, and $h \left(q \left(c^{-1} \left(\frac{\xi(t_W) \cdot [1 - 2\gamma]}{2} \right) \right) \right) > 0$, the Intermediate Value Theorem implies that $g(\cdot; t_W, \gamma)$ has at least one fixed point in $(0, q(\hat{e}(t_W)))$ and at least one fixed point in $I(t_W, \gamma)$. Let x_1 and x_2, respectively, denote these two fixed points. Suppose that there is a third fixed point, x_3. Clearly $x_3 \notin \{0, q(\hat{e}(t_W))\}$. If $x_3 \in (0, q(\hat{e}(t_W)))$, then, by Rolle's Theorem, there exists $x_0 \in (\min\{x_1, x_3\}, \max\{x_1, x_3\})$ such that $h'(x_0) = 0$. This conclusion, though, contradicts the fact that $h'(x) < 0$ for $x \in [0, q(\hat{e}(t_W)))$. The premise that $x_3 \in \left(q(\hat{e}(t_W)), \frac{1}{2} \right)$ leads to a similar contradiction, since $h'(x) > 0$ for $x \in \left(q(\hat{e}(t_W)), \frac{1}{2} \right)$. Thus the two fixed points are unique.

$\square$

Definition A.1 $(\bar{p}(t_W, \gamma))$. Fix a worker type $t_W \equiv (\pi_W, \lambda_W, q(\cdot), c(\cdot))$ and a constant $\gamma \in \mathbb{R}$, and let $g(\cdot; t_W, \gamma)$ denote the function defined in Lemma A.4. Define

$$ P(t_W, \gamma) \equiv \left\{ x \in \left[0, \frac{1}{2} \right) : g(x; t_W, \gamma) = x \right\} $$

to be the set of fixed points of $g(\cdot; t_W, \gamma)$. Now, define

$$ \bar{p}(t_W, \gamma) \equiv \begin{cases} 0 & \text{if } P(t_W, \gamma) = \varnothing, \\ \max P(t_W, \gamma) & \text{otherwise.} \end{cases} $$

That is, $\bar{p}(t_W, \gamma)$ is the largest fixed point of $g(\cdot; t_W, \gamma)$, if it has any fixed points. If $g(\cdot; t_W, \gamma)$ has no fixed point, then $\bar{p}(t_W, \gamma) \equiv 0$.

Lemma A.5. $\overline{p}(t_W, \gamma) \leq q(\hat{e}(t_W))$ *if and only if both of the conditions* $\gamma > 0$ *and* $q(\hat{e}(t_W)) \leq g(q(\hat{e}(t_W)); t_W, \gamma)$ *hold simultaneously.*

Proof. By Lemma A.4, $\overline{p}(t_W, \gamma) < q(\hat{e}(t_W))$ (and, in fact, $\overline{p}(t_W, \gamma) = 0$) if and only if $\gamma > 0$ and $q(\hat{e}(t_W)) < g(q(\hat{e}(t_W)); t_W, \gamma)$. Furthermore, $\overline{p}(t_W, \gamma) = q(\hat{e}(t_W))$ if and only if $\gamma > 0$ and $q(\hat{e}(t_W)) = g(q(\hat{e}(t_W)); t_W, \gamma)$. The result follows. $\qquad\square$

Lemma A.6. *If* $\overline{p}(t_W, \gamma) > 0$, *then* $\overline{p}(t_W, \gamma) \geq q(\hat{e}(t_W))$. *The inequality becomes strict if* $\gamma < 0$.

Proof. The result follows directly from the cases analyzed in the proof of Lemma A.4. $\qquad\square$

Lemma A.7. *Fix a worker type* $t_W \equiv (\pi_W, \lambda_W, q(\cdot), c(\cdot))$ *and a constant* $\gamma \in \mathbb{R}$, *and let* $g(\cdot\,; t_W, \gamma)$ *denote the function defined in Lemma A.4.*

   (i) $g(x\,; t_W, \gamma) < x$ *for every* $x \in (q(\hat{e}(t_W)), \overline{p}(t_W, \gamma))$.

   (ii) $g(x\,; t_W, \gamma) > x$ *for every* $x \in \left(\overline{p}(t_W, \gamma), \frac{1}{2}\right)$.

Proof. Let t_W and γ be given, and define the following function, which is twice differentiable and strictly convex:

$$h : \left[0, \frac{1}{2}\right) \to \mathbb{R}$$

$$x \mapsto g(x\,; t_W, \gamma) - x.$$

I prove each of the two statements individually.

   (i) When $\overline{p}(t_W, \gamma) \leq q(\hat{e}(t_W))$, the conclusion holds vacuously, since $(q(\hat{e}(t_W)), \overline{p}(t_W, \gamma)) = \varnothing$. By Lemma A.5, this condition occurs if and only if both of the conditions $\gamma > 0$ and $q(\hat{e}(t_W)) \leq g(q(\hat{e}(t_W)); t_W, \gamma)$ hold simultaneously. The remainder of the proof proceeds by contradiction. Suppose that either $\gamma \leq 0$ or $q(\hat{e}(t_W)) > g(q(\hat{e}(t_W)); t_W, \gamma)$ and that there exists $x_0 \in (q(\hat{e}(t_W)), \overline{p}(t_W, \gamma))$ for which $g(x_0\,; t_W, \gamma) \geq x_0$, meaning that $h(x_0) \geq 0$. There are three cases to consider.

      Case 1: $\boldsymbol{\gamma < 0.}$ In this case, $h(0) = \gamma < 0$. By the Intermediate Value Theorem, there exists $x_1 \in (0, x_0]$ such that $h(x_1) = 0$. Note that $x_1 < \overline{p}(t_W, \gamma)$ and that x_1 is a fixed point of $g(\cdot\,; t_W, \gamma)$. This conclusion contradicts Lemma A.4, which asserts that $\overline{p}(t_W, \gamma)$ is the only fixed point of $g(\cdot\,; t_W, \gamma)$.

      Case 2: $\boldsymbol{\gamma = 0.}$ Here, $h(0) = 0$, and, since $h'(0) < 0$, there exists $x_1 > 0$ such that $h(x_1) < 0$. Hence the Intermediate Value Theorem implies that there exists $x_2 \in (x_1, x_0]$ such that $h(x_2) = 0$. Note that $0 < x_2 < \overline{p}(t_W, \gamma)$ and that x_2 is a fixed point of $g(\cdot\,; t_W, \gamma)$. This conclusion contradicts Lemma A.4, which asserts that 0 and $\overline{p}(t_W, \gamma)$ are the only fixed points of $g(\cdot\,; t_W, \gamma)$.

Case 3: $\gamma > 0$ and $q(\hat{e}(t_W)) > g(q(\hat{e}(t_W)); t_W, \gamma)$. In this case, $h(q(\hat{e}(t_W))) < 0$, so, according to the Intermediate Value Theorem, there exists $x_1 \in (q(\hat{e}(t_W)), x_0]$ such that $h(x_1) = 0$. Thus x_1 is a fixed point of $g(\cdot; t_W, \gamma)$ that belongs to $(q(\hat{e}(t_W)), \overline{p}(t_W, \gamma))$, which is impossible, since, as Lemma A.4 states, $g(\cdot; t_W, \gamma)$ has exactly two fixed points: $\overline{p}(t_W, \gamma)$ and another that belongs to $(0, q(\hat{e}(t_W)))$.

(ii) Suppose that there exists $x_0 \in \left(\overline{p}(t_W, \gamma), \frac{1}{2}\right)$ such that $x_0 \geq g(x_0; t_W, \gamma)$. By the definition of $\overline{p}(t_W, \gamma)$ (i.e., Definition A.1),

$$x_0 \neq g(x_0; t_W, \gamma) \quad \forall x \in \left(\overline{p}(t_W, \gamma), \frac{1}{2}\right). \tag{A.26}$$

Thus $x_0 > g(x_0; t_W, \gamma)$. Because $\lim_{x \to \infty} g(x; t_W, \gamma) = \infty$, there exists $x_1 \in \left(x_0, \frac{1}{2}\right)$ such that $g(x_1; t_W, \gamma) > \frac{1}{2} > x_1$, which implies that $h(x_1) > 0$. On the other hand, $h(x_0) < 0$, so the Intermediate Value Theorem implies that there exists $x_2 \in (x_0, x_1) \subset \left(\overline{p}(t_W, \gamma), \frac{1}{2}\right)$ such that $h(x_2) = 0$, meaning that $x_2 \in \left(\overline{p}(t_W, \gamma), \frac{1}{2}\right)$ is a fixed point of $g(\cdot; t_W, \gamma)$. This conclusion contradicts (A.26).

$\square$

Definition A.2 (Tolerance for skepticism). Fix a worker type $t_W \equiv (\pi_W, \lambda_W, q(\cdot), c(\cdot))$.

(i) $\phi_\alpha(t_W) \equiv \overline{p}\left(t_W, -\frac{\tau(t_W)}{2\xi(t_W)}\right)$ is the worker's *tolerance for skepticism regarding α*.

(ii) $\phi_\beta(t_W) \equiv \overline{p}\left(t_W, \frac{\tau(t_W)}{2\xi(t_W)}\right)$ is the worker's *tolerance for skepticism regarding β*.

(iii) $\phi(t_W) \equiv \max\{\phi_\alpha(t_W), \phi_\beta(t_W)\}$ is the worker's *tolerance for skepticism*.

Lemma A.8. *Consider a candidate worker of type $t_W \in \mathcal{T}_W$.*

(i) *If $\tau(t_W) < 0$, then $\phi_\alpha(t_W) < \phi_\beta(t_W)$.*

(ii) *If $\tau(t_W) = 0$, then $\phi_\alpha(t_W) = \phi_\beta(t_W)$.*

(iii) *If $\tau(t_W) > 0$, then $\phi_\alpha(t_W) > \phi_\beta(t_W)$.*

Proof. Let $t_W = (\pi_W, \lambda_W, q(\cdot), c(\cdot))$. I prove the statements individually:

(i) Suppose that $\tau(t_W) < 0$. Then, by Lemma A.4, $\phi_\beta(t_W) > q(\hat{e}(t_W))$, and there are three possibilities for $\phi_\alpha(t_W)$. The first two possibilities are the ones in which $\phi_\alpha(t_W) \in \{0, q(\hat{e}(t_W))\}$. In either of these cases, the result follows immediately. The third possibility is that $\phi_\alpha(t_W) > q(\hat{e}(t_W))$. In this case,

$$g\left(\phi_\alpha(t_W); t_W, \frac{\tau(t_W)}{2\xi(t_W)}\right) = \frac{\tau(t_W)}{2\xi(t_W)} + \frac{c\left(q^{-1}\left(\phi_\alpha(t_W)\right)\right)}{2\xi(t_W)}$$

$$< -\frac{\tau(t_W)}{2\xi(t_W)} + \frac{c\left(q^{-1}\left(\phi_\alpha(t_W)\right)\right)}{2\xi(t_W)}$$
$$= \phi_\alpha(t_W) \text{ by Definition A.2.}$$

The result now follows from Lemma A.7.

(ii) This result is clear from Definition A.2.

(iii) The proof is analogous to that of (i).

$\square$

Appendix B

Supplement to Chapter 2

Lemma B.1. *Let $w \in W$ be given.*

    (i) $w < 2c(q^*) - q^* + \frac{1}{2}$ *if and only if* $\overline{q}(w) = 0$.

    (ii) $w = 2c(q^*) - q^* + \frac{1}{2}$ *if and only if* $\overline{q}(w) = q^*$.

    (iii) $w > 2c(q^*) - q^* + \frac{1}{2}$ *if and only if* $\overline{q}(w) > q^*$.

Proof. Given that the left hand side statements are mutually exclusive and exhaustive and that the right hand side statements are mutually exclusive, it suffices to prove only the three "only if" statements. Fix $w \in W$, and observe that

$$\arg\min_{q \in \left[0, \frac{1}{2}\right)} \psi_w(q) = \{q^*\}, \tag{B.1}$$

where ψ_w is as defined in the proof of Lemma 2.2. Also, observe that

$$\overline{q}(w) = \begin{cases} 0 & \text{if } \left\{q \in \left[0, \frac{1}{2}\right) : \psi_w(q) = 0\right\} = \varnothing, \\ \max\left\{q \in \left[0, \frac{1}{2}\right) : \psi_w(q) = 0\right\} & \text{otherwise.} \end{cases}$$

    (i) Suppose that $w < 2c(q^*) - q^* + \frac{1}{2}$. Then, by (B.1),

$$\min_{q \in \left[0, \frac{1}{2}\right)} \psi_w(q) = \psi_w(q^*) > 0,$$

    and $\psi_w(q) = 0$ has no solution. Hence, φ_w has no fixed point: $\overline{q}(w) = 0$.

    (ii) Suppose that $w = 2c(q^*) - q^* + \frac{1}{2}$. In this case,

$$\min_{q \in \left[0, \frac{1}{2}\right)} \psi_w(q) = \psi_w(q^*) = 0.$$

    Since q^* is the unique minimizer, it is the unique fixed point: $\overline{q}(w) = q^*$.

(iii) Suppose that $w > 2c(q^*) - q^* + \frac{1}{2}$. In this case,

$$\min_{q \in \left[0, \frac{1}{2}\right)} \psi_w(q) = \psi_w(q^*) < 0.$$

Since $\lim_{q \uparrow \frac{1}{2}} \psi_w(q) = \infty$, the Intermediate Value Theorem implies that there exists $q' \in \left(q^*, \frac{1}{2}\right)$ such that $\psi_w(q') = 0$. Thus $\overline{q}(w) \geq q' > q^*$.[1]

$\square$

Lemma B.2. *Suppose that* $\overline{q}(w) > q^*$.

(i) *For all* $q \in (q^*, \overline{q}(w))$, $c(q) - \frac{q}{2} + \frac{1}{4} < \frac{w}{2}$.

(ii) *For all* $q \in \left(\overline{q}(w), \frac{1}{2}\right)$, $c(q) - \frac{q}{2} + \frac{1}{4} > \frac{w}{2}$.

Proof. Suppose that $\overline{q}(w) > q^*$. Lemma B.1 implies that $\psi_w(q^*) < 0$, where ψ_w is defined in the proof of Lemma 2.2.

(i) Suppose that there exists $q' \in (q^*, \overline{q}(w))$ such that $c(q') - \frac{q'}{2} + \frac{1}{4} \geq \frac{w}{2}$. Then $\psi_w(q') \geq 0$. By the Intermediate Value Theorem, there exists $q'' \in (q^*, q']$ such that $\psi_w(q'') = 0$. Recall that $\psi_w(\overline{q}(w)) = 0$ as well. Rolle's Theorem implies that there exists $q''' \in (q'', \overline{q}(w))$ such that $\psi'_w(q''') = 0$, which contradicts the fact that q^* is the unique solution to $\psi'_w(q) = 0$.

(ii) Suppose that there exists $q' \in \left(\overline{q}(w), \frac{1}{2}\right)$ such that $c(q') - \frac{q'}{2} + \frac{1}{4} \leq \frac{w}{2}$. Then $\psi_w(q') \leq 0$. Since $\lim_{q \uparrow \frac{1}{2}} \psi_w(q) = \infty$, the Intermediate Value Theorem implies that there exists $q'' \in \left[q', \frac{1}{2}\right)$ such that $\psi_w(q'') = 0$. Since $\psi_w(\overline{q}(w)) = 0$ as well, Rolle's Theorem implies that there exists $q''' \in (\overline{q}(w), q'')$ such that $\psi'_w(q''') = 0$. This conclusion contradicts the fact that q^* is the unique solution to $\psi'_w(q) = 0$.

$\square$

Lemma B.3. *Consider any* $w \in W$.

(i) *If* $\overline{q}(w) = 0$, *then* $\overline{q}(w') = 0$ *for every* $w' \in W$ *that satisfies* $w' < w$.

(ii) *If* $\overline{q}(w) > 0$, *then* $\overline{q}(w') < \overline{q}(w)$ *for every* $w' \in W$ *that satisfies* $w' < w$.

(ii) *If* $\overline{q}(w) > 0$, *then* $\overline{q}(w') > \overline{q}(w)$ *for every* $w' \in W$ *that satisfies* $w' > w$.

Proof. Let $w \in W$ be given.

---

[1]In fact, since $\psi'_w(q) > 0$ for $q \in \left(q^*, \frac{1}{2}\right)$, Rolle's Theorem implies that $\psi_w(q) \neq 0$ for all $q \in (q^*, q') \cup \left(q', \frac{1}{2}\right)$. Thus $\overline{q}(w) = q'$.

(i) Take any $w' \in W$ that satisfies $w' < w$. Since $\bar{q}(w) = 0$, Lemma B.1 implies that

$$w' < w < 2c(q^*) - q^* + \frac{1}{2},$$

and that $\bar{q}(w') = 0$.

(ii) Take any $w' \in W$ that satisfies $w' < w$. Suppose that $\bar{q}(w) > 0$. If $\bar{q}(w') = 0$, the result follows immediately. Suppose that $\bar{q}(w') > 0$. Recall the function ψ_w, defined in the proof of Lemma 2.2. Note that

$$\psi_w(\bar{q}(w')) = \psi_w(\bar{q}(w')) - \underbrace{\psi_{w'}(\bar{q}(w'))}_{=0} = w' - w < 0.$$

Since $\lim_{q \uparrow \frac{1}{2}} \psi_w(q) = \infty$, the Intermediate Value Theorem implies that there exists $q' \in \left(\bar{q}(w'), \frac{1}{2} \right)$ such that $\psi_w(q') = 0$. Thus $\bar{q}(w) \geq q' > \bar{q}(w')$.

(iii) Take any $w' \in W$ that satisfies $w' > w$. Suppose that $\bar{q}(w) > 0$. In this case,

$$\psi_{w'}(\bar{q}(w)) = \psi_{w'}(\bar{q}(w)) - \underbrace{\psi_w(\bar{q}(w))}_{=0} = w - w' < 0.$$

Since $\lim_{q \uparrow \frac{1}{2}} \psi_{w'}(q) = \infty$, the Intermediate Value Theorem implies that there exists $q' \in \left(\bar{q}(w), \frac{1}{2} \right)$ such that $\psi_{w'}(q') = 0$. Thus $\bar{q}(w') \geq q' > \bar{q}(w)$.

$\square$

Bibliography

ACEMOGLU, D., V. CHERNOZHUKOV, AND M. YILDIZ (2007): "Learning and Disagreement in an Uncertain World," Mimeo, Massachusetts Institute of Technology.

——— (2009): "Fragility of Asymptotic Agreement under Bayesian Learning," Mimeo, Massachusetts Institute of Technology.

AGHION, P., AND J. TIROLE (1997): "Formal and Real Authority in Organizations," *Journal of Political Economy*, 105(1), 1–29.

ALI, S. N., J. GOEREE, N. KARTIK, AND T. PALFREY (2008): "Information Aggregation In Standing and Ad Hoc Committees," *American Economic Review: Papers and Proceedings*, 98(2), 181–186.

AMBRUS, A., E. AZEVEDO, AND Y. KAMADA (2013): "Hierarchical Cheap Talk," *Theoretical Economics*, 8(1), 233–261.

AUMANN, R. (1976): "Agreeing to Disagree," *The Annals of Statistics*, 4(6), 1236–1239.

AUSTEN-SMITH, D., AND J. S. BANKS (1996): "Information Aggregation, Rationality, and the Condorcet Jury Theorem," *American Political Science Review*, 90(1), 34–45.

BADAWY, M. K. (1982): *Developing Managerial Skills in Engineers and Scientists: Succeeding as a Technical Manager.* Van Nostrand Reinhold Company.

BAKER, G., R. GIBBONS, AND K. J. MURPHY (1999): "Informal Authority in Organizations," *Journal of Law, Economics, and Organization*, 15(1), 56–73.

BANERJEE, A., AND R. SOMANATHAN (2001): "A Simple Model of Voice," *Quarterly Journal of Economics*, 116(1), 189–227.

BATTAGLINI, M., R. MORTON, AND T. PALFREY (2008): "Information Aggregation and Strategic Abstention in Large Laboratory Elections," *American Economic Review: Papers and Proceedings*, 98(2), 194–200.

BEGGS, A. W. (2001): "Queues and Hierarchies," *Review of Economic Studies*, 68(2), 297–322.

BLACKWELL, D., AND L. DUBINS (1962): "Merging of Opinions with Increasing Information," *Annals of Mathematical Statistics*, 33(3), 882–886.

BLOOM, N., B. EIFERT, A. MAHAJAN, D. MCKENZIE, AND J. ROBERTS (2013): "Does Management Matter? Evidence from India," *Quarterly Journal of Economics*, 128(1), 1–51.

BLOOM, N., AND J. VAN REENEN (2007): "Measuring and Explaining Management Practices across Firms and Countries," *Quarterly Journal of Economics*, 122(4), 1351–1408.

BOLTON, P., AND M. DEWATRIPONT (1994): "The Firm as a Communication Network," *Quarterly Journal of Economics*, 109(4), 809–839.

BYLUND, P. L. (2011): "The Role of Specialization in Transaction Cost Analyses of Economic Organizing," Mimeo, University of Missouri.

CAI, H. (2009): "Costly Participation and Heterogeneous Preferences in Informational Committees," *RAND Journal of Economics*, 40(1), 173–189.

CALVERT, R. L. (1985): "The Value of Biased Information: A Rational Choice Model of Political Advice," *Journal of Politics*, 47(2), 530–555.

CALVO, G. A., AND S. WELLISZ (1978): "Supervision, Loss of Control, and the Optimum Size of the Firm," *Journal of Political Economy*, 86(5), 943–952.

——— (1979): "Hierarchy, Ability, and Income Distribution," *Journal of Political Economy*, 87(5), 991–1010.

CHE, Y.-K., AND N. KARTIK (2009): "Opinions as Incentives," *Journal of Political Economy*, 117(5), 815–860.

CHWE, M. S. (1999): "Minority Voting Rights Can Maximize Majority Welfare," *American Political Science Review*, 93(1), 85–97.

COASE, R. H. (1937): "The Nature of the Firm," *Economica*, 4(16), 386–405.

COHEN, R. (2012): "Obama's Team of Idolizers," `http://www.nytimes.com/2012/08/28/opinion/roger-cohen-obamas-team-of-idolizers.html`, Accessed: 20 September 2012.

COUGHLAN, P. J. (2000): "In Defense of Unanimous Jury Verdicts: Mistrials, Communication, and Strategic Voting," *American Political Science Review*, 94(2), 375–393.

CRAWFORD, V. P., AND J. SOBEL (1982): "Strategic Information Transmission," *Econometrica*, 50(6), 1431–1451.

CRÉMER, J. (1980): "A Partial Theory of the Optimal Organization of a Bureaucracy," *Bell Journal of Economics*, 11(2), 683–693.

——— (1995): "Arm's Length Relationships," *Quarterly Journal of Economics*, 110(2), 275–295.

DEGAN, A., AND A. MERLO (2009): "Do Voters Vote Ideologically?," *Journal of Economic Theory*, 144(5), 1868–1894.

DEGROOT, M. H. (1970): *Optimal Statistical Decisions*. McGraw-Hill Book Company.

DEKEL, E., AND B. L. LIPMAN (2010): "How (Not) to Do Decision Theory," *Annual Review of Economics*, 2(1), 257–282.

DESSEIN, W. (2002): "Authority and Communication in Organizations," *Review of Economic Studies*, 69(4), 811–838.

DEWATRIPONT, M., AND J. TIROLE (1999): "Advocates," *Journal of Political Economy*, 107(1), 1–39.

DIXIT, A. K., AND J. W. WEIBULL (2007): "Political Polarization," *Proceedings of the National Academy of Sciences*, 104(18), 7351–7356.

DOBB, M. H. (1928): *Russian Economic Development since the Revolution*. E. P. Dutton and Company.

DOPSON, S., AND R. STEWART (1990): "What Is Happening to Middle Management?," *British Journal of Management*, 1(1), 3–16.

DRUCKER, P. F. (1988): "The Coming of the New Organization," *Harvard Business Review*, 66(1), 45–53.

FEDDERSEN, T. J. (2004): "Rational Choice Theory and the Paradox of Not Voting," *Journal of Economic Perspectives*, 18(1), 99–112.

FEDDERSEN, T. J., AND W. PESENDORFER (1996): "The Swing Voter's Curse," *American Economic Review*, 86(3), 408–424.

——— (1997): "Voting Behavior and Information Aggregation in Elections with Private Information," *Econometrica*, 65(5), 1029–1058.

——— (1998): "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting," *American Political Science Review*, 92(1), 23–35.

——— (1999): "Abstention in Elections with Asymmetric Information and Diverse Preferences," *The American Political Science Review*, 93(2), 381–398.

FLOYD, S. W., AND B. WOOLDRIDGE (1996): *The Strategic Middle Manager: How to Create and Sustain Competitive Advantage.* Jossey-Bass Publishers.

——— (1997): "Middle Management's Strategic Influence and Organizational Performance," *Journal of Management Studies*, 34(3), 465–485.

FRIEBEL, G., AND M. RAITH (2004): "Abuse of Authority and Hierarchical Communication," *RAND Journal of Economics*, 35(2), 224–244.

GAILMARD, S., AND J. W. PATTY (2007): "Slackers and Zealots: Civil Service, Policy Discretion, and Bureaucratic Expertise," *American Journal of Political Science*, 51(4), 873–889.

——— (2012): *Learning While Governing: Information, Accountability, and Executive Branch Institutions.* Book manuscript, University of California at Berkeley.

GARICANO, L. (2000): "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 108(5), 874–904.

GEANAKOPLOS, J., AND P. MILGROM (1991): "A Theory of Hierarchies Based on Limited Managerial Attention," *Journal of the Japanese and International Economies*, 5(3), 205–225.

GEANAKOPLOS, J., AND H. POLEMARCHAKIS (1982): "We Can't Disagree Forever," *Journal of Economic Theory*, 28(1), 192–200.

GERARDI, D., AND L. YARIV (2008): "Costly Expertise," *American Economic Review: Papers & Proceedings*, 98(2), 187–193.

GERLING, K., H. P. GRÜNER, A. KIEL, AND E. SCHULTE (2005): "Information Acquisition and Decision Making in Committees: A Survey," *European Journal of Political Economy*, 21(3), 563–597.

GERSHKOV, A., AND B. SZENTES (2009): "Optimal Voting Schemes with Costly Information Acquisition," *Journal of Economic Theory*, 144(1), 36–68.

GOLTSMAN, M., J. HÖRNER, G. PAVLOV, AND F. SQUINTANI (2009): "Mediation, Arbitration and Negotiation," *Journal of Economic Theory*, 144(4), 1397–1420.

GOODWIN, D. K. (2005): *Team of Rivals: The Political Genius of Abraham Lincoln.* Simon & Schuster.

GORDON, D. M. (1996): *Fat and Mean: The Corporate Squeeze of Working Americans and the Myth of Managerial "Downsizing".* The Free Press.

GRATTON, L. (2011): "The End of the Middle Manager," *Harvard Business Review*, 89(1–2), 36.

Gulick, L. (1937): "Notes on the Theory of Organization," in *Papers on the Science of Administration*, ed. by L. Gulick, and L. Urwick, chap. 1, pp. 1–45. Institute of Public Administration.

Guth, W. D., and I. C. MacMillan (1986): "Strategy Implementation versus Middle Management Self-Interest," *Strategic Management Journal*, 7(4), 313–327.

Harris, M., and A. Raviv (2005): "Allocation of Decision-making Authority," *Review of Finance*, 9(3), 353–383.

Hayek, F. A. (1945): "The Use of Knowledge in Society," *American Economic Review*, 35(4), 519–530.

Hirsch, A. V. (2011): "Experimentation and Persuasion in Political Organizations," Mimeo, Princeton University.

Holmström, B., and J. Tirole (1989): "The Theory of the Firm," in *Handbook of Industrial Organization*, ed. by R. Schmalensee, and R. D. Willig, vol. 1, chap. 2, pp. 63–133. Elsevier Science Publishers B.V.

Huy, Q. N. (2001): "In Praise of Middle Managers," *Harvard Business Review*, 79(8), 72–79.

Ivanov, M. (2010): "Communication via a Strategic Mediator," *Journal of Economic Theory*, 145(2), 869–884.

Kainen, T. L., and D. P. Boyd (2007): "Who Succeeds in the Murky Middle?," *Journal of Applied Business Research*, 23(4), 61–68.

Kamenica, E., and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101(6), 2590–2615.

Keren, M., and D. Levhari (1979): "The Optimum Span of Control in a Pure Hierarchy," *Management Science*, 25(11), 1162–1172.

Koriyama, Y., and B. Szentes (2009): "A Resurrection of the Condorcet Jury Theorem," *Theoretical Economics*, 4(2), 227–252.

Krishna, V., and J. Morgan (2012): "Voluntary Voting: Costs and Benefits," *Journal of Economic Theory*, 147(6), 2083–2123.

Ladha, K. K. (1992): "The Condorcet Jury Theorem, Free Speech, and Correlated Votes," *American Journal of Political Science*, 36(3), 617–634.

Lazear, E. P., K. L. Shaw, and C. T. Stanton (2012): "The Value of Bosses," NBER Working Paper 18317.

Leavitt, H. J. (2003): "Why Hierarchies Thrive," *Harvard Business Review*, 81(3), 96–102.

Leavitt, H. J., and T. L. Whisler (1958): "Management in the 1980's," *Harvard Business Review*, 36(6), 41–48.

Li, H. (2001): "A Theory of Conservatism," *Journal of Political Economy*, 109(3), 617–636.

McLennan, A. (1998): "Consequences of the Condorcet Jury Theorem for Beneficial Information Aggregation by Rational Agents," *American Political Science Review*, 92(2), 413–418.

McMurray, J. C. (2013): "Aggregating Information by Voting: The Wisdom of the Experts versus the Wisdom of the Masses," *Review of Economic Studies*, 80(1), 277–312.

Melumad, N. D., and D. Mookherjee (1989): "Delegation as Commitment: The Case of Income Tax Audits," *RAND Journal of Economics*, 20(2), 139–163.

Mollick, E. (2011): "People and Process, Suits and Innovators: Individuals and Firm Performance," Mimeo, Wharton School of Management, University of Pennsylvania.

Mookherjee, D. (2006): "Decentralization, Hierarchies, and Incentives: A Mechanism Design Perspective," *Journal of Economic Literature*, 44(2), 367–390.

——— (2012): "Incentives in Hierarchies," in *Handbook of Organizational Economics*, ed. by R. Gibbons, and J. Roberts, chap. 19, pp. 764–799. Princeton University Press.

Muller, R. A. (2011): "The Case against Global-Warming Skepticism," `http://online.wsj.com/article/SB10001424052970204422404576594872796327348.html`, Accessed: 5 November 2012.

——— (2012): "The Conversion of a Climate-Change Skeptic," `http://www.nytimes.com/2012/07/30/opinion/the-conversion-of-a-climate-change-skeptic.html`, Accessed: 5 November 2012.

Myerson, R. B. (1998): "Population Uncertainty and Poisson Games," *International Journal of Game Theory*, 27(3), 375–392.

Osterman, P. (2008): *The Truth about Middle Managers*. Harvard Business Press.

Prendergast, C. (2007): "The Motivation and Bias of Bureaucrats," *American Economic Review*, 97(1), 180–196.

Qian, Y. (1994): "Incentives and Loss of Control in an Optimal Hierarchy," *Review of Economic Studies*, 61(3), 527–544.

Radner, R. (1992): "Hierarchy: The Economics of Managing," *Journal of Economic Literature*, 30(3), 1382–1415.

———— (1993): "The Organization of Decentralized Information Processing," *Econometrica*, 61(5), 1109–1146.

Rogoff, K. (1985): "The Optimal Degree of Commitment to an Intermediate Monetary Target," *Quarterly Journal of Economics*, 100(4), 1169–1189.

Rosoff, M. (2011): "So Microsoft Is Cutting Its Bloated Middle Management? Great!," `http://www.businessinsider.com/so-microsoft-is-cutting-its-bloated-middle-management-great-2011-9`, Accessed: 21 October 2012.

Rotemberg, J. J., and G. Saloner (1995): "Overt Interfunctional Conflict (and Its Reduction through Business Strategy)," *RAND Journal of Economics*, 26(4), 630–653.

———— (2000): "Visionaries, Managers, and Strategic Direction," *RAND Journal of Economics*, 31(4), 693–716.

Sah, R. K., and J. E. Stiglitz (1985): "Human Fallibility and Economic Organization," *American Economic Review*, 75(2), 292–297.

Samuelsohn, D. (2010): "John Shimkus Cites Genesis on Climate Change," `http://www.politico.com/news/stories/1110/44958.html`, Accessed: 5 November 2012.

Shin, H. S. (1998): "Adversarial and Inquisitorial Procedures in Arbitration," *RAND Journal of Economics*, 29(2), 378–405.

Silverman, R. E. (2012): "Who's the Boss? There Isn't One," `http://online.wsj.com/article/SB10001424052702303379204577474953586383604.html`, Accessed: 30 October 2012.

Smith, A. (1776): *An Inquiry into the Nature and Causes of the Wealth of Nations.* W. Strahan and T. Cadell.

Sobel, J. (2010): "Giving and Receiving Advice," Mimeo, University of California at San Diego.

Souder, W. E. (1987): "Stimulating and Managing Ideas," *Research Management*, 30(3), 13–17.

Sutton, R. I. (2010): "If You're the Boss, Start Killing More Good Ideas," `http://blogs.hbr.org/sutton/2010/08/if_youre_the_boss_start_killin.html`, Accessed: 22 October 2012.

Szalay, D. (2005): "The Economics of Clear Advice and Extreme Options," *Review of Economic Studies*, 72(4), 1173–1198.

TEECE, D. J. (1996): "Firm Organization, Industrial Structure, and Technological Innovation," *Journal of Economic Behavior & Organization*, 31(2), 193–224.

TORRINGTON, D., AND J. WEIGHTMAN (1982): "Technical Atrophy in Middle Management," *Journal of General Management*, 7(4), 5–17.

USHER, A. P. (1920): *An Introduction to the Industrial History of England.* Houghton Mifflin Company.

VAN DEN STEEN, E. J. (2005): "Organizational Beliefs and Managerial Vision," *Journal of Law, Economics, and Organization*, 21(1), 256–283.

——— (2010a): "Culture Clash: The Costs and Benefits of Homogeneity," *Management Science*, 56(10), 1718–1738.

——— (2010b): "Disagreement and the Allocation of Control," *Journal of Law, Economics, and Organization*, 26(2), 385–426.

VAN ZANDT, T. (1999): "Real-Time Decentralized Information Processing as a Model of Organizations with Boundedly Rational Agents," *Review of Economic Studies*, 66(3), 633–658.

WANG, H., AND K. F. E. WONG (2012): "The Effect of Managerial Bias on Employees' Specific Human Capital Investments," *Journal of Management Studies*, 49(8), 1435–1458.

WILLIAMSON, O. E. (1967): "Hierarchical Control and Optimum Firm Size," *Journal of Political Economy*, 75(2), 123–138.

WOOLDRIDGE, B., AND S. W. FLOYD (1990): "The Strategy Process, Middle Management Involvement, and Organizational Performance," *Strategic Management Journal*, 11(3), 231–241.