

**The Democratic Dilemma**

**Can Citizens Learn What They Need to Know?**

**Arthur Lupia**

**Mathew D. McCubbins**

## **Contents**

---

*Table of Contents*

*List of Tables*

*List of Figures*

*Series Editors' Preface*

*Acknowledgments*

### **1 Knowledge and the Foundation of Democracy**

*Democracy, Delegation, and Reasoned Choice*

*A Preview of Our Theory*

*Plan of the Book*

## **PART I: Theory**

### **2 How People Learn**

***Attention! This is How we Learn***

***The Cognitive Stock Market***

***Attention and Connections***

***Conclusion***

**3 How People Learn From Others**

***The Aristotelian Theories of Persuasion***

***Our Theory of Persuasion***

***Dynamic Implications***

***Persuasive Implications***

***Conclusion***

***Food for Thought***

**4 What People Learn From Others**

***The Conditions for Enlightenment***

***The Conditions for Deception***

***Discussion: How We Choose Whom to Believe***

***Conclusion***

**5 Delegation and Democracy**

*The Dilemma of Delegation*

*A Theory of Delegation with Communication*

*What It all Means*

*Conclusion*

## **PART II: Experiments**

### **6 Theory, Predictions, and the Scientific Method**

### **7 Laboratory Experiments on Information, Persuasion, and Choice**

*Experimental Design*

*Experiments on Persuasion and Reasoned Choice*

*Summary*

### **8 Laboratory Experiments on Delegation**

*Experimental Design*

*Experiments on Delegation*

*Conclusion*

## **9 A Survey Experiment on the Conditions for Persuasion**

*Description of the Experiment*

*Analysis*

*Conclusion*

## **PART III: Implications for Institutional Design**

## **10 The Institutions of Knowledge**

*Electoral Institutions*

*Legislative Institutions*

*Bureaucratic Institutions*

*Legal Institutions*

*Unenlightening Democratic Institutions*

*Conclusion*

*Afterword*

**Appendix to Chapter 2**

**Appendix to Chapter 3**

**Appendix to Chapter 5**

*References*

*Author Index*

*Subject Index*

***List of Tables***

Table 3-1:	Numerical Examples
Table 6-1:	Predictions
Table 7-1:	Summary of Results
Table 7-2:	Theoretical Premises and Experimental Analogies
Table 7-3:	Speaker Behavior in the Benchmark Trials
Table 7-4:	Speaker Behavior in the Control Condition
Table 7-5:	Speaker Behavior in the Treatment Conditions
Table 7-6:	A Summary of Speaker and Principal Behavior
Table 8-1:	Theoretical Premises and Experimental Analogies
Table 8-2:	Expectations about Agent Behavior in the Knowledge Experiment
Table 8-3:	Agent Behavior in the Speaker Knowledge Experiment
Table 8-4:	Expectations about Agent Behavior in the Speaker Interest, Penalty for Lying, and Verification Experiments
Table 8-5:	Agent Behavior in the Speaker Interest Experiment
Table 8-6:	Agent Behavior in the Penalty for Lying Experiment
Table 8-7:	Agent Behavior in the Verification Experiment

- Table 9-1: Percent reporting that spending money to build more prisons is a good idea
- Table 9-2: Percent who report that spending money to build more prisons is a “good” idea, separated by perceived speaker knowledge or trust
- Table 9-3: Percent who report that spending money to build more prisons is a “good” idea, separated by perceived speaker knowledge *and* trust
- Table 9-4: Respondents who heard either “Rush Limbaugh supports” or “Phil Donahue supports.” Percent who reported that stated issue is a good idea, separated by knows, agrees, and feeling thermometer score
- Table 9-5: Respondents who heard either “Rush Limbaugh opposes” or “Phil Donahue opposes.” Percent who reported that stated issue is a good idea, separated by knows, agrees, and feeling thermometer score
- Table 9-6: Percent reporting that stated issue is a good idea, separated by knows, agrees, and speaker ideology

## List of Figures

- Figure 3-1: Our Basic Model of Persuasion
- Figure 3-2: The Effect of Costly Action
- Figure 4-1: The Conditions for Persuasion and Enlightenment
- Figure 4-2: The Conditions for Deception
- Figure 5-1: The Basic Model of Delegation
- Figure 5-2: Delegation with Communication
- Figure 5-3: The Incentive Condition with the Persuasive Speaker as a Verifier
- Figure 5-4a: Steps to the Knowledge Condition
- Figure 5-4b: Conditions for Successful Delegation
- Figure 7-1: Knowledge Experiment: Control Condition
- Figure 7-2: Knowledge Experiment: Treatment Condition
- Figure 7-3: Knowledge Experiment: Treatment Condition with 70% Knowledge
- Figure 7-4: Knowledge Experiment: Conflicting Interests and No Knowledge
- Figure 7-5: Interest Experiment: Control Condition
- Figure 7-6: Interest Experiment: Treatment Condition
- Figure 7-7: Interest Experiment: Variations on the Treatment Condition

- Figure 7-8: Penalty for Lying Experiment: Treatment Condition
- Figure 7-9: Verification Experiment: Treatment Condition
- Figure 8-1: Knowledge Experiment
- Figure 8-2: Interest Experiment
- Figure 8-3: Penalty for Lying Experiment
- Figure 8-4: Verification Experiment
- Figure 10-1: How a Proposal Becomes a Policy in the House of  
Representatives, Highlighting Aspects of Party Control

## **Series Editors' Preface**

## Acknowledgments

*The Democratic Dilemma* draws from the lessons of a wide array of scholarship to provide a cohesive and positive statement about the political consequences of limited information. Skilled practitioners in many different disciplines have challenged us to give the strongest possible argument while expressing ourselves to a wide academic audience. As a result, we supplement our formal arguments, mathematical models, laboratory experiments, and national surveys with metaphors, analogies, and anecdotes that specialists and nonspecialists alike can understand.

We owe many debts of gratitude. We acknowledge the support of the National Science Foundation and its Political Science Program, through grant SBR-9422831. Dr. Lupia acknowledges a UCSD COR grant for partial support of the research reported here, and for a grant, administered by Paul Sniderman, for the 1994 Multi-Investigator Survey, from the National Science Foundation, number SBR-9309946. Early drafts of this book were written while Dr. McCubbins was a Fellow at the Center for Advanced Study in the Behavioral Sciences. Dr. McCubbins is grateful for the financial support for this fellowship, provided by the National Science Foundation, grant SBR-9022192.

We presented portions of this book at many conferences and seminars and received comments and advice for which we are grateful. Two seminars, lasting several days each, were especially useful to us in revising this book. The first was at the Department of Government, Harvard University, in May 1995; the second was at the Hoover Institution of Stanford University in January 1996. We thank Jim Alt, Barry Weingast, and Doug North for these opportunities.

We received comments on this manuscript from nearly one hundred people. We thank all of them and will repay our debt in kind. Of these, several scholars showed tremendous determination and ingenuity of argument in their efforts to steer the course of this project. To these individuals we owe a special debt: Jim Alt, Randall Calvert, Gary Cox, Vince Crawford, John Ferejohn, Elisabeth Gerber, James Kuklinski, Douglass North, Samuel Popkin, Kenneth Shepsle, Joel Sobel, Paul Sniderman, Tracy Strong, Peter Tyack, Mark Turner, Barry Weingast, and Oliver Williamson.

We also owe a great debt to a number of hardworking research assistants: Scott Basinger, Morganne Beck, Greg Bovitz, Andrea Campbell, Carola Clift, Michael Ensley, Jennifer Nicoll, Julie Pope, and

Robert Schwartz. Perhaps more than any other individual's, we are grateful for the hard work and keen insight of James Druckman.

Finally, we owe a debt of gratitude to our children, Francesca, Colin, and Kenny. They were often our first experimental subjects and were the source of many of our examples.

## Chapter 1

### Knowledge and the Foundation of Democracy

Knowledge will forever govern ignorance, and a people who mean to be their own governors, must arm themselves with the power knowledge gives. A popular government without popular information or the means of acquiring it, is but a prologue to a farce or a tragedy or perhaps both.

--James Madison<sup>1</sup>

The founders of the American republic, and many of their contemporaries around the world, believed that democracy requires voters, legislators and jurors to make reasoned choices. Reasoned choice, in turn, requires that people know the consequences of their actions.

Are voters, legislators and jurors capable of making reasoned choices? Many observers conclude that they are not. The evidence for this conclusion is substantial -- study after study documents the breadth and depth of citizen ignorance. Making matters worse is the manner in which people acquire the little information they have. Most voters, for example, get their information from thirty-minute news summaries, thirty-second political advertisements or eight-second sound bites. As a consequence of their ignorance, a strong possibility exists that "Men of factious tempers, of local prejudices, or of sinister designs,

may, by intrigue, by corruption, or by any other means, first obtain the suffrages, and then betray the interests of the people” (Madison, *Federalist* 10).

It is widely believed that there is a mismatch between the requirements of democracy and most citizens’ ability to meet these requirements. If this mismatch is too prevalent, then effective self governance is impossible. The *democratic dilemma* is that the people who are called upon to make reasoned choices may not be capable of doing so.

In this book, we concede that people lack political information. We also concede that this ignorance can allow people “of sinister designs” to deceive and betray the underinformed. We do not concede, however, that democracy *must* succumb to these threats. Rather, we conclude that:

- Reasoned choice does not require full information; rather, it requires the ability to predict the consequences of actions. We define this ability as knowledge.<sup>2</sup>
- People *choose* to disregard most of the information they could acquire and base virtually all of their decisions on remarkably little information.
- People often *substitute* the advice of others for the information they lack. This substitution can give people the capacity for reasoned choice.

- Relying on the advice of others involves tradeoffs. While it decreases the costs of acquiring knowledge, it also introduces the possibility of deception.
- A person who wants to gain knowledge from the advice of others must choose to follow some advice while ignoring other advice. People make these choices in systematic and predictable ways.
- Political institutions can help people choose which advice to follow and which advice to ignore. Institutions do this when they *clarify* the incentives of advice givers.
- Understanding when and how people learn, and understanding how political institutions affect learning, help us to understand the resolution of the democratic dilemma.

In the remainder of this chapter we foreshadow our argument, and provide a roadmap of the rest of the book.

## **Democracy, Delegation, and Reasoned Choice**

Democracy is a method of government based upon the choices of the people. In all modern democracies, the people elect or appoint others to represent them. Legislative assemblies, executives, commissions, judges, and juries are

empowered by the people to make collective decisions on their behalf. These delegations form the foundation of democracy.

But there are dangers. As Dahl (1967: 21) warns, the principal danger is that uninformed decision makers, by failing to delegate well, will transform democracy into a *tyranny of experts*: “there are decisions that require me to *delegate* authority to others...but if I delegate, may I not, in practice, end up with a kind of aristocracy of experts, or even false experts?”

Must democracy become a tyranny of experts? Many observers answer yes, because those who delegate seem uninformed when compared to those to whom they delegate.

The principal democratic delegation, that of the people electing their governors, seems most susceptible to tyranny. Cicero’s observation that “in the common people there is no wisdom, no penetration, no power of judgment,” is an apt summary of modern voting studies (see Berelson 1952, Campbell et. al. 1960, Converse 1964, Lane and Sears 1964, Kinder and Sears 1985, Luskin 1987, McClosky 1964, Neuman 1986, Schattschneider 1960, Schumpeter 1942, Zaller 1992, Zaller and Feldman 1992; for a survey see Delli Carpini and Keeter 1996). Many scholars argue that voters, because of their obstinance or inability to educate themselves, become the unwitting puppets of campaign and media puppetmasters (Bennett 1992, Sabato 1991). Iyengar (1987: 816) summarizes the literature on

voting and elections, “the low level of political knowledge and the absence of ideological reasoning has lent credence to the charges that popular control of government is illusory.”

Other observers make similar arguments about elected representatives. Weber, for example, argues that bureaucrats use expertise to overwhelm legislators:

Under normal conditions, the power position of a fully developed bureaucracy is always overtowering. The ‘political master’ finds himself in the position of the ‘dilettante’ who stands opposite the ‘expert,’ facing the trained official who stands within the management of administration. This holds whether the ‘master’ whom the bureaucracy serves is a ‘people,’ equipped with the weapons of ‘legislative initiative,’ the ‘referendum,’ and the right to remove officials, or a parliament, elected on a more aristocratic or more ‘democratic’ basis and equipped with the right to vote a lack of confidence, or with the actual authority to vote it (Weber quoted in Gerth and Mills 1946: 232).

Niskanen (1971) continues that public officials’ inability to contend with the complexities of modern legislation places them at the mercy of self-serving special interests and bureaucrats. Lowi (1979: xii) concludes “actual policy-making will not come from voter preferences or congressional enactments but

from a process of tripartite bargaining between the specialized administrators, relevant members of Congress, and the representatives of self-selected organized interests.”

Jurors also seem to lack the information they need. Posner (1995: 52), for example, argues, “As American law and society become ever more complex, the jury’s cognitive limitations will become ever more palpable and socially costly.” Other observers characterize the legal system, not as a forum where citizens make reasoned choices, but as a stage for emotional appeals where style and deception overwhelm knowledge. As Abramson (1994: 3) laments,

The gap between complexity and modern litigation and the qualification of jurors has widened to frightening proportions. The average jury rarely understands the expert testimony in an anti-trust suit, a medical malpractice case, or an insanity defense. Nor do most jurors know the law or comprehend the judge’s crash course of instruction on it. Trial by jury has thus become trial by ignorance.

While the critiques of democracy’s delegations are myriad and diverse, they all share a common conclusion -- *reasoned choice does not govern delegation*. As Schumpeter (1942: 262) argues, “the typical citizen drops down to a lower level of mental performance as soon as he enters the political field. He

argues and analyzes in a way which he would readily recognize as infantile within the sphere of his real interests. He becomes a primitive again. His thinking is associative and affective...this may prove fatal to the nation.”

If voters, legislators, and jurors lack the capability to delegate effectively, then democracy may be “but a prologue to a farce or a tragedy.” Like the scholars just quoted, we find this possibility alarming. Unlike these scholars, however, we will argue that the capabilities of the people and the requirements of democracy are not as mismatched as these critics would have us believe. In what follows, we will identify the conditions when this mismatch does and does not exist.

## **A Preview of Our Theory**

### ***Knowledge and Information***

We show that *limited information need not prevent people from making reasoned choices*. We begin in Chapter Two by asking how humans cope with complexity and scarcity. As Simon (1979: 3) argues, “human thinking powers are very modest when compared with the complexities of the environments in which human beings live.” Making matters worse is the fact that many of the resources people need to survive are scarce.

*Ironically, for many political issues, information is not scarce; rather it is the cognitive resources that a person can use to process information that are scarce.* For example, political information appears in the newspapers, in the mail,

on community bulletin boards, on television and radio, and is relayed to us in person by friends and family. People often lack the time and energy needed to make sense of all this information. As a consequence, people have only incomplete information. Fortunately, reasoned choice does not require complete information. Instead, it requires **knowledge: the ability to predict the consequences of actions.**<sup>3</sup>

Implicit in many critiques of democracy is the claim that people who lack *information* are incapable of reasoned choice. By contrast, we argue that people who lack information solve enormously complex problems every day. They do so by making effective use of the information available to them, sorting that which is useful from that which is not.

*Information is useful only if it helps people avoid costly mistakes.* By contrast, if more information does not lead people to change their decisions, then it provides no instrumental benefit and they should ignore it. Indeed, ignoring useless information is necessary for humans and other species to survive and prosper (Churchland and Sejnowski 1992).

Those who find such statements surprising should consider the almost boundless range of actions, both mundane and grand, for which people ignore available information. For example, people take medication without knowing all of the conditions under which it is harmful. They also buy houses based on limited

information about the neighborhoods around them and with little or no information about the neighbors. People make choices in this way not because the information is unavailable, but because the costs of paying attention to it exceed the value of its use.<sup>4</sup>

While reasoned choice does not require complete information, it does require the ability to predict the consequences of actions. In many cases, simple pieces of information can provide the knowledge people need. For example, to successfully navigate a busy intersection, you must *know* where all of the other cars are going to be sure that you can avoid crashing into them. Advocates of complete information might argue that successful navigation requires as much information as you can gather about the intentions of other drivers and the speed, acceleration, direction, and mass of their cars. At many intersections, however, there is a simple substitute for *all* of this information -- a traffic light. At these intersections, traffic lights are substitutes for more complex information and reduce the amount of information required to make a reasoned choice. At intersections without working traffic lights or other simple cues, reasoned choices require more information. Using similar logic, it follows that limited information precludes reasoned choice only if people appear to be stuck at complex political intersections and lack access to effective political traffic lights.

***Persuasion, Enlightenment, and Deception***

People who want to make reasoned choices need knowledge. There are two ways to acquire knowledge. The first way is to draw from personal experience. People who exercise this option use their own observations of the past to derive predictions about the future consequences of their actions. The second way is to learn from others. People who exercise this option substitute other peoples' observations of the past for the personal experience they lack.

In many political settings, only the second option is available. This is true because politics is often abstract and its consequences remote. In these settings, personal experience does not provide sufficient knowledge for reasoned choice. Therefore, for many political decisions, reasoned choice requires learning from others.

There are many explanations of how people learn from others. Indeed, a generation of scholars, starting with Knight, Simon, Berelson et al., and Downs, suggest numerous heuristics -- simple means for generating information substitutes.<sup>5</sup> Examples include opinion leaders (Berelson et al. 1954), party identification (Downs 1957), biased information providers (Calvert 1985), campaign events (Popkin 1991), campaign information (Lodge et al. 1995), history (Downs 1957, Key 1966, Fiorina 1981), polls (McKelvey and Ordeshook 1986), costly action (Lupia 1992), "fire alarms" (McCubbins and Schwartz 1984), people who have similar interests (Sniderman et al. 1991, Krehbiel 1991),

demographics (Popkin et al. 1976), competition (Milgrom and Roberts 1986), interest group endorsements (Lupia 1994), and the media (Iyengar and Kinder 1987, Page et al. 1987).

Individually, each of these explanations of how we learn from others is valuable and enlightening. Each reveals a source of the judgmental shortcuts that people undoubtedly use. However, as Sniderman et al. (1991: 70) argue, “The most serious risk is that...every correlation between independent and dependent variables [is] taken as evidence of a new judgmental shortcut.” We agree. We need a theory that explains when or how people choose among the shortcuts listed above. To understand how people learn from others, we must be able to explain *how people choose whom to believe*.

In Chapter Three, we explain *how* people learn from others. This explanation answers questions such as “Who can learn from whom?” In Chapter Four, we explain *what* people learn from others. This explanation answers questions such as “When is learning from others a sufficient substitute for personal experience as the basis of reasoned choice?” and “When does relying on the testimony of others prevent reasoned choice?”

In Chapters Three and Four, we show that learning from others is no trivial matter. To see why, notice that any attempt to learn from others leads to one of three possible outcomes.

- **The first outcome is *enlightenment*.** When someone furnishes us with knowledge, we become enlightened. Enlightenment, then, is the process of becoming enlightened. If we initially lack knowledge sufficient for reasoned choice and can obtain such knowledge only from others, then we can make reasoned decisions only if others enlighten us.
- **The second outcome is *deception*.** Deception is the process by which the testimony we hear reduces our ability to predict accurately the consequences of our actions.
- **The third outcome is that we *learn nothing*.** When we learn nothing, our beliefs go unchanged and we gain no knowledge.

Both enlightenment and deception, in turn, require **persuasion: a successful attempt to change the beliefs of another**. The key to understanding whether people become enlightened or deceived by the testimony of others is to understand the conditions under which they can persuade each other.

Most scholars of communication and politics, dating back to Aristotle, focus on a speaker's *internal character* (e.g., honesty, ideology, or reputation) as a necessary condition for persuasion. If a speaker lacks the right character, then these scholars conclude that the speaker will not be persuasive. In Chapter Three, we derive a different set of necessary and sufficient conditions for persuasion. We show that persuasion is not contingent upon personal character; rather, *persuasion*

*requires that a listener perceive a speaker to be both knowledgeable and trustworthy.* While a perception of trust can arise from a positive evaluation of a speaker's character, we show that *external forces can substitute for character*, and can thus generate persuasion in contexts where it would not otherwise occur.

An example of an external force that generates trust and persuasion is a listener's observation of a speaker's costly effort. From this observation, the listener can learn about the intensity of a speaker's preferences. This particular condition is also very much like the old adage that actions speak louder than words. When speaker costs have this effect, they can provide a basis for trust by providing listeners with a window to speaker incentives. For example, suppose that a listener knows a speaker to have one of three possible motivations – he is a conservative with intense preferences, a conservative with non-intense preferences, or a liberal with non-intense preferences – but does not know which he actually has. Moreover, suppose that the listener can make a reasoned choice only if she knows whether the speaker is liberal or conservative. Suppose further that, if she observes that the speaker paid a quarter of his income to affect a policy outcome, then she can conclude that the speaker has intense preferences. Therefore, she can infer that the speaker is a conservative and can make a reasoned choice.

Another example of a trust-inducing external force is a penalty for lying. Penalties for lying, whether explicit, such as fines for perjury, or implicit, such as the loss of a valued reputation, can also generate trust by revealing a speaker's incentives. That is, while a listener may believe that a speaker has an interest in deception, the presence of a penalty for lying may lead the listener to believe that certain types of lies are prohibitively costly, rendering certain types of statements very likely to be true.

Our conditions for persuasion show when forces like these are, and are not, effective substitutes for a speaker's character.<sup>6</sup> These conditions reveal that you do not necessarily learn more from people who are like you, nor do you learn more from people you like. This is why most people turn to financial advisors, instead of their mothers, when dealing with mutual funds, and back to Mom when seeking advice about child-rearing.

Another way to frame this lesson is as follows: our conditions for persuasion show why some statements are persuasive and others are not. The obvious reason for these differences is that statements vary in content. The less obvious reason is that the context under which a speaker makes a statement also affects persuasion considerably. Two people making precisely the same statement may not be equally persuasive if only one is subject to penalties.

Our conditions for persuasion further imply that not everyone can persuade. People listen to some speakers and not others. They read some books and not others. They buy some products even though the manufacturers spend very little money on advertising, while refusing to buy others supported by celebrity endorsements. Similarly, people respond to the advice of some experts or interest groups and not others. Our conditions for persuasion explain how people make these choices.

Our results also reveal the bounds on the effectiveness of the heuristics mentioned earlier. Consider, for example, the use of ideology as a heuristic. When there is a high correlation between a speaker's ideology and that speaker's knowledge and trustworthiness, then people are likely to find ideological cues useful. By contrast, when there is no clear correlation, ideology is useless. Similar arguments can be made about other heuristics, such as party, reputation, and likability. In sum, *concepts such as reputation, party or ideology are useful heuristics only if they convey information about knowledge and trust. The converse of this statement is not true* – knowledge and trust are the fundamental factors that make cues persuasive, the other factors are not.

In Chapter Four, we shift our focus from identifying the conditions for persuasion to identifying the conditions for enlightenment and deception. The key to enlightenment is that a listener has accurate beliefs about the speaker's

knowledge and incentives. The key to deception is that the listener has inaccurate beliefs about these factors.<sup>7</sup> When nature, cultural norms, or the structure of political institutions provide listeners with a window to a speaker's interests, knowledge, and incentives, then the context is ripe for enlightenment. Otherwise people who attempt to learn from others are likely to be deceived. We conclude Chapter Four by arguing that reasoned choice is therefore impossible only when there is limited information and the conditions for enlightenment do not exist and cannot be created. Together, Chapters Three and Four clarify the relationship between limited information and reasoned choice.

### ***Successful Delegation and The Institutions of Knowledge***

In Chapter Five we use the lessons of Chapters Two, Three, and Four to clarify the political consequences of limited information. We begin with the observation that modern democracy requires delegation. We then show that delegation has three possible consequences -- it can succeed, it can fail, or it can have no effect. We say delegation succeeds when an agent, the person or persons to whom authority is delegated, enhances the welfare of a principal, the person or persons who delegated. We say delegation fails when an agent reduces a principal's welfare. Delegation has no effect when an agent's actions do not affect a principal's welfare.

Two reasons are commonly cited for the failure of delegation: principals and agents have conflicting interests over the outcome of delegation, and agents have expertise regarding the consequences of the delegation that principals do not (for a survey, see Kiewiet and McCubbins 1991, Miller 1992). When delegation occurs under these conditions, agents are free to take any action that suits them, irrespective of the consequences for the principal, and the principal cannot cause them to do otherwise.

We find that delegation succeeds if two conditions are satisfied: the knowledge condition and the incentive condition. The knowledge condition is satisfied in one of two ways. First, it is satisfied when the principal's personal experience allows her to distinguish beneficial from detrimental agent actions. Second, it is satisfied when the principal can obtain this knowledge from others. Therefore, the knowledge condition does not require the principal to know everything the agent knows. It requires only that the principal know enough to distinguish welfare-enhancing from welfare-decreasing agent actions.

The incentive condition is satisfied when the agent and the principal share at least some goals in common. In many cases, satisfaction of the knowledge condition is sufficient for satisfaction of the incentive condition: if a principal can become enlightened with respect to the consequences of delegation, then she can

either motivate the agent to take actions that enhance her welfare, or she can reject the agent's actions that do not enhance her welfare.

We find that the outcome of delegation is not determined by whether or not the principal can match the agent's technical expertise. Instead, it is determined by the principal's ability to use the testimony of others effectively. If the principal has this ability, then delegation can succeed despite the information she lacks. If the principal lacks information about the agent and lacks the ability to learn from others, then delegation is doomed.

Moreover, we argue that, if democratic principals can create the context in which knowledgeable and persuasive speakers can inform them of the consequences of their agent's actions, then they can facilitate successful delegation. We argue that institutions, such as administrative procedure, rules of evidence, and statutory law, provide the context under principals can learn about their agent's actions. Institutions can, if properly structured, offer principals a way to better judge their agent's actions. When institutions are poorly designed, or the incentives they induce are opaque, then the political consequence of limited information is likely to be failed delegation. By contrast, when these institutions properly and clearly structure incentives, then they facilitate enlightenment, reasoned choice, and successful delegation even in complex circumstances.

**Conclusion**

The mismatch between what delegation demands and citizens' capability constitutes the democratic dilemma. If people are not capable of reasoned political choices, then effective self governance is an illusion. Observing that voters, legislators, and jurors are ignorant of many of the details of the decisions they face, many scholars and political commentators conclude that the illusion is real and argue for some type of reform. If their conclusion is correct, then effective self governance may indeed require political reform. If, however, their conclusion is incorrect, their reforms may restrain the truly competent and do more harm than good.

Other scholars have made the observation that people are quite capable of making complex decisions with very little information. They point to instances where people use heuristics and conclude that such heuristics are sufficient for reasoned choice. If these conclusions are correct, then successful delegation does not require reform and the critics mentioned above are akin to democracy's Chicken Littles. If, however, these latter conclusions are incorrect, then these latter scholars are akin to democracy's PollyAnnas, leading us to perpetuate an ineffective and harmful system of governance.

Both sides of this debate recognize that people are often ignorant about the details of the choices they make. They also both recognize the existence of information short cuts, cues, and heuristics. What is missing from this debate is

an understanding of when ignorance prevents reasoned choice, how people choose among potential heuristics, and when these heuristics provide effective substitutes for the information people lack. Only when we know these things will we be able to make constructive use of the observation that people lack information. At that point, we can separate the Chicken Littles from the PollyAnnas and build solutions to the Democratic Dilemma.

### **Plan of the Book**

This book has two parts. In Part One, containing Chapters Two through Five, we develop the theories just described. In Part Two, which contains Chapters Six through Ten, we test the crucial hypotheses about learning, persuasion, reasoned choice and delegation that we produce in Part One.

In Chapter Six, we define a set of standards for empirical research that motivates the experiments we conducted. In Chapters Seven and Eight, we describe a series of laboratory experiments. In Chapter Seven, we use laboratory experiments to evaluate the predictive strength of our conditions for persuasion, enlightenment and deception. In Chapter Eight, we use laboratory experiments to evaluate our theoretical predictions about delegation. In Chapter Nine, we describe a survey experiment about persuasion. In Chapter Ten, we examine democratic institutions from the United States and elsewhere, and show how they do or do not provide the context for successful delegation. Finally we discuss how

to reform institutions to stack the deck in favor of reasoned choice and successful delegation.

### Footnotes

---

1 From Hunt (1910: 103). Madison expressed similar beliefs in *Federalist 57*, and in a speech before the Virginia Ratifying Convention where he argued that it is necessary that the people possess the “virtue and intelligence to select men of virtue and wisdom” (Riemer 1986: 40).

2 There exists a centuries-old debate about what democracy *should* do. This debate has involved many great minds, is wide-ranging, and is totally unresolved. We do not believe ourselves capable of resolving this debate. However, we strongly believe that we can make the debate more constructive. We can do so by clarifying the relationship between what information people have, what they know, and what types of decisions they can make. Our book is firmly about determining the capabilities of people who lack political information. It is designed to resolve debates about what skills voters, legislators, and jurors have and debates about how much information people need to have these skills. So, while our book may help to clarify debates about what democracy should do, it will not resolve these debate.

We mention this because our relationship to the debate about what democracy should do motivates our definition of reasoned choice. Our definition of reasoned choice allows any of us to define what amount of knowledge is required for reasoned choice. So, some of us may argue that a reasoned choice

---

requires knowledge of very technical matters, while others may argue that a reasoned choice requires less knowledge. Our definition of reasoned choice is purposefully precise with respect to the relationship between information, knowledge, and choice and is purposefully vague with respect to the debate about what democracy should do. Thus, whatever we want democracy to be, our effort here will help us decide whether or not people are capable of accomplishing it.

3 For example, knowing which of two products is “better” than the other is often sufficient for us to make the same choice we would have made had we been completely informed about each product.

4 Furthermore, beyond being useless, some types of information cause people to make the wrong (i.e., welfare-reducing) choices when they would have otherwise made the right (i.e., welfare-increasing) ones with less information. For example, if a person votes for Jones instead of Smith because a newspaper endorses Jones, she may regret having attended to this additional information when Jones later opposes a policy that both she and Smith support.

5 Also, see Key (1966) and Tversky and Kahneman (1974).

6 A third external force that can induce a listener to trust a speaker arises when the speaker’s statements are subject to some chance of being externally verified.

---

7 If you know a false statement is coming, then it is optimal to ignore what the speaker is saying. Therefore, you can only be deceived if you mistake a false statement for a true one.

## Appendix to Chapter 2

We now present a formal version of our theory. We do this to show how our conclusions follow from our premises. Our goal is to make the theory's logic accessible to a wide audience. At the request of the editors, we have moved some of the proofs to our web site (<http://poliscilab.ucsd.edu>). They are also available from the authors upon request.

We proceed in the following manner. Instead of presenting one all-encompassing model of attention, communication, and delegation, we present a series of smaller, more focused models. We begin by presenting our model of attention. This is followed by models of communication and delegation, respectively. Individually, these models reveal important relationships between complexity, scarcity, and choice. Collectively, they constitute our theory.

There is a hierarchical relationship between our models. Our delegation models are based on our communication models, and both are based on our attention model. This hierarchy allows us to generate relevant conclusions with minimal losses of generality.

Presenting the theory in this way reduces the number of logical steps required to derive any one of our conclusions from a well-defined set of premises. As a result, more readers should be able to follow our logic than would be the case if we presented an all-encompassing model. Readers who are interested in

examining related all-encompassing models should consult our previous publications (Lupia and McCubbins 1994a, b; Lupia and McCubbins 1995).

The appendix is partitioned into three sections. The sections correspond to our theories of attention, communication, and delegation, respectively. We begin each section by stating our assumptions about player objectives, opportunities and knowledge. We end each section with our conclusions about attention, persuasion, enlightenment, deception, reasoned choice, and delegation. In between, we trace the logical steps that link premises to conclusions. Unless otherwise stated, and there will be important exceptions, we assume that all elements of every model are common knowledge.

### ***1A Model of Attention***

#### **PREMISES**

A *principal* has an ideal point,  $p \in \mathbb{R}^d$ . Her task is to choose one of two exogenously and independently determined alternatives,  $x \in \mathbb{R}^d$  or  $y \in \mathbb{R}^d$ .

Note that the assumption “ $x \in \mathbb{R}^d$  and  $y \in \mathbb{R}^d$ ” is without a loss of generality to the assumption that  $x$  and  $y$  are points in any finite-dimensional space.

Unlike our models of communication and delegation, there is only one actors in our model of attention. Therefore, we use decision theory and not game theory to derive conclusions in about her actions.

The principal's choice of  $x$  or  $y$  depends on her objectives, knowledge and opportunities. We now describe each in turn.

The principal's objective is to maximize her expected utility. We say that the principal maximizes *ex ante* expected utility when she must make a decision before she has exhausted her pre-payoff opportunities to acquire information, and that she maximizes *ex post* expected utility when she must make after she has exhausted these opportunities (see Holmstrom and Myerson 1983.)

From her choice of  $x$  or  $y$  she derives utility  $u(x|p)$  or  $u(y|p)$ , respectively. That is, the principal prefers the alternative whose spatial location is closest to  $p$ . Note that our conclusions on trivial variants of them remain valid if the shape of the principal's utility function is in a large class of quasi-concave utility functions. Our conclusions merely require that decreases in utility are weakly monotonic with respect to increases in the spatial distance between ideal points and outcomes and with respect to cost increases (introduced below).

The principal has incomplete information about the personal consequence of her actions. That is, she may not *know* whether  $x$  or  $y$  is closer to  $p$ . She does, however, have *beliefs* about which is closer. Specifically, she knows that the true spatial location of  $x$  is determined by a single random draw from the distribution  $X$ , where  $X$  has support on a known subset of  $\mathcal{D}$ . Put another way, the principal has beliefs about the range of possible locations of  $x$  and about the likelihood that

each possible location is the true one. She does not, however, know which location was actually drawn. The principal has similar information about  $y$ . That is, she knows that  $y$  is the result of a single draw from the distribution  $Y$ , where  $Y$  has support on a subset of  $\theta_I$

The principal's opportunities are implicit in the model's sequence of events. First, the principal decides whether to acquire information (described in greater detail below). Next, the principal chooses  $x$  or  $y$ . Finally, the game ends and she receives a payoff.

With respect to her information acquisition decision, the principal has three options. She can:

- Pay nothing and learn nothing.
- Pay  $c_{xy} \geq 0$  for the opportunity to learn the location of  $x$  and  $y$  with probability  $q_{xy}$  and learn nothing with probability  $1 - q_{xy}$ . This is equivalent to the decision to purchase the most detailed information available.
- Pay  $c_x \geq 0$  for the opportunity to learn the location of  $x$  with probability  $q_x$  and learn nothing with probability  $1 - q_x$ . When the principal chooses this option, she purchases a relatively vague signal even though the option to purchase more detailed information is available.

We think of  $c_x$  and  $c_{xy}$  as cognitive opportunity costs and  $q_x$  and  $q_{xy}$  as cognitive transactions costs. That is, the  $c$  terms are the scarce energies required to acquire information and the  $1 - q$  terms are the frictions associated with processing a stimulus into a helpful form.  $c_x$ ,  $c_{xy}$ ,  $q_x$  and  $q_{xy}$  are fixed and exogenous.

For simplicity and without a loss of generality, we make the following assumptions about how the principal breaks ties: if  $x$  and  $y$  provide the principal with the same expected utility, then the principal chooses  $y$ ; if paying for information and not paying for it provide equal expected utility, then the principal does not pay; if two pieces of information provide equal expected utility at the time of purchase, then the principal pays for the less expensive piece.

Let  $CL_p(x)$  be the set of  $y \in \mathcal{I} \setminus \{x\}$  that is closer to  $p$  than  $x$  and let  $CL_p(y)$  be the set of  $x \in \mathcal{I} \setminus \{y\}$  that is closer to  $p$  than  $y$ . Then,  $\int_{CL_p(y)} |y - p| - |x - p| dX$  is the expected return from learning  $x$  for a given value of  $y$  and

$\iint_{CL_p(y)} |y - p| - |x - p| dXY$  is the expected return from learning  $x$  given beliefs  $Y$ .

Note that the set of  $x$  over which the integral is taken is the set of  $x$  that is closer to the principal's ideal point than  $y$ . Put another way, paying for information can be beneficial only if it prevents the principal from making a costly mistake (choosing  $y$  when  $x$  is closer to  $p$ ).

If  $-\int |y - p| dY \geq -\int |x - p| dX$ , then the return from paying  $c_{xy}$  is

$q_{xy} \times \left( \iint_{CL_p(y)} |y-p| |x-p| d\mathbf{X}Y \right)$ . Otherwise, it is  $q_{xy} \times \left( \iint_{CL_p(x)} |x-p| |y-p| d\mathbf{X}Y \right)$ .

Let  $CL_p\mathcal{Y}$  be the set of  $x$  for which  $-|x-p| > -\int |y-p| dY$ .  $CL_p\mathcal{X}$  is the set of  $x$  that provide higher utility than the expected value of  $y$ , given beliefs  $Y$ . If the principal knows  $y$ , then  $CL_p(y) = CL_p\mathcal{Y}$ . Let  $\neg CL_p\mathcal{Y}$  be the set of  $x$  for which  $-|x-p| \leq -\int |y-p| dY$ .

Let  $B_x = \int_{CL_p(y)} \int_{CL_p(y)} |y-p| |x-p| d\mathbf{X}Y$ ,  $L_y = \int_{CL_p(x)} \int_{CL_p(y)} |y-p| |x-p| d\mathbf{X}Y$ ,

$\int_{CL_p(x)} \int_{CL_p(y)} |y-p| |x-p| d\mathbf{X}Y$ ,  $B_y = \int_{CL_p(x)} \int_{\neg CL_p(y)} |y-p| |x-p| d\mathbf{X}Y$ ,

and

$L_x = \int_{CL_p(y)} \int_{\neg CL_p(y)} |y-p| |x-p| d\mathbf{X}Y$ . In words,  $B_x$  is the expected benefit

of observing a value of  $x$  that correctly suggests that  $x$  is closer to  $p$  than  $y$ .  $L_y$  is

the expected loss from observing a value of  $x$  that incorrectly suggests that  $x$  is

closer to  $p$  than  $y$ . The definitions of  $B_y$  and  $L_x$  parallel those of  $B_x$  and  $L_y$ ,

respectively. The terms  $L_x$  and  $L_y$  identify the circumstances under which acquiring

incomplete information - in this case learning only  $x$ , can *cause* the principal to

make a costly mistake. The existence of such circumstances is but one reason why

the principal may rationally choose not to acquire more information.

Note that if  $-\int |y-p| dY \geq -\int |x-p| dX$ , then the expected benefit to the principal

of paying  $c_x$  is  $B_x - L_y$ . Otherwise, it is  $B_y - L_x$ .

**Proposition 2** The principal uses the following rule to determine her actions.

$$\text{If } -\int |y - p| dY \geq -\int |x - p| dX,$$

• and  $[q_{xy} \cdot (\iint_{CLP(y)} |y - p| |x - p| dXY) \not\geq \max_{xy} B_{x - L_y - c_x}]$  then the principal pays  $c_{xy}$ .

• and  $B_{x - L_y - c_x} \geq [q_{xy} \cdot (\iint_{CLP(y)} |y - p| |x - p| dXY) \not\geq \max_{xy} B_{x - L_y - c_x}]$  then the principal pays  $c_x$ .

$$\text{If } -\int |x - p| dX > -\int |y - p| dY$$

• and  $[q_{xy} \cdot (\iint_{CLP(x)} |x - p| |y - p| dXY) \not\geq \max_{xy} B_{y - L_x - c_x}]$  then the principal pays  $c_{xy}$ .

• and  $B_{y - L_x - c_x} \geq [q_{xy} \cdot (\iint_{CLP(x)} |x - p| |y - p| dXY) \not\geq \max_{xy} B_{y - L_x - c_x}]$  then the principal pays  $c_x$ .

Otherwise, the principal does not pay.

If, as a result of paying  $c_{xy}$ , the principal observes both  $x$  and  $y$ , then she chooses

$y$  if and only if  $-\int |y - p| dY > -\int |x - p| dX$  If, as a result of paying  $c_x$  the principal

observes  $x$ , then she chooses  $y$  if and only if  $-\int |y - p| dY \geq -\int |x - p| dX$  If the

principal observes neither  $x$  nor  $y$ , then she chooses  $y$  if and only if  $-\int |y - p| dY$

$\geq -\int |x - p| dX$  Otherwise, she chooses  $x$ .

As the game is decision theoretic, the proof of the proposition is trivial. To falsify it, one would have to either contradict the assumption of utility maximization, change the definition of  $p$ ,  $x$ ,  $y$ ,  $XY$ ,  $c_x$ ,  $c_{xy}$ ,  $q_x$  or  $q_{xy}$ , or change the sequence of events.

### CONCLUSIONS FROM OUR ATTENTION MODEL

We will now prove the validity of the Chapter 2 theorems. To unite the mathematics of the model with the theory as presented in the text, we make the following analogies.

- If  $-\int |y - p|dY \geq -\int |x - p|dX$ , then the expected cost of a mistake is  $\iint_{CLp(y)} |y - p| - |x - p| dXY$ . Otherwise, it is  $\iint_{CLp(x)} |x - p| - |y - p| dXY$ .
- Attention to a particular stimulus is either the decision to pay  $c_{xy}$  or the decision to pay  $c_x$ .
- The probability that attending to a particular stimulus is sufficient to prevent a mistake is  $[q_{xy} \times \iint_{CLp(y)} dXY]$  or  $[q_x \times \int_{CLp(y)} \int_{CLp(Y)} dXY]$  if  $-\int |y - p|dY \geq -\int |x - p|dX$ . Otherwise, it is  $[q_{xy} \times \iint_{CLp(x)} dXY]$  or  $[q_x \times \int_{CLp(x) - CLp(Y)} dXY]$ .

- The probability that attending to a particular stimulus is sufficient to cause a

costly mistake is  $[q_x \times \int_{CL_p(x)} \int_{CL_p(Y)} d\mathbb{X}Y]$  if  $-\int |y-p| dY \geq -\int |x-p| dX$ . Otherwise, it is

$$q_x \times \int_{CL_p(y)} \int_{CL_p(Y)} d\mathbb{X}Y .$$

- Induces attention to a particular stimulus is a change from not paying  $c_x$  (or  $c_{xy}$ ) to paying for it.
- Reduces attention to a particular stimulus is a change from paying  $c_x$  (or  $c_{xy}$ ) to not paying for it.

***Proof of Theorem 1*** (*The information is neither necessary nor sufficient for reasoned choice.*)

By contradiction. Suppose that more information is necessary for reasoned choice and that  $Y$  has all of its support in  $CL_p(x)$ . Then, by supposition, the principal cannot make a reasoned choice without learning the precise location of  $y$ .

However, it is also true, by the supposition, that the principal can infer that  $y$  is closer to  $p$  than  $x$ , without learning  $y$ . Therefore, learning the exact location of  $y$  cannot be necessary for reasoned choice.

Now suppose that more information is sufficient for reasoned choice. Since the principal can choose not to learn  $x$  and  $y$ , then knowing either  $x$  or  $y$  must be

sufficient for a reasoned choice. Suppose that  $p = 3$ ,  $x = 8$ , the principal learns  $x$ , and  $Y$  is uniformly distributed. Therefore, the expected utility of  $y$  is greater than the expected value of  $x$  for the principal. Suppose that  $y = 9$ . Thus, learning  $x$  is not sufficient to make a reasoned choice.  $\square$

**Proof of Corollary 2 to Theorem 1** (An increase in either the expected cost of a mistake or the probability that attending to a particular stimulus is sufficient to prevent a mistake, either induces attention to the stimulus or has no effect.)

Consider the case  $-\int |y - p| dY \geq -\int |x - p| dX$  (the other case follows straightforwardly). Next suppose that there is an increase in the expected cost of a mistake. Then, by definition there is an increase in  $\int\int_{CLP(y)} |y - p| - |x - p| dXY$  and an increase in  $[q_{xy} \times \int\int_{CLP(y)} |y - p| - |x - p| dXY] - c_{xy}$ . From Proposition 2-1, it follows that such an increase could only induce the principal to pay for the opportunity to pay to learn  $x$  and  $y$ . Similar reasoning shows why an increase in the probability that attending to a stimulus is sufficient to prevent a mistake has the same effect.

$\square$  .

**Proof of Corollary 3 to Theorem 1** (If an increase in the probability that attending to a particular stimulus is sufficient to cause a mistake is larger than a

*consequent increase in the probability that attending to a particular stimulus is sufficient to prevent a mistake, then attention to that stimulus is either reduced or has no effect.)*

Consider the case  $-\int |y - p|dY \geq -\int |x - p|dX$  (the other case follows straightforwardly). Next suppose that there is an increase in the probability that attending to a particular stimulus is sufficient to cause a mistake that is larger than a consequent increase in the probability that attending to that stimulus is sufficient to prevent a mistake. Then, by definition there is an increase in

$$\int_{CLp(x)} \int_{CLp(Y)} dXY - \int_{CLp(y)} \int_{CLp(Y)} dXY . \text{ This increase implies a decrease in } B_x - L_y - c_x.$$

From Proposition 2-1, it follows that such an increase could only induce the principal not to pay  $c_x$ . Otherwise, the increase has no effect.  $\square$

As noted in the text, Theorem 2-2 (*People have an incentive to ignore many stimuli.*) requires two additional premises. The first premise is that there are thousands of stimuli to which a person could attend. The second premise is that opportunity cost of paying attention in the general case is substantially higher than if there are one or two stimuli competing for attention. Within the context of the model, we represent these premises with the assumption that  $c_x$  and  $c_{xy}$  are high relative to all of the model's other parameters. Theorem 2-2 and its corollary follow straightforwardly.

## Chapter Three Appendix

First, we present our basic communication model. Then, we extend it in three ways. Our basic model and each of its extensions allow us to derive the necessary and sufficient conditions for persuasion presented in the text. All mathematical terms are in italics.

### ***The Basic Model***

We model communication as a game between two players, a speaker and a principal. The principal chooses one of two alternatives, called  $x$  and  $y$ . The speaker sends a signal to the principal about her choice. We depict the extensive form in Figure 3-1.

The sequence of events begins with three probabilistic choices by nature. We denote these choices  $n \in \{n_b, n_c, n_k\}$ . The order of these choices is irrelevant. In Figure 3-1, we order the choices in the way that makes the extensive form (notably the speaker's information sets) easiest to draw. Below, we order nature's choices in the way that makes them easiest to describe.

One of nature's three choices determines the "state of the world." We denote this choice  $n_b \in \{better, worse\}$ . This choice determines whether  $x$  is *better* or *worse* than  $y$  for the principal. Nature chooses the state  $n_b = better$  with probability  $b \in [0, 1]$  and the state  $n_b = worse$  with probability  $1 - b$ . The principal

does not know  $n_b$ . The principal does, however, have prior beliefs about the “state of the world.” These beliefs are represented by the probability  $b$ .

If  $n_b = \textit{better}$  and the principal chooses  $x$ , then the principal earns utility  $U$   
 $\text{ } 0$  If  $n_b = \textit{worse}$  and if the principal chooses  $x$ , then she earns utility  $U - \text{ } 0$ . We  
 assume, without a loss of generality, that if the principal chooses  $y$ , then she earns  
 utility  $0$ . So if  $n_b = \textit{better}$ , then it is better for the principal to choose  $x$  than it is for  
 her to choose  $y$ . If  $n_b = \textit{worse}$ , then it is worse for the principal to choose  $x$  than it  
 is for her to choose  $y$ .

One of nature’s three choices determines what the speaker knows about  
 the “state of the world.” We denote this choice  $n_k \in \{n_b, \emptyset\}$ . Nature allows the  
 speaker to know  $n_b$  ( $n_k = n_b$ ) with probability  $k$  and makes no such  
 revelation (chooses  $n_k = \emptyset$ ) with probability  $1 - k$ . We assume that the speaker  
 knows  $n_k$  and that the principal does not. So, when  $k > 0$  the speaker has  
 private information about his knowledge of “the state of the world.” The  
 principal’s beliefs about the speaker’s knowledge are represented by the probability  
 $k$ .

The third of nature’s three choices determines whether the speaker and  
 principal have common or conflicting interests. We denote this choice  $n_c$   
 $\in \{\textit{common}, \textit{conflicting}\}$ . Nature chooses  $n_c = \textit{common}$  with probability  $c$  and  
 and  $n_c = \textit{conflicting}$  with probability  $1 - c$ . We assume that the speaker knows  $n_c$  and



### ***The Equilibrium Concept***

To derive precise deductive inferences about player beliefs and strategies from the model, we employ an additional premise: we assume that a refined version of sequential equilibrium is the appropriate solution concept for our model. We use this section to motivate both the sequential equilibrium concept and the refinement we use.

A sequential equilibrium has two components. The first component is a strategy profile that prescribes, for every information set, a probability distribution over available actions. We use the vector  $\pi$  to denote a typical strategy profile and the scalar  $h$  to denote a typical information set.

A sequential equilibrium's second component is a system of beliefs. A system of beliefs assigns probabilities to the decision nodes within every information set. We use the vector  $\mu$  to denote a typical system of beliefs. So,  $\mu(h)$  is a player's beliefs about which of several unobservable events -- the decision nodes within information set  $h$  -- has lead to his present observable situation -- the information set  $h$ . Formally,  $\mu$  is a function from  $d\hat{I}D$ , the set of decision nodes, to  $[0,1]$  such that for every information set  $h$ ,  $\sum_{d \in d\hat{I}h} \mu(d) = 1$ . We make the usual assumption that the game's information sets collectively partition  $D$ .

When is a strategy profile and a set of beliefs a sequential equilibrium? We now describe a sequential equilibrium's necessary and sufficient conditions. In

short, a sequential equilibrium consists of a strategy profile that is “sequentially rational” and a system of beliefs that is “consistent.”

### ***Sequential Rationality***

Kreps (1990: 427) described sequential rationality as follows: “Roughly speaking, a *sequential equilibrium* is a profile of strategies  $\pi$  and beliefs  $\mu$  such that starting from every information set  $h$ ... [each player] plays optimally from then on, given that what has transpired previously is given by  $h$  and what will transpire at subsequent nodes belonging to other players is given by  $\pi$ . This condition is called *sequential rationality*.”

A sequentially rational strategy profile is a necessary condition for a sequential equilibrium. It is the requirement that everyone makes what they believe to be utility-maximizing choices at every stage of the game. In this sense, it is closely related to the better-known Nash equilibrium concept.

Neither sequential rationality nor the Nash equilibrium concept is sufficient for our purposes. The source of the insufficiency is that neither sequential rationality nor the Nash equilibrium concept restrict player beliefs. Restrictions on beliefs are important because communication games can have many sequentially rational (or Nash equilibrium) strategy profiles that require players to have nonsensical belief systems (Kreps 1990, Chapter 12 contains several examples of such equilibria.) The sequential equilibrium concept improves on sequential

rationality and Nash equilibrium by posing the minimal requirement that strategies and beliefs be somehow “consistent” with each other.

### *Consistency*

Consistency requires that a set of beliefs  $\mu$  be based on strategy profile  $\pi$ , that is itself based on beliefs  $\mu$ , and so on. Kreps (1990: 429-430) describes consistency as follows, noting an important set of circumstances where consistency is difficult to define: “Sequential rationality is one part of the definition of a sequential equilibrium. In addition, we want strategies and beliefs to make sense together... At a minimum, we would want to insist that the strategy profile  $\pi$  and the beliefs  $\mu$  are consistent at the level of Bayes’ rule in the following sense: Given the strategy profile  $\pi$ , for any information set  $h$  that will be reached with positive probability (if players use the strategies given by  $\pi$ ) beliefs at  $h$  are computed from the strategies via Bayes’ rule... But Bayes’ rule will not apply to information sets that are not reached with positive probability in the course of play. And it is precisely at such information sets that beliefs are important. So we might want to insist on rather more consistency than just consistency with Bayes’ rule when it applies.”

Consistency, while generally easy to define, is problematic when a player finds himself at an information set that she believes should occur with zero

probability in equilibrium. Such a finding implies that beliefs and strategies are out of sync.

Within the game theory literature, there is substantial disagreement about what to assume about player beliefs at zero-probability information sets. However, if we want to avoid the nonsensical equilibrium possibilities associated with sequential rationality and Nash Equilibrium, then we must impose some restriction. The game theorists' disagreement is about which restriction to impose. One popular restriction is to assume that zero-probability information sets cannot occur. For example, the following definition requires "strictly mixed" strategy profiles that preclude zero-probability information sets (as well as some pure strategies) and allow Bayes' rule to be invoked everywhere. "[A set of beliefs]  $(\mu, \pi)$  is consistent (Kreps and Wilson [1982]) if there is a sequence of totally mixed strategy profiles  $\mathbf{p}^n \rightarrow \pi$  such that the beliefs  $\mathbf{m}^n$  computed from  $\mathbf{p}^n$  using Bayes' rule converge to  $\mu$ . A strategy profile is totally mixed if at every information set the associated behavioral strategy puts strictly positive probability on every action. Thus the beliefs associated with a totally mixed strategy profile are completely determined by Bayes' rule." (Fudenberg and Tirole 1991: 241.) This restriction, while effective in other contexts, is far too strong for our purposes. Other restrictions are contained within equilibrium refinements, such as Cho and Kreps' (1987) intuitive criterion, Banks and Sobel's (1987) concept of divinity, and

Fudenberg and Levine's (1993) steady-state equilibrium. Each refinement requires players to base beliefs at zero-probability information sets on other information available within the game.

We adopt a refinement (developed in the context of a study of both sequential and perfect Bayesian equilibria) offered by Fudenberg and Tirole (1991). We choose this refinement because it was developed using models that are structurally similar to our own. Fudenberg and Tirole's refinement requires that deviations from equilibrium that lead to zero-probability information sets not be treated as containing information about things that the deviating player does not know.

In our model, the logic of the Fudenberg and Tirole refinement reduces to the following, innocuous assumption --- "if the principal is at a zero-probability information set, then she ignores the speaker's signal." To see why this refinement is innocuous in our model, recall that the speaker can take one of two actions -- he can signal  $B$  or he can signal  $W$ . Also recall that the principal does not observe any of nature's three choices. These two facts imply that there are only two information sets at which the principal can find herself -- the information set following the principal's observation that  $sB$  and the information set following  $sW$ . If there exists an information set that is arrived at in equilibrium with probability  $0$ , then the other information set must be arrived at with probability  $1$ .

In essence, the speaker has a dominant strategy -- he does not base his signal on  $n_b$ ,  $n_c$ , or  $n_k$ . As a result, the principal cannot infer anything about  $n_b$ ,  $n_c$ , or  $n_k$  -- and, hence, the consequences of her own actions -- from the speaker's deviation. This inference is equivalent to assuming that the principal ignores signals that are off the equilibrium path. Thus, our assumption is innocuous.

### **The Definition**

We now provide a formal definition of a sequential equilibrium in our model. A sequential equilibrium is a strategy profile  $\pi$  and system of beliefs  $\mu$  that are consistent with each other in the sense above and satisfy sequential rationality at every information set.

We use the vector  $\mathbf{p}_s$  to denote the speaker's component of strategy profile  $\pi$ .  $\mathbf{p}_s$  has six scalar elements, one for each speaker information set  $h_s \in \{h_1, \dots, h_6\}$ . Note that the speaker's information sets are completely determined by nature's choice vector  $n$  and that the information set labels we use,  $h_1, \dots, h_6$  match the labels used in Figure 3-1. Each element,  $p_s(h_j)$  is the probability that the speaker signals  $s$  if he is at information set  $h_j$ . We require that these probabilities sum to 1 for each information set.

We use the vector  $\mathbf{p}_r$  to denote the principal's component of strategy profile  $\pi$ . This vector has two scalar elements, one for each principal information set  $h_r \in \{h_b, h_w\}$ . Note that the principal's information sets are completely

determined by the speaker's signal. Each element,  $p_r(x; s)$ , is the probability that the principal chooses  $x$  having heard the signal  $s \in \mathcal{S}$ .  $p_r(y; s)$  is the probability that the principal chooses  $y$  given the same signal. A signal  $s$  is "along the path of play" if there exists an information set at which  $p_s(h; s) > 0$ .

**Definition:** A pair of strategy profiles  $(p_r, p_s)$  is a *sequential equilibrium* if:

- (a) For each  $h_s, p_s(h; s)$  maximizes expected speaker utility given  $p_r(x; s)$  for all  $s \in \mathcal{S}$ .
- (b) For each  $s$  that is along the path of play,  $p_r(x; s)$  maximizes the principal's expected utility given  $\mu(\text{better})$  and  $\mu(\text{worse})$  where  $\mu$  is computed from  $p_s$  by Bayes' rule.
- (c) For any  $s$  that is not along the path of play,  $p_r(x; s)$  maximizes expected principal utility given  $\mu(\text{better})$  and  $\mu(\text{worse})$ .

### ***A Note on Equilibrium Selection***

In what follows, we identify the set of non-babbling sequential equilibria. A babbling equilibrium requires either a principal who ignores all signals or a speaker who sends only uninformative signals. In our model, a babbling equilibrium is an equilibrium in which either the speaker does not base his signal on  $n_b$  or the principal does not base her response on  $s$ .

Babbling equilibria generally exist alongside non-babbling equilibria in economic signaling models (Crawford and Sobel 1982, Farrell and Gibbons 1987).

These equilibria exist in our model and provide an important insight -- you have no incentive to persuade if you are certain to be ignored and you have no incentive to be persuaded if you are certain that communication cannot provide the useful knowledge. However, there are many circumstances in which babbling equilibria exist even though it is not clear how players would reach such an equilibrium. For example, in any case where a speaker and principal can benefit from the speaker's sending a particular signal to the principal, there exists a babbling equilibrium that leads to a worse outcome for both players than a non-babbling equilibrium. Now, if an accident of nature leads the speaker in this example to babble and the principal to ignore the speaker's signal, then the "babble-ignore" strategy profile is sustainable in this case. However, we concur with Farrell (1993; 518) who claims that cases like this the "babbling equilibrium is implausible. It requires [the speaker] to randomize extensively, saying some very unnatural things, not for his own sake but for the sake of the equilibrium."

More generally, we focus on non-babbling equilibria because we are interested in determining the conditions under which people *can* persuade each other when they attempt to communicate with each other. This focus is justified by our model of attention. Recall that people who face cognitive opportunity or transactions costs should not attend to stimuli that promise zero benefit. Now consider the plight of a person who has an opportunity to communicate and

anticipates a babbling equilibrium. When compared to not communicating, communication in a babbling equilibrium promises zero benefit to both principal and speaker. If communication entails any opportunity or transactions costs -- as we assert that it generally does -- then players who anticipate babbling equilibria should make no attempt to communicate. So while we acknowledge their theoretical existence, we do not further pursue babbling equilibria."

We also focus on non-neologistic equilibria. In our model, a neologistic equilibria *requires* the speaker and principal to agree that the signal  $B$  means "worse" and not "better" and that the signal  $W$  means "better" and not "worse." Focusing on non-neologistic equilibria is equivalent to assuming that words have focal meanings (Farrell 1993; 319). Since we allow people to lie, and context to affect a signal's persuasiveness, focusing on non-neologistic equilibria is quite unrestrictive.

### **Conclusions**

For notational simplicity, let  $p \in \{p_r, p_s\}$ ,  $p_i \in \{p_s^B, p_s^W\}$ ,  $p_{s^B} \in \{p_{s^B}^B, p_{s^B}^W\}$ ,  $p_{s^W} \in \{p_{s^W}^B, p_{s^W}^W\}$  and  $p_r \in \{p_r^B, p_r^W\}$ .

### **Proposition 3**

The only non-babbling, non-neologistic sequential equilibria in the basic model is:

$$\begin{aligned}
 & p_r \in \{0, 1\} \\
 & p_5 \neq 0 \quad \text{if } b \geq \frac{1}{2} \quad \text{and } p_5 = 0 \text{ otherwise.} \\
 & p_6 \neq 0 \quad \text{if } b \geq \frac{1}{2} \quad \text{and } p_6 = 0 \text{ otherwise.} \\
 & p_r \in \{0, 1\}
 \end{aligned}$$

This equilibrium requires

Condition A:

$$\frac{U(c_k) - U(c_l)}{U(c_k) - U(c_l)} \geq \frac{[p_s B h_{5l} + p_s B h_{6l}]}{[p_s B h_{5l} + p_s B h_{6l}]} \quad \text{and} \quad U$$

and Condition B:

$$\frac{U(c_k) - U(c_l)}{U(c_k) - U(c_l)} \geq \frac{[p_s B h_{5l} + p_s B h_{6l}]}{[p_s B h_{5l} + p_s B h_{6l}]} \quad \text{and} \quad U$$

where at least one of the inequalities is strict.

### **Proof**

We proceed as follows. First, we define the expected value of every pure strategy at every speaker information set. Second, we identify the boundaries of the set of potential non-babbling, non-neologistic sequential equilibria. Third, we identify the sequentially rational strategy profiles within this set. We find that the named equilibrium is this set's only member. Finally, we evaluate the consistency of the sequentially rational strategy profiles.

To see the expected value of every pure strategy at every speaker information set, consider the following relationships. At  $h_1$ , the expected utility from  $p_s B h_{1l}$  is  $p_r(c; B)$ . The expected utility from  $p_s B h_{1r}$  is  $p_r(c; W)$ . If  $p_r(c; B) \geq p_r(c; W)$ , then  $p_s B h_{1l}$  is the best response. At  $h_2$ , the expected utility from  $p_s W h_{2l}$  is  $p_r(c; B)$ . The expected utility from  $p_s W h_{2r}$  is  $p_r(c; W)$ . If  $p_r(c; B) \geq p_r(c; W)$ , then  $p_s B h_{2l}$  is the best response. At  $h_3$ , the expected utility from  $p_s B h_{3l}$  is  $p_r(c; B)$ . The expected utility from  $p_s B h_{3r}$  is  $p_r(c; W)$ . If

$p_r(\kappa; B) \geq p_r(\kappa; W)$ , then  $p_s(h_3) = \theta$  is the best response. At  $h_4$ , the expected utility from  $p_s(h_4) = \theta$  is  $p_r(\kappa; B)Z$ . The expected utility from  $p_s(h_2) = \theta$  is  $p_r(\kappa; W)Z$ . If  $p_r(\kappa; B) \geq p_r(\kappa; W)$  then  $p_s(h_4) = \theta$  is the best response. At  $h_5$ , the expected utility from  $p_s(h_5) = \theta$  is  $b p_r(\kappa; B)Z + (1-b)p_r(\kappa; W)Z$ . The expected utility from  $p_s(h_5) = \theta$  is  $b p_r(\kappa; W)Z + (1-b)p_r(\kappa; B)Z$ . If  $p_r(\kappa; B) \geq p_r(\kappa; W)$  and  $bZ \geq (1-b)Z$ , then  $p_s(h_5) = \theta$  is the best response. At  $h_6$ , the expected utility from  $p_s(h_6) = \theta$  is  $b p_r(\kappa; B)Z + (1-b)p_r(\kappa; W)Z$ . The expected utility from  $p_s(h_6) = \theta$  is  $b p_r(\kappa; W)Z + (1-b)p_r(\kappa; B)Z$ . If  $p_r(\kappa; B) \geq p_r(\kappa; W)$  and  $bZ \geq (1-b)Z \leq 0$ , then  $p_s(h_6) = \theta$  is the best response.

*Lemma 1* All mixed strategy sequential equilibria in the model are babbling equilibria.

*Proof of Lemma 1* A mixed strategy equilibrium requires that each player choose a strategy that makes the other player indifferent between their two pure strategies. A necessary and sufficient condition for rendering the speaker indifferent between his pure strategies at information sets  $h_1$  through  $h_4$  is to set  $p_r(\kappa; B) = p_r(\kappa; W)$ . Setting  $p_r(\kappa; B) = p_r(\kappa; W)$  is also necessary and sufficient to make the speaker indifferent between her two strategies at  $h_5$  if  $bZ \geq (1-b)Z \geq 0$  and at  $h_6$  if  $bZ \geq (1-b)Z \leq 0$ . Setting  $p_r(\kappa; B) = p_r(\kappa; W)$  implies that the principal is not conditioning her

strategy on the signal. Anticipating such behavior, the speaker can choose any strategy he likes. These speaker strategies will either make the principal indifferent between her pure strategies, in which case we have a babbling equilibrium, or they will not, in which case we do not have an equilibrium.

If  $b \geq \frac{1}{2}$  or  $b \leq \frac{1}{2}$  then any principal strategy, including  $p_r(B) = p_r(W)$  makes the speakers at  $h_5$  and  $h_6$  indifferent. Note, however, the principal has an incentive to choose a mixed strategy other than  $p_r(B) = p_r(W)$  only if she can induce the speaker at  $h_5$  and  $h_6$  to take distinct and knowledge transferring actions. Since the speaker at  $h_5$  and  $h_6$  has no useful private information at either of these information sets, by definition, the requirement cannot be met. Therefore, only equilibria that could result from such an adaptation is a babbling equilibrium. *QED.*

From similar logic it follows that all equilibria for which  $p_r(B) = p_r(W)$  are babbling equilibria. Moreover, any non-babbling equilibrium for which  $p_r(B) < p_r(W)$  and  $p_r(B) > p_r(W)$  requires neologisms (i.e., both players know that  $B$  means *worse* and  $W$  means *better*.) Therefore, non-babbling, non-neologistic sequential equilibria must include  $p_r(B) = p_r(W)$

Since non-babbling, non-neologistic sequential equilibria must include  $\pi_r \in (1,0)$ , they must also include  $p_r \neq 0,0,1$ . The reason for this is that the expected speaker utility at  $h_1$  through  $h_4$  reveal  $p_r \neq 0,0,1$  to be the unique profile of best responses when  $p_r(\kappa; B) > p_r(\kappa; W)$ . Therefore, the set of non-babbling, non-neologistic sequential equilibria must be contained within  $p \neq 0,0,1,0$

where  $0,1$  within strategy profile  $\pi$  is read as “either 0 or 1.” It remains to first identify the sequentially rational strategy profiles within this set and then evaluate these profiles’ consistency.

At  $h_B$ , the expected utility from  $p_r(\kappa; B) \neq$  is:

$$\begin{aligned}
 & E_{\kappa} [u_B(p_r(\kappa; B) \neq)] = p_s(\kappa; 1)U(\kappa, \kappa) + p_s(\kappa; 2)U(\kappa, b) + p_s(\kappa; 3)U(\kappa, \kappa) + p_s(\kappa; 4)U(\kappa, +) \\
 & + p_s(\kappa; 5)U(\kappa, \kappa) + p_s(\kappa; 5)U(\kappa, \kappa) + p_s(\kappa; 6)U(\kappa, \kappa) + p_s(\kappa; 6)U(\kappa, \kappa) \\
 & E_{\kappa} [u_B(p_r(\kappa; B) \neq)] = p_s(\kappa; 1)U(\kappa, \kappa) + p_s(\kappa; 2)U(\kappa, b) + p_s(\kappa; 3)U(\kappa, \kappa) + p_s(\kappa; 4)U(\kappa, +) \\
 & + p_s(\kappa; 5)U(\kappa, \kappa) + p_s(\kappa; 5)U(\kappa, \kappa) + p_s(\kappa; 6)U(\kappa, \kappa) + p_s(\kappa; 6)U(\kappa, \kappa)
 \end{aligned}$$

At  $h_W$ , the expected utility from  $p_r(\kappa; W) \neq$  is

$$\begin{aligned}
 & E_{\kappa} [u_W(p_r(\kappa; W) \neq)] = p_s(\kappa; 1)U(\kappa, \kappa) + p_s(\kappa; 2)U(\kappa, b) + p_s(\kappa; 3)U(\kappa, \kappa) + p_s(\kappa; 4)U(\kappa, +) \\
 & + p_s(\kappa; 5)U(\kappa, \kappa) + p_s(\kappa; 5)U(\kappa, \kappa) + p_s(\kappa; 6)U(\kappa, \kappa) + p_s(\kappa; 6)U(\kappa, \kappa) \\
 & E_{\kappa} [u_W(p_r(\kappa; W) \neq)] = p_s(\kappa; 1)U(\kappa, \kappa) + p_s(\kappa; 2)U(\kappa, b) + p_s(\kappa; 3)U(\kappa, \kappa) + p_s(\kappa; 4)U(\kappa, +) \\
 & + p_s(\kappa; 5)U(\kappa, \kappa) + p_s(\kappa; 5)U(\kappa, \kappa) + p_s(\kappa; 6)U(\kappa, \kappa) + p_s(\kappa; 6)U(\kappa, \kappa)
 \end{aligned}$$

Recall that the principal earns utility zero for choosing  $y$ . Therefore,  $p_r(\kappa; B) \neq$  is the best response only if the expected utility from  $p_r(\kappa; B) \neq$  is  $\geq 0$  and  $p_r(\kappa; W) \neq$  is a best response only if the expected utility from  $p_r(\kappa; W) \neq$  is  $\geq 0$ . Since a non-

babbling equilibrium requires that the expected utility from  $p_r(\cdot; B) \neq$  is  $\geq 0$ , that the expected utility from  $p_r(\cdot; W) \neq$  is  $\leq 0$ , and that one of these inequalities is strict, it requires that one of the inequalities in Conditions A or B be strict.

We now prove that  $p_r(0, 0, 1, 0)$  is a sequential equilibrium under the conditions that we specify in Proposition 3-1. The other cases --  $p_r(0, 0, 1, 1)$ ,  $p_r(1, 0, 1, 0)$ , and  $p_r(1, 0, 1, 1)$  -- follow equivalent logic. From the expected utility at information sets  $h_5$  and  $h_6$ , we know that this equilibrium holds only if  $b \geq 0$  and  $b \leq 0$ . This requirement matches the related requirement of Proposition 3-1. From the expected utility at information sets  $h_1$  through  $h_4$ , we know that this equilibrium requires the expected utility of  $p_r(\cdot; B) \neq \geq 0 \geq$  the expected utility of  $p_r(\cdot; W) \neq$ . We evaluate the conditions under which this inequality holds below.

If  $p_r(0, 0, 1, 0)$  then the numerator of the expected utility from  $p_r(\cdot; W) \neq$  reduces to:  $c \frac{U(b) - U(0)}{c}$ . Since the denominator of this expectation is  $\Theta$ , by assumption, it is trivial to show that this quantity is  $\leq 0$  iff  $[k - \epsilon k] \geq b$ , which is true iff Condition B is true. Similarly, if  $p_r(0, 0, 1, 0)$  then the expected utility from  $p_r(\cdot; B) \neq$  reduces to:  $c \frac{U(b) - U(0)}{c}$ . It is trivial to show that

this quantity is  $\geq 0$  iff  $b \geq \frac{p_b}{p_b + p_w}$ , which is true iff Condition A is true.

Therefore,  $\pi = (1, 0, 0, 1, 0, 0, 1, 0)$  is sequentially rational under the conditions that we specify in Proposition 3-1.

If the beliefs required to support this profile are consistent, then the profile and beliefs together constitute a sequential equilibrium. Beliefs are consistent iff

$$m_{\text{better}|\mathcal{B}} = \frac{b \cdot (\text{the probability that } s_{\mathcal{B}} \text{ is } b\text{-better})}{b \cdot (\text{the probability that } s_{\mathcal{B}} \text{ is } b\text{-better}) + (1-b) \cdot (\text{the probability that } s_{\mathcal{W}} \text{ is } b\text{-better})}$$

In the proposed equilibrium  $m_{\text{better}|\mathcal{B}} = 1$ ; the probability that  $s_{\mathcal{W}}$  is  $b$ -better is zero; and the probability that  $s_{\mathcal{B}}$  is  $b$ -better is non-zero. Therefore, beliefs are consistent. Equivalent logic proves consistency for  $m_{\text{better}|\mathcal{W}}$ ,  $m_{\text{worse}|\mathcal{B}}$ , and  $m_{\text{worse}|\mathcal{W}}$ .  $\square$

**Theorem 3**

The equilibrium in Proposition 3-1 exists only if  $c > \frac{1}{2}$

**Proof:**

For notational simplicity, let  $f = \frac{1}{2} [p_s B_h + p_w B_h]$  and let  $g =$

$$\frac{1}{2} [p_s B_h + p_w B_h] \quad \text{A necessary, but not sufficient,$$

condition for the satisfaction of Proposition 3-1 is that Conditions A and B hold. A

necessary condition for Conditions A and B to hold is that  $(k - ck) \geq k f$   $\square$

$(k - g) - c(k - g)$       Multiplying each side of the inequality by its denominator  
 and dividing everything by  $k$ , which requires  $k > 0$ , produces  $k > g$        $\frac{1}{k} > \frac{1}{k} +$   
 $\frac{1}{k} - g$ .      Dividing each side of the inequality by  $\frac{1}{k} - g$       produces the  
 requirement that  $c > \frac{1}{k} - g$        $\square$

### **Theorem 3**

The equilibrium in Proposition 3-1 exists only if  $k > 0$

#### **Proof:**

If  $k \leq 0$ , then both the expected utility from  $(p, k; B)$  and the expected utility from  
 $(p, k; W)$  equal 0. Therefore, neither of the above mentioned inequalities can be  
 strict.  $\square$

### **Corollary to Theorem 3**

The equilibrium in Proposition 3-1 does not require  $n_c < common$ .

### **Corollary to Theorem 3**

The equilibrium in Proposition 3-1 does not require  $n_k < b$ .

The proofs of these corollaries are simple -- it is  $c$  and  $k$ , not the status of  
 $n_c$  or  $n_k$ , that determine the equilibria in Proposition 3-1.

### **Extension 1 Verification**

#### **Premises**

The verification extension differs from the basic model in just one way.

Now, nature makes a fourth choice  $(n_v, \hat{I}, h_b, s)$ . As the following time line shows,

nature makes this choice after the speaker sends his signal and before the principal makes her choice.

Action: Nature makes first three choices.    The speaker sends a signal.    Nature verifies with probability  $v$ . (New)    The principal chooses.

\*-----\*

Sequence: First, Second, Third    Fourth    Fifth    Last

Nature replaces the speaker's signal with  $n_b$  (the true "state of the world") with probability  $\theta < 1$ . With probability  $1 - \theta$ , no replacement occurs. Neither player knows  $n_v$  at the time they make their choice. So, if  $\theta < 1$ , then the speaker does not know whether or not nature will "verify" his signal and the principal does not know whether the signal she has received is the speaker's statement or nature's verification of the true "state of the world." Note that the case where the principal knows  $n_v$  is a trivial variant of the basic model, that the case where  $v = 1$  is the basic model, and that the case where  $v = 0$  is trivial.

The current definition of a sequential equilibrium differs from the basic model's equilibrium only in that we now replace the strategy  $p_s(\xi; h_s)$  with the strategy  $p_s(\xi; h_s, v)$ ; the strategy  $p_r(\xi; s)$  with the strategy  $p_r(\xi; s, v)$ ; and the beliefs  $m(better; \xi)$  and  $m(worse; \xi)$  with the beliefs  $m(better; \xi, v)$  and  $m(worse; \xi, v)$  respectively.

**Conclusions**

Proposition 3-2 describes the set of non-babbling, non-neologistic equilibria for the basic communication model with verification. The main difference between this set of equilibria and the equilibria of the basic model is that  $c > 5$  is no longer a requirement for equilibria. Put another way,  $v$  is a substitute for  $c$ .

**Proposition 3**

The only non-babbling, non-neologistic sequential equilibria in the basic model with verification are (with differences from Proposition 3-1 in **bold**):

$$\begin{aligned}
 p_1 & \neq 0 \\
 p_5 & \neq 0 \text{ if } b \geq 0 \text{ and } = 0 \text{ otherwise.} \\
 p_6 & \neq 0 \text{ if } b \geq 0 \text{ and } = 0 \text{ otherwise.} \\
 p_r & \neq 0
 \end{aligned}$$

These equilibria require:

Condition A':

$$\frac{p_5 \cdot (k - 1) \cdot [p_5 B h_{5\phi} + p_6 B h_{6\phi}]}{p_6 \cdot (k - 1) \cdot [p_5 B h_{5\phi} + p_6 B h_{6\phi}]} \geq \frac{b}{c} \quad \text{and} \quad U$$

Condition B':

$$\frac{p_5 \cdot (k - 1) \cdot [p_5 B h_{5\phi} + p_6 B h_{6\phi}]}{p_6 \cdot (k - 1) \cdot [p_5 B h_{5\phi} + p_6 B h_{6\phi}]} \geq \frac{b}{c} \quad \text{and} \quad U$$

where at least one of the inequalities is strict.

**Proof**

That the set of non-babbling, non-neologistic sequential equilibria is contained

within  $p \neq 0, 1, 0, 1, 0$  follows from the same logic as in the proof of Proposition 3-1.

We proceed as follows. First, we define the expected value of every pure strategy at every information set. Second, we identify the sequentially rational strategy profiles within  $\mathcal{P}$ . Finally, we evaluate the consistency of the sequentially rational strategy profiles.

To see the expected value of every pure strategy at every information set, consider the following relationships. At  $h_1$ , the expected utility from  $p_s^1$  is  $v(p_r^1; B)$ . The expected utility from  $p_s^1$  is  $v(p_r^1; B)$ . If  $v(p_r^1; B) \geq v(p_r^1; W)$  then  $p_s^1$  is the best response. At  $h_2$ , the expected utility from  $p_s^2$  is  $v(p_r^2; W)$ . The expected utility from  $p_s^2$  is  $v(p_r^2; W)$ . If  $v(p_r^2; B) \geq v(p_r^2; W)$  then  $p_s^2$  is the best response. At  $h_3$ , the expected utility from  $p_s^3$  is  $v(p_r^3; B)$ . The expected utility from  $p_s^3$  is  $v(p_r^3; B)$ . If  $v(p_r^3; B) \geq v(p_r^3; W)$  then  $p_s^3$  is the best response. At  $h_4$ , the expected utility from  $p_s^4$  is  $v(p_r^4; W)$ . The expected utility from  $p_s^4$  is  $v(p_r^4; W)$ . If  $v(p_r^4; B) \geq v(p_r^4; W)$  then  $p_s^4$  is the best response. At  $h_5$ , the expected utility from  $p_s^5$  is  $v(p_r^5; B)$ . The expected utility from  $p_s^5$  is  $v(p_r^5; B)$ . If  $v(p_r^5; B) \geq v(p_r^5; W)$  then  $p_s^5$  is the best response.

response. At  $h_6$ , the expected utility from  $p_s(B) = 0$  is  $v p_r(x; B) = 0$

$p_r(x; W) = 0$ . The expected utility from  $p_s(W) = 0$  is  $v p_r(x; B) = 0$ . If  $p_r(x; B) \geq p_r(x; W)$  and  $b \geq 0$ , then  $p_s(B) = 0$  is the best response.

At  $h_B$ , the numerator of the expected utility from  $p_r(x; B) = 0$  is  $v U$

$p_s(B) = 0$ ,  $p_s(B) = 1$ ,  $p_s(B) = 0$ ,  $p_s(B) = 1$ ,  $p_s(B) = 0$ . The denominator is the same equation without the  $U$  and  $v$ 's and is  $\theta$  by definition..

Note that the first component of this equation,  $vU$  is the probability that the signal came from nature, times  $U$ , the utility that the principal earns from choosing  $x$  when the state of the world is *better*. The remaining component reflects the principal's expectations about when the speaker would signal  $B$  -- it is  $\theta$  times the expected value at  $h_B$  in the basic model. Since the expected utility from choosing  $y$  is  $0$ ,  $p_r(x; B) = 0$  is a best response only if the expected utility from  $p_r(x; B) = 0$  is  $\geq 0$ . The expected utility at  $h_W$  has an equivalent structure.

At  $h_W$ , the numerator of the expected utility from  $p_r(x; W) = 0$  is  $v U$

$p_s(B) = 0$ ,  $p_s(B) = 1$ ,  $p_s(B) = 0$ ,  $p_s(B) = 1$

$$\begin{aligned}
 & p_r(h_4, v) U < p_r(h_5, v) U & p_r(h_5, v) U > p_r(h_6, v) U \\
 & p_r(h_6, v) U > p_r(h_5, v) U & p_r(h_6, v) U > p_r(h_5, v) U
 \end{aligned}$$

The denominator is the same equation without the  $U$  and  $U$ 's and is  $\theta$  by definition.

Therefore,  $p_r(h_4; W)$  is a best response only if the expected utility from  $p_r(h_4; W)$  is  $\geq 0$ . Since a non-babbling equilibrium requires that the expected utility from  $p_r(h_4; B)$  is  $\geq 0$ , that the expected utility from  $p_r(h_4; W)$  is  $\geq 0$ , and that one of these inequalities is strict, it requires that one of the inequalities in Conditions A' or B' be strict.

We now prove that  $p = (0, 0, 1, 0, 0, 1, 0)$  is a sequential equilibrium under the conditions that we specify in Proposition 3-2. The other cases --  $p = (0, 0, 1, 0, 1, 0, 0)$ ,  $p = (0, 0, 1, 0, 1, 1, 0)$  and  $p = (0, 0, 1, 1, 1, 1, 0)$  -- follow equivalent logic. From the expected utility at information sets  $h_5$  and  $h_6$ , we know that this equilibrium holds only if  $bZ > 0$  and  $bZ > 0$ . This requirement matches the related requirement of Proposition 3-2. From the expected utility at information sets  $h_1$  through  $h_4$ , we know that this equilibrium requires the expected utility of  $p_r(h_4; B) \geq 0 \geq$  the expected utility of  $p_r(h_4; W)$ . We evaluate the conditions under which this inequality holds below.

If  $p_s = (0, 0, 1, 0, 0)$  then the numerator of expected utility from  $p_r$  ( $k; W$ ) reduces to:  $v(U(k)) - U(kb)U(kb)$  This quantity is  $\geq 0$  iff  $U(k) \geq U(kb)$  which is true iff Condition B' is true.

If  $p_s = (0, 0, 0, 1, 0)$  then the numerator of the expected utility from  $p_r$  ( $k; B$ ) similarly reduces to:  $v(U(k)) - U(kb)U(kb)$  This quantity is  $\geq 0$  iff  $U(k) \geq U(kb)$  which is true iff Condition A' is true. Therefore,  $p = (0, 0, 1, 0, 0, 1, 0)$  is sequentially rational under the conditions that we specify in Proposition 3-2.

If the beliefs required to support this strategy profile are consistent, then the profile and beliefs together constitute a sequential equilibrium. Beliefs are consistent iff:

$\frac{p(better|B)}{p(better|B) + p(worse|B)}$  = the probability that  $s(B|n_{b=better})$  = the probability that  $s(B|n_{b=better})$  In the proposed equilibrium,  $\frac{p(better|B)}{p(better|B) + p(worse|B)}$  = 1, the probability that  $s(W|n_{b=better})$  is zero, and the probability that  $s(B|n_{b=better})$  is non-zero. Therefore, beliefs are consistent. Equivalent logic proves consistency for  $\frac{p(better|W)}{p(better|W) + p(worse|W)}$  and  $\frac{p(worse|W)}{p(better|W) + p(worse|W)}$

**Corollary to Proposition 3**

The equilibria in Proposition 3-2 do not require  $c > 5$

The proof of Corollary 2-1 is trivial (i.e., set  $v \neq 0$ ).

**Corollary 2 Proposition 3**

If  $Z \neq 0$ , then, in equilibrium, increases in  $v$  makes participation in the game less valuable for speakers at information sets where there is an incentive to deceive. It does not have the same effect on other speakers.

**Proof**

In equilibrium, the speaker has an incentive to send a truthful signal at  $h_1$  and  $h_2$ , and incentive to send an untruthful signal at  $h_3$  and  $h_4$ . At  $h_5$  and  $h_6$ , which incentive he has depends on conditions outlined in Proposition 3-2. We focus on the relationship between  $v$  and the value of the game at  $h_1$  and  $h_3$  and note that the other cases follow similar logic.

In equilibrium,  $p_s(h_1) = v \neq 0$ , for all values of  $v$ . The expected utility from  $p_s(h_1) = v \neq 0$  is  $p_s(x; B)Z$ . The derivative of this expectation with respect to  $v$  is 0.

Therefore,  $v$  has no effect on the value of the game for the speaker at information set  $h_1$ .

In equilibrium,  $p_s(h_3) = v \neq 0$ , for all values of  $v$ . The expected utility from  $p_s(h_3) = v \neq 0$  is  $v p_s(x; B)Z - p_r(x; W)Z$ . The derivative of this expectation with respect to  $v$  is  $p_s(x; B)Z - p_r(x; W)Z$ . In equilibrium this value equals  $Z$ , which is,

by definition, a negative value. Therefore, any increase in  $v$  decreases value of the game to the speaker at information set  $h_3$ .  $\square$

### **Extension 2: Penalties for Lying**

#### **Premises**

Our penalties for lying extension differs from the basic model in only one way. Now, if the the speaker sends a false signal, then he must pay penalty  $pen \geq 0$ . This extension directly effects the speaker's utility. If  $n_c = common$  and the speaker lies, then the speaker receives utility  $Z-pen$  when the principal receives utility  $U \geq 0$  and receives utility  $Zpen \leq 0$  when the principal receives utility  $U \leq 0$ . If  $n_c = conflicting$  and the speaker lies, then the speaker receives utility  $Z-pen \leq 0$  when the principal receives utility  $U \geq 0$  and receives utility  $Zpen$  when the principal receives utility  $U \leq 0$ . Note that if  $pen > Z$ , then  $Z-pen < 0$ . If the speaker tells the truth, then the speaker's utility is the same as in the basic model.

A sequential equilibrium in this extension is equivalent to the sequential equilibrium in the basic model. It is a strategy profile  $\pi$  and system of beliefs  $\mu$  that are consistent with each other in the sense above and that satisfy sequential rationality at every information set. The only differences are that we replace the strategy  $p_s(\cdot; h_s)$  with the strategy  $p_s(\cdot; h_s, pen)$  the strategy  $p_r(\cdot; s)$  with the strategy  $p_r(\cdot; s, pen)$ ; and the beliefs  $m(better)$  and  $m(worse)$  with the beliefs  $m(better, pen)$  and  $m(worse, pen)$  respectively.

### Conclusions

Like the previous two cases, there exist non-babbling, non-neologistic sequential equilibria that contain  $p_1 \neq 0$  and  $p_2 \neq 0$ . These are equilibria where knowledgeable speakers with common interests tell the truth while the knowledgeable speakers with conflicting interests lie. However, these equilibria are now possible only if the penalty is less than the smaller of  $U$  and  $|U|$  (i.e., all possible lies are “worthwhile.”)

In addition, penalties for lying induce a second set of non-babbling, non-neologistic sequential equilibria. In these equilibria, the penalty *induces* a speaker to send a truthful signal at information set  $h_3$  and/or  $h_4$ .

There are two types of equilibria. In Type I equilibria, penalties for lying can affect persuasion. In Type II equilibria, penalties for lying affect speaker strategies, yet extreme circumstances allow for equilibria where the principal ignores the speaker nevertheless. For notational simplicity, let  $p_1 \in \{p_s(h_1), p_s(h_2)\}$  let  $p_3 \in \{p_s(h_3), p_s(h_4)\}$  and let  $p_4 \in \{p_s(h_3), p_s(h_4)\}$ .

### Proposition 3

The only *type I* non-babbling, non-neologistic sequential equilibria in the basic model with penalties for lying are (with differences from Proposition 3-1 in **bold**):

$p_1 \neq 0$   
 $p_3 \neq 0$  if  $Z \geq \frac{3}{2} pen$  and otherwise.  
 $p_4 \neq 0$  if  $Z \geq \frac{3}{2} pen$  and  $=0$  otherwise.  
 $p_5 \neq 0$  if  $bZ \geq \frac{3}{2} pen$  and  $p_5 = 0$  otherwise.  
 $p_6 \neq 0$  if  $bZ \geq \frac{3}{2} pen$  and  $p_6 = 0$  otherwise.

$p_r \neq 0$

These equilibria require:

$$\frac{[k - p_s B_4) + k] \cdot [p_s B_5) + p_s B_6)]}{[k - p_s B_3) + k] \cdot [p_s B_5) + p_s B_6)]} \geq b \quad \underline{U} \quad \text{and}$$

$$\frac{[k - p_s B_4) + k] \cdot [p_s B_5) + p_s B_6)]}{[k - p_s B_3) + k] \cdot [p_s B_5) + p_s B_6)]} \leq b \quad \underline{U}$$

where at least one of the inequalities is strict.

**Proof**

We proceed as follows. First, we define the expected value of every pure strategy at every information set. Second, we identify the boundaries of the set of Type I non-babbling, non-neologistic equilibria. Third, we identify the sequentially rational strategy profiles within this set. Finally, we examine whether the sequentially rational strategy profiles are consistent.

To see the expected value of every pure strategy at every information set, consider the following relationships. Notice first that the expected utilities of principal strategies at information sets  $h_B$  and  $h_W$  in this extension are identical to those in the basic model. At  $h_1$ , the expected utility from  $p_s B_1) \neq$  is  $p_r(k; B) \geq$  The expected utility from  $p_s B_1) \neq$  is  $p_r(k; W) \geq$ pen. If  $p_r(k; B) \geq p_r(k; W)$ , then  $p_s B_1) \neq$  is the best response. At  $h_2$ , the expected utility from  $p_s W_2) \neq$  is  $p_r(k; B) \geq$ pen. The expected utility from  $p_s W_2) \neq$  is  $p_r(k; W) \geq$ . If  $p_r(k; B) \geq p_r(k; W)$ , then  $p_s B_2) \neq$  is the best response. At  $h_3$ , the expected utility from

$p_r(\kappa; B) \geq p_r(\kappa; W)$  and  $Z \geq \text{pen}$ , then  $p_s(h_3)$  is the best response.

At  $h_4$ , the expected utility from  $p_s(h_4)$  is  $p_r(\kappa; B) \geq \text{pen}$ . The expected utility from  $p_s(h_2)$  is  $p_r(\kappa; W) \geq Z$ . If  $p_r(\kappa; B) \geq p_r(\kappa; W)$  and  $Z \geq \text{pen}$ , then

$p_s(h_4)$  is the best response. At  $h_5$ , the expected utility from  $p_s(h_5)$  is  $b p_r(\kappa; B) \geq \text{pen}$ . The expected utility from  $p_s(h_5)$  is  $(b p_r(\kappa; W) \geq \text{pen})$ .

If  $p_r(\kappa; B) \geq p_r(\kappa; W)$  and  $b Z \geq \text{pen}$ , then  $p_s(h_5)$  is the best response. At  $h_6$ , the expected utility from

$p_s(h_6)$  is  $b p_r(\kappa; B) \geq \text{pen}$ . The expected utility from  $p_s(h_6)$  is  $b(p_r(\kappa; W) \geq \text{pen})$ .

If  $p_r(\kappa; B) \geq p_r(\kappa; W)$  and  $b Z \geq \text{pen}$ , then  $p_s(h_6)$  is the best response.

As was true in the previous cases, all mixed strategy equilibria are babbling equilibria. To see why this is true, recall that mixed strategy equilibria require each player to choose a strategy that makes the other player indifferent between their two pure strategies. Because of the presence and placement of the penalty for lying, there is no principal strategy profile capable of accomplishing this when  $\text{pen} > 0$ . At  $\text{pen} = 0$ , this extension of the model is the basic model and Lemma 1 applies.

From similar logic it follows that all equilibria for which  $p_r(k; B) = p_r(k; W)$  are babbling equilibria. Moreover, any non-babbling equilibrium for which  $p_r(k; B) > p_r(k; W)$  and  $p_r(k; W) > p_r(k; B)$  requires neologisms (i.e., both players know that  $B$  means *worse* and  $W$  means *better*.) Therefore, non-babbling, non-neologistic Type I sequential equilibria require  $p_r(k; B) > p_r(k; W)$ .

Since non-babbling, non-neologistic sequential equilibria must include  $p_r(k; B) > p_r(k; W)$  they must also include  $p_r(k; W) > p_r(k; B)$  because the expected values at  $h_1$  and  $h_2$  reveal this vector to be the unique profile of best responses when  $p_r(k; B) > p_r(k; W)$ . Therefore, set of non-babbling, non-neologistic sequential equilibria must be contained within  $\pi \in (1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0)$ . It remains to first identify the sequentially rational strategy profiles and then evaluate these profiles' consistency. We will prove that  $\pi \in (1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0)$  is a sequential equilibrium under the conditions that we specify in Proposition 3-3. The other cases follow equivalent logic.

From the expected utility at information sets  $h_3$  through  $h_6$ , it is trivial to verify that this equilibrium holds iff the conditions specified in Proposition 3-3 are true. If  $p_r(k; B) > p_r(k; W)$  then the numerator of the expected utility from  $p_r(k; W) > p_r(k; B)$  reduces to:

$$U(k; B) - (1-k)U(k; W) = U(k; B) - (1-k)U(k; W)$$

$c \in \mathbb{R}$ )  $\underline{U}$ . Since the denominator is  $\theta$ , by assumption, this quantity is  $\geq 0$  iff  
 $\frac{c - \epsilon_k}{\theta} \geq \frac{b}{\theta} \underline{U}$ , which is true iff the final condition of  
 Proposition 3-3 is true. Similarly, if  $p_s \in (0, 1)$  then the expected utility  
 from  $p_r \in (0, 1)$  reduces to:  $\frac{c - \epsilon_k}{\theta} \geq \frac{b}{\theta} \underline{U}$ . This quantity is  $\geq 0$  iff  $b \leq$   
 $\frac{c - \epsilon_k}{\theta} \underline{U}$ , which is true iff the final condition of Proposition 3-3 is true.  
 Therefore,  $\pi = (1, 0, 0, 1, 0, 0, 1, 0)$  is sequentially rational under the conditions  
 that we specify in Proposition 3-3.

If the beliefs required to support this equilibrium are consistent, then this profile is  
 also a sequential equilibrium. Beliefs are consistent iff

$$\mu(\text{better} | B) = \frac{b \cdot \text{the probability that } s = \text{better}}{b \cdot \text{the probability that } s = \text{better} + \theta \cdot \text{the probability that } s = \text{worse}}$$

In the proposed equilibrium,  $\mu(\text{better} | B) = 1$ , the probability that  $s = \text{better}$   
 is zero, and the probability that  $s = \text{better}$  is non-zero. Therefore, beliefs are  
 consistent. Equivalent logic proves consistency for  $\mu(\text{better} | W)$ ,  $\mu(\text{worse} | B)$ , and  
 $\mu(\text{worse} | W)$ .  $\square$

Unlike the equilibria in the basic model, but the equilibrium of the  
 verification extension,  $c > \bar{c}$  is not a requirement for non-babbling equilibria. As a

result, even a penalty for lying that keeps speakers at some information sets from lying is sufficient to induce persuasion in a range of cases where it is not otherwise possible.

**Corollary to Proposition 3**

The equilibria in Proposition 3-3 do not require  $c > 5$

**Proof**

By example (there are many). Recall the final condition of Proposition 3-3

$$\frac{[k - p_3 B_4) + k] \cdot [p_s B_5 + p_s B_6]}{[k - p_3 B_3) + k] \cdot [p_s B_5 + p_s B_6]} \leq b \text{ and } \underline{U}$$

$$\frac{[k - p_3 B_4) + k] \cdot [p_s B_5 + p_s B_6]}{[k - p_3 B_3) + k] \cdot [p_s B_5 + p_s B_6]} \geq b \text{ and } \underline{U}$$

Suppose  $Z_{pen} = 1$  and  $Z_{pen} = 1$ . If  $p_4, B = p_4, W$ , then the speaker's best

response at information sets  $h_1$  through  $h_4$  is  $p_4 \notin \emptyset$ . For notational

simplicity, let  $f = [p_s B_5 + p_s B_6]$  and let  $g = [p_s B_5 + p_s B_6]$

$[p_s B_5 + p_s B_6]$ . Then the preceding inequalities reduce to:  $f < g +$

$f \leq b \text{ and } \underline{U}$  and  $k \geq g \geq b \text{ and } \underline{U}$ . **Suppose that**  $c > k$  and  $b \leq \underline{U}$

$\geq 1$ . Then,  $f < g$ . Suppose further that  $f < g$ . Then the preceding inequalities

reduce to:  $0 \leq 1$  and  $k < g \geq 1$ . Since  $k \in \emptyset$ , Proposition 3-3's conditions

are satisfied.  $\square$

**Corollary to Proposition 3**

The only  $p \in I$  non-babbling, non-neologistic sequential equilibria are:

- $p \in \{0, 1, 1, 0, 0\}$  iff  $pen > b \text{ and } \underline{U} \leq k$  and  $b \geq 2$
- $p \in \{0, 1, 0, 0, 1\}$  iff  $pen > b \text{ and } \underline{U} \geq k$  and  $b \leq 2$  and,

- The following equilibria require  $b \geq \frac{1}{2}$ 
  - $p \in (0, 1/2)$  if  $p \in (0, 1/2)$   $\underline{U} = \min\{c, k\}$
  - $p \in (1/2, 1)$  if  $p \in (1/2, 1)$   $\underline{U} = \max\{c, k\}$
  - $p \in (0, 1/2)$  if  $p \in (0, 1/2)$   $\underline{U} = \min\{k, c\}$
  - $p \in (1/2, 1)$  if  $p \in (1/2, 1)$   $\underline{U} = \max\{k, c\}$

We omit the proof of the Type II equilibria as their logic mirrors that of the Type I equilibria.

### **Extension 3 Costly Effort with Continuous Types**

We introduce costly effort to represent cases where the speaker signals through actions instead of words. This extension features continuous types, that is, nature determines player interests and knowledge by selecting one element from a continuum of elements. Previously, nature made each of her choices by choosing one element from a set of two (e.g., previously,  $n_b \in \{\text{better}, \text{worse}\}$ ). We extend the model in this way to illuminate an important consequence of costly effort that is not easily replicated in the non-continuous format. Note that our previous theorems hold for the continuous as well as the non-continuous case (see, e.g., Lupia 1993, 1996 and Lupia and McCubbins 1994.)

#### **Premises**

As before, there are two players, the principal and the speaker. The principal has ideal point  $p \in [0, 1]$ . The speaker has ideal point  $s \in [0, 1]$ . The principal chooses one of two alternatives, now defined as  $x \in [0, 1]$  and  $y \in [0, 1]$ .

*J* The speaker takes an action that may provide the principal with information about her choice.

The sequence of events begins with two probabilistic moves by nature. The order of these moves is irrelevant. One of nature's choices determines the "state of the world." We denote this choice  $n_b = x \hat{\mathbf{I}} \mathbf{X} \subseteq \mathcal{O}$ . This choice not only determines whether  $x$  is "better" or "worse" for the principal, but also how much better or worse  $x$  is. For parsimony in discovering the consequences of costly effort, we examine the case where the speaker knows  $n_b$  and the principal does not (i.e., in the notation of the previous models,  $k \neq$  and  $n_k \neq b$ ). By contrast, the principal may not know, but does have beliefs about, which alternative is better for her.

Specifically, it is common knowledge that the true spatial location of  $x$  is determined by a single random draw from the distribution  $\mathbf{X}$ , where  $\mathbf{X}$  has support on a known subset of  $\mathcal{O}$ . Therefore, the principal has beliefs about the range of possible locations of  $x$  and the likelihood that each possible location is the true one. She does not, however, know which location was actually drawn. Thus, she may not *know* whether  $x$  or  $y$  is closer to  $p$ .

Nature's other choice,  $n_c$ , determines the relationship between the players' interests ( $n_c \rightarrow p \hat{\mathbf{I}} \mathbf{B} \subseteq \mathcal{O}$ ). As before, the speaker knows  $n_c$  and the principal does not. In other words, the speaker knows her interests while the principal

merely has beliefs about the speaker's interests. Specifically, the principal knows that the true spatial location of  $s$  is determined by a single random draw from the distribution  $\mathcal{B}$ , where  $\mathcal{B}$  has support on a known subset of  $\mathcal{D}$ .

Third, the speaker makes a decision,  $s \in \{cost, \emptyset\}$ . He must choose whether or not to pay  $cost \geq 0$ . If he pays, then the principal gets to choose  $x$  or  $y$ . If he does not pay, then  $y$  is the outcome.

The speaker's sole objective is to maximize his utility. If he pays and the principal chooses  $x$ , then the speaker earns  $x - sp - cost$ ; if he pays and the principal chooses  $y$ , then the speaker earns  $y - sp - cost$ ; and if he does not pay, then he earns  $y - sp$ . To break ties, we make the following innocuous assumption: if paying and not paying provide the speaker with the same expected utility given  $\mathbf{p}$ , then the principal does not pay.

If the speaker pays  $cost$ , then the principal makes the game's final move by choosing  $x$  or  $y$ . From her choice she earns utility  $x - p$  or  $y - p$  respectively. To break ties, we make the following innocuous assumption: if  $x$  and  $y$  provide the principal with the same expected utility, given  $\mathbf{p}_s$  and  $m(cost)$  then the principal chooses  $y$ .

A sequential equilibrium in this extension is equivalent to the basic model's sequential equilibrium. The differences are: there are now many more information sets; each speaker information set contains only one decision node;  $s \in \mathcal{B}\mathcal{W}$  is

now  $s\hat{I}(cost, \emptyset)$  and we replace the beliefs  $m(\text{better})$  and  $m(\text{worse})$  with the belief  $m(cost)$

### Conclusions

This model has only one non-babbling sequential equilibrium. Neologisms have no meaning in this model. The key insight of the non-babbling equilibrium is as follows. The speaker's objective is to induce the principal to choose the alternative that is closest to  $sp$ . However, it is profitable for him to do so only if his action will actually change what the principal does and  $y - sp < -sp/cost$ .

That is, if  $x \hat{I}(y - cost, y - cost)$  then it is not worthwhile for the speaker to influence the principal's choice. Therefore, if the speaker pays cost, then the principal can infer that  $x \hat{I}(y - cost, y - cost)$ . If this inference is different than the principal's priors, then persuasion can occur.

### Proposition 3

The only non-babbling sequential equilibrium for this extension is:

$s \leq cost$  and the principal chooses  $x$  if and only if:

$$-\int_{\min(0, y - cost)}^{y - cost} (x - p/d) X - \int_{\min(y + cost, 1)}^1 (x - p/d) X > -y - p/ \quad \text{and} \\ -x - sp/cost > y - sp/$$

Otherwise, the speaker does not pay and the outcome is  $y$ .

### Proof

For a given value of  $x$ , the expected utility to the speaker of paying  $cost$  is -

$$p_r(x; cost) - sp/ - (p_r(x; cost) - sp/cost). \quad \text{For any value of } x, \text{ the expected}$$

utility to the speaker of not paying  $cost$  is  $-y - sp/d$ . Note that the expected utility of paying  $cost$  minus the expected utility of not paying it reduces to  $p_r(x; cost) - y - sp/d - (sp/d - cost)$ . The expected utility of paying  $cost$  is higher than the expected utility from not paying it only if this quantity is positive.

Now suppose that  $x \hat{I}(y - cost, y + cost)$ . For example, suppose that  $x = y - cost + \epsilon$ , where  $cost > \epsilon > 0$  and  $\epsilon$  small -- the case where  $x = y - cost - \epsilon$  is equivalent. Then the expected utility of paying  $cost$  minus the expected utility of not paying it is  $p_r(x; cost) - y - sp/d - cost + \epsilon - (sp/d - cost)$ . Since  $cost > \epsilon > 0$ , we have  $p_r(x; cost) - y - sp/d - cost - \epsilon - (sp/d - cost)$ , which is greater than zero only if  $(p_r(x; cost) - p_r(x; cost) - cost) / \epsilon$ . Since  $p_r(x; cost) \in [0, 1]$ ,  $-(p_r(x; cost) - p_r(x; cost)) / cost > \epsilon > 0$ ,  $\min cost/\epsilon \geq 1$ . Therefore, if  $x \hat{I}(y - cost, y + cost)$  then the expected utility of paying  $cost$  minus the expected utility of not paying it must be less than zero. Thus, if  $x \hat{I}(y - cost, y + cost)$  then  $meost$  is a dominated strategy for the speaker.

In addition to  $x \hat{I}(y - cost, y + cost)$  a second necessary condition for  $meost$  is  $\pi_r(x; cost) > 0$ . This condition is satisfied if  $\xi_0 \leq x - p/d$ ,  $X > -y - p/d$ . From the fact that  $x \hat{I}(y - cost, y + cost)$  makes  $meost$  a dominated strategy for

the speaker, the principal can infer from  $\text{specost}$  that  $x \in (y - \text{cost}, y + \text{cost})$ . Thus,

$$\pi_i(x; \text{cost}) > 0 \quad \text{only if} \quad - \int_{\min(0, y - \text{cost})}^{y - \text{cost}} k - p/d \cdot X - \int_{\min(y + \text{cost}, 1)}^1 k - p/d \cdot X > -y - p/d$$

$$\text{If} \quad - \int_{\min(0, y - \text{cost})}^{y - \text{cost}} k - p/d \cdot X - \int_{\min(y + \text{cost}, 1)}^1 k - p/d \cdot X > -y - p/d \quad \text{then}$$

$\pi_i(x; \text{cost}) > 0$ , is a dominant strategy for the principal. Otherwise, and given the tie-

breaking rule,  $\pi_i(x; \text{cost}) = 0$  is a dominant strategy. If  $\pi_i(x; \text{cost}) = 0$ , then the

speaker's best response is not to pay  $\text{cost}$ . If  $\pi_i(x; \text{cost}) > 0$  and  $x \in (y - \text{cost}, y +$

$\text{cost})$  then the speaker's best response is to pay  $\text{cost}$ .  $\square$

### ***Corollary to Proposition 3***

Persuasion in Proposition 3-4 does not require that the probability of common interests between speaker and principal is  $> 0$

### ***Proof***

By example (there are many). Suppose that  $y = 0.5$ , that  $S$  and  $X$  are independent distributions with 99 percent of their respective masses at .2 and 1 percent of their respective masses at .9, that  $\text{cost} = 0.5$  and that  $sp = 0.8$ . If  $x = 0.2$  or  $sp = 0.2$ , then paying  $\text{cost}$  is a dominated strategy for the speaker. By contrast, if  $x = 0.9$  and  $sp = 0.9$ , then the speaker will pay cost if he believes that the principal will choose  $x$ . If the principal believes that the speaker will pay cost only if  $x = 0.9$ , then

her best response is to choose  $x$  when  $sp \geq st$ . Thus, persuasion occurs even though the likelihood of common interests was  $.0$  (the players have common interests only if  $sp \geq st$ ).  $\square$

### **Theorem 3**

The following conditions are individually necessary and collectively sufficient for persuasion: the principal must perceive the speaker to be trustworthy; the principal must perceive the speaker to be knowledgeable; and the signal, if true, must imply that the principal's prior beliefs are insufficient for reasoned choice.

In the absence of all external forces, persuasion requires perceived common interests ( $c \geq .5$ ) and perceived speaker knowledge ( $k \geq .5$ ). In the presence of external forces, these requirements can be reduced. As the likelihood of verification, the magnitude of the penalty for lying, or the magnitude of costly effort increases, the extent to which common interests is required decreases ( $c$  can be less than  $.5$ ). In other words, with respect to persuasion, the external forces can be substitutes for common interests.

### **Proof:**

The individual necessity and collective sufficiency of the three conditions is proven in each of Propositions 3-1 through 3-4. The statement about persuasion in the absence of external forces is proven as Theorems 1 and 2. The statement about

persuasion in the presence of external forces is proven as Corollary 1 to Proposition 3-2, Corollary 1 to Proposition 3-3, and Corollary 1 to Proposition 3-

4. *Q*

## Appendix to Chapter 5

### **Premises**

The agent makes the game's first move by choosing whether or not to propose an alternative to the policy status quo,  $sq \in \mathcal{X}$ . To propose, the agent must pre-commit to pay a cost  $k_p(\varnothing)$ . If the agent chooses not to pay this cost, then the game ends with the status quo determining each player's payoff. If the agent pays, then he next chooses the proposal's content. We model this choice as the selection of a single point  $x \in \mathcal{X}$ . We assume that once  $x$  is chosen, the agent and speaker know its location, while the principal does not. The agent's sole objective is to maximize his utility, which we represent as  $-|outcome - \hat{I}|$ , where  $outcome \in \{sq, x\}$ .

If the agent makes a proposal, then the speaker makes a statement to the principal. The speaker says either "better" or "worse," where both statements refer to the relative proximity of the proposal to the principal's ideal point. The speaker need not tell the truth.

To identify the dynamics of delegation without verifying our conditions for persuasion, we draw on the lessons from Chapter Three to simplify the current model. We examine two cases. In the first case, we assume that the principal has no basis for trusting the speaker. As a result, the principal treats the speaker's statement as totally uninformative. In the second case, we assume that the speaker's statement is true and that the principal believes it. Examining these two cases accomplishes three things. First, it provides a simple way to incorporate a substantive reality of delegation—some speakers are not credible to some principals. Second, this simplification is sufficient to show the

endpoints of the range of effects that the speaker's statement can have on delegation. Third, this variation keeps us from having to rederive the conditions for persuasion derived in the previous section. That is, in this model, we assume that the speaker's credibility has already been established (exogenous to the interaction described here and, presumably, in the manner described in Chapter Three). This simplification also allows us to employ the subgame perfect Nash equilibrium concept to describe the relationship between the agent and the principal.

If the agent makes a proposal, and after the speaker makes a statement, the principal must choose either the proposal,  $x$ , or the status quo,  $sq$ . The principal's sole objective is to maximize her utility. We represent her utility as  $U(x)$ , where  $U(x) \in \{U(x), U(sq)\}$ . After making this choice, the game ends and the principal receives a payoff in utils.

We also make two innocuous tie-breaking assumptions. First, we assume that if making a proposal and not making a proposal provide the agent with the same expected utility, then he makes no proposal. Second, we assume that if  $x$  and  $sq$  provide the principal with the same expected utility, then she chooses  $sq$ .

The principal knows neither  $x$  nor the agent's ideal point  $c$ . She can, however, form beliefs about the location of  $x$ , which can affect her utility. That is, she knows that  $x$  was chosen by the agent. She also knows the agent's utility function and that  $c$  is the result of a single draw from the distribution  $C$ , which has density  $f_C$  and support on a subset of  $[0,1]$ . She can use her initial knowledge to form beliefs about the range of possible locations for  $x$  as well as the likelihood of each.

## Conclusions

We now present two propositions that describe equilibrium behaviors and outcomes. Proposition 5-1 applies to the case where the speaker is not persuasive. It is equivalent to the case where the game is played without a speaker. Proposition 5-2 applies to the case where the speaker's statement is true and the principal believes it.

**Proposition 5** Suppose that the speaker is not persuasive. Then, the agent proposes his ideal point  $(x, \epsilon)$ , the principal accepts the proposal, and  $c$  is the outcome if and only if  $c \in [sq - k_p, sq + k_p]$  and  $-\int_0^{\max(o, sq - k_p)} |c - p| dC - \int_{\min(sq + k_p, 1)}^1 |c - p| dC > -|p - sq|$ . Otherwise, the agent does not participate and  $sq$  is the outcome.

In words, if the principal correctly perceives her and the agent's interests to be sufficiently similar and there is either no speaker or a non-persuasive speaker, then delegation succeeds. If, instead, this perception is incorrect, the delegation fails. By contrast, if the principal perceives her and the agent's interests to be dissimilar, then the consequence of delegation is the status quo.

A proof of Proposition 5-1 is provided in Lupia 1992. The crux of the proof is that the principal cannot tell whether the agent sets  $x = \epsilon$  or whether he chooses a point that is closer to the principal's ideal point. While the principal would like to induce the agent to make a proposal that is more favorable to her, she is not sufficiently knowledgeable to induce such behavior. Therefore, both the agent and principal know that if he makes a proposal, the agent can commit to no other proposal strategy than  $x = \epsilon$ .

What determines whether the agent makes a proposal is whether  $|c - p| > |sq - p| + k_p$ , which occurs when  $c \in [sq - k_p, sq + k_p]$ , and whether the principal, who can infer that  $x \in [c - \epsilon, c]$ , will approve  $x$ , which occurs when the principal's expected payoff from the proposal

$$\int_{\min(sq+k_p, 1)}^c |c - p| dC' - \int_0^{\max(0, sq-k_p)} |c - p| dC' > -|p - sq|.$$

The agent makes no proposal if his ideal point is within  $k_p$  of the status quo. The reason for this choice is for all points within  $k_p$  of  $sq$ , the agent cannot gain enough from making a proposal to recover the cost of doing so. The logic underlying this part of the result is the same as the logic of Proposition 3-4.

We state Proposition 5-2 for the case where  $p \leq sq$ . The case  $p > sq$  is equivalent. Let  $\epsilon > 0$  be a very small number and let  $k_p > \epsilon$ .

**Proposition 5** Suppose that the speaker's statement is true and that the principal believes it. Then, if  $2 \times |sq - p| > k_p$  and  $c \in [sq - k_p, sq + k_p]$ , then the agent proposes his deal point  $(x = c)$ , and  $c$  is the outcome. If  $2 \times |sq - p| < k_p$  and  $c \in [sq - k_p, sq + k_p]$ , then the agent proposes  $x = sq - k_p + \epsilon$ , and  $sq - (2 \times |sq - p|) + \epsilon$  is the outcome. Otherwise, the agent does not participate and  $sq$  is the outcome.

**Proof:** The agent makes a proposal only if he anticipates that the principal will accept it and that the gain in agent utility from such acceptance is greater than  $k_p$ . If the agent either expects the principal to reject the proposal, or the gain in utility is less than  $k_p$ , then

the agent's best response is to make no proposal. Otherwise, the agent's best response is to choose  $x = c$  when  $c \hat{I} (\max\{sq - \ell, \hat{I}(sq - p) - k - p\})$  and to choose  $x = sq - \ell - \hat{I}(sq - p) + e$  when  $c \hat{I} (\max\{sq - \ell, \hat{I}(sq - p)\})$ .

In the case where the speaker's statement is believed, the necessary and sufficient condition for the principal to accept the proposal is to hear the statement "better" and the necessary and sufficient condition for the principal to hear this statement is  $x \hat{I} (sq - \ell - \hat{I}(sq - p) + sq)$ .  $\square$

Had we assumed, instead, that the speaker was persuasive and deceptive, then it follows that the persuasive speaker's effect would be to induce the speaker to make proposals that are less favorable to the principal.

**Corollary to Proposition 5-2**. The introduction of the speaker can influence the agent's proposal. Moreover, if either the agent and principal have common interests or the speaker speaks in a context where the conditions for enlightenment are satisfied, communication can prevent the failure of delegation and lead to its success.

**Proof:** To see the validity of this statement consider the set of cases in Proposition 5-2 where the agent chooses  $x = c$  or decides not to challenge even though there exists an  $x$  that is better for him than  $sq$ . Had the exact same circumstances applied in the non-persuasive context, Proposition 5-1 implies that the agent would have chosen  $x = c$ , which could only be worse for the principal.  $\square$

The proofs of the Chapter Five theorems follow straightforwardly. The proofs of Theorems 5-1 and 5-3 follow directly from Propositions 5-1 and 5-2 and their Corollary. The proof of Theorem 5-2 follows directly from the Theorems 3-1 through 3-3.