

# Leadership, Followership, and Beliefs About the World: An Experiment

Eric S. Dickson  
Assistant Professor  
Department of Politics and  
Center for Experimental Social Science  
New York University

## **Abstract**

One key function of political leadership is to facilitate coordinated action by “followers.” The need to fulfill this function can affect leaders’ strategic incentives to depict the political world in one way as opposed to another when choosing political communications. This paper describes a laboratory experiment that was conducted to explore the effects of a leader’s communications on followers’ beliefs about the world. Interactions take place within a simple game-theoretic framework, in which a leader communicates with followers who are *ex ante* uncertain about the state of the world. In this framework, followers’ preferences over coordination outcomes are aligned in some states, but not in others. As a result, leaders sometimes have incentives to misrepresent the state of the world in order to make coordination more likely. The key experimental finding is that leaders’ communications strongly influence followers’ beliefs about the world even under conditions when Bayesian-rational followers would not find these communications to be credible; followers appear not fully to account for leaders’ strategic incentives to misrepresent the world in forming their posterior beliefs. This result suggests one mechanism through which members of different political groups may come to have different beliefs about the world.

# 1 Introduction

Among the many challenges faced by political leaders is the need to coordinate the actions of “followers.” From Schelling (1960) onward, scholars have emphasized that groups must often overcome coordination problems in order to be effective in the political sphere. Decisions as to how best to allocate group resources may in the end lead to costly delays and other losses if groups fail to coordinate smoothly in making collective decisions (Calvert 1992). A political party may successfully win power if its activists coordinate in advocating some reasonably good policy, but may fail to do so if divisions among activists inhibit the prospects for effective group action (Dewan and Myatt 2007). Coordination problems may be particularly difficult to overcome when individuals within a group have different preferences over the potential outcomes of collective action – a state of affairs that is more the norm than the exception in politics.

While leaders face the challenge of coordinating followers’ behavior, they also have the responsibility to inform followers about the political world. Indeed, considerable traditions of experimental and empirical research within political science suggest that individuals *can* learn about the political world through the communications they receive from leaders (e.g., Lupia and McCubbins 1998), and that they actually *do* so in practice (e.g., Berelson, Lazarsfeld, and McPhee 1954; Zaller 1992). This responsibility, however, also offers leaders an *opportunity* that is relevant to the kinds of coordination problems political groups must frequently overcome. Given that followers learn about the political world from leaders, it may be possible for leaders sometimes to shape followers’ understanding of the world in a way that makes group coordination easier. In the words of Levi (2006), “leadership aligns incentives...provides the learning environment that enables individuals to transform or revise beliefs, and plays a major role in inducing preferences.”

Taken together, these observations provoke several key questions about the function of political leadership and the way in which communications from leaders affect followers’ beliefs and behavior. Does the need to coordinate follower behavior affect leaders’ incentives to depict the political world in

one way, as opposed to another? By what mechanism do leaders' communications about the political world affect prospects for successful coordination among followers? How do followers update their beliefs about the world in the light of what leaders tell them? And, if leaders sometimes have incentives to misrepresent the state of the world in order to enhance the ultimate likelihood of coordination, do followers accurately account for these incentives when deciphering the informational content of leaders' messages?

This paper explores these questions in the context of a game-theoretic laboratory experiment. Existing literature demonstrates that a game-theoretic approach can be fruitfully applied towards understanding the potential role of leadership in coordinating followers' behavior, both from a theoretical (e.g., Calvert 1992; Banks and Calvert 1992; Dewan and Myatt 2007; Dewan and Myatt 2008) and from a laboratory-experimental perspective (e.g., Wilson and Rhodes 1997).<sup>1</sup> In the context of the research questions posed above, a stylized, game-theoretic setting also offers distinct advantages. In such a setting, leaders' incentives can be clearly specified, and followers' patterns of posterior beliefs can be directly compared to benchmarks associated with fully-rational agency.

In Section 2, the paper describes a "Leadership and Coordination Game." In the game, the interests of leader and followers alike depend on coordination being successfully achieved. Three key features of the model are that (1) followers, unlike the leader, are *ex ante* uncertain about the state of the world; (2) all actors' preferences over outcomes are state-dependent; (3) and the actors' interests are fully aligned in some states of the world ("Agreement States") but not in others (the "Disagreement State"). Because preferences are state-dependent, followers' *ex ante* uncertainty about the state of the world induces them to be initially uncertain about their own underlying interests. In the game, communications from the leader take the form of a cheap talk "message," which can be thought of as the leader's strategically-chosen depiction of the world to her followers. Followers then choose coordination alternatives after receiving the leader's message.

---

<sup>1</sup>Of course, an immense game-theoretic literature focuses on the strategic transmission of information in many other kinds of settings. With respect to the study of leadership, Canes-Wrone, Herron, and Shotts (2001) present an incomplete-information model that explores a very different context.

Given the structure of the game, a multiplicity of Perfect Bayesian Equilibria exist; regardless of followers' posterior beliefs about the state of the world, there will always be multiple sets of mutual best responses available to the followers. Section 3 explicates one specific mechanism through which a leader's message might influence the prospects for coordination by fully-rational followers. Under this mechanism, "Leadership-Related Equilibrium," followers use the leader's message as a simple correlation device, selecting mutual best responses in a way that is directly conditional on the message the leader chose to send. A key feature of the incentive structure under this mechanism is that leaders sometimes have incentives strategically to misrepresent the state of the world. This is true in part because messages indicating an Agreement State are more likely to induce successful coordination than messages indicating the Disagreement State, and in part because the leader herself has preferences over specific coordination outcomes.

Section 4 then describes the protocol for a laboratory experiment that involves numerous rounds of play of a Leadership and Coordination stage game. After each play of the stage game, information on followers' posterior beliefs about the state of the world was obtained using an incentive-compatible mechanism rewarding followers based on accuracy. Under assumptions that are well-suited to the laboratory scenario, Leadership-Related Equilibrium is associated with an unambiguous prediction about the message that a given type of leader would choose to send in each state of the world. This feature of the scenario makes it possible to define a rational-choice benchmark specifying the posterior beliefs that would be held by Bayesian followers after particular communications from the leader were received.

Section 5 presents the experimental results. The data indicate that patterns of behavior in the Leadership and Coordination game are largely consistent with the expectations of Leadership-Related Equilibrium. As predicted by the model, followers achieved coordination at extremely high rates when leaders sent messages indicating that the state of the world was an Agreement State, but did so at much lower rates when leaders sent Disagreement State messages. Most leaders took considerable advantage of this property of follower behavior, and chose routinely to misrepresent the state of the world when it

was in their interests to do so. Despite the relative transparency of leaders' incentives to misrepresent the state of the world, however, the evidence indicates that leaders' messages strongly influenced not only followers' behavior, but also their beliefs about the state of the world, even under circumstances when such messages would not have been considered credible as statements of fact by Bayesian-rational followers. As a result, followers' posterior beliefs about the world were biased strongly in the direction of leaders' communications, relative to the rational-choice benchmark. In short, followers appeared not fully to account for leaders' strategic incentives to misrepresent the state of the world in forming their posterior beliefs.

At first glance, this result appears to stand somewhat in contrast with some related experimental literature on political communications. In Lupia and McCubbins (1998), for example, receivers of strategically-chosen communications about a state variable are quite good at factoring in senders' interests when inferring the likely meaning of such communications. Notably, however, in the most relevant treatments of their study, speakers with an incentive to dissemble had interests that were diametrically opposed to listeners' interests. In the scenario described here, the leader's interests are instead at least *partially* aligned with those of each follower, because leader and followers alike share a common interest in coordination being successfully achieved. This contrast highlights the extent to which informational conditions and the structure of preferences may affect the dynamics of leadership and the extent to which leaders' messages may be effective in shaping followers' beliefs about the world.

While the experimental results draw particular contrasts with existing experimental literature, they exhibit suggestive commonalities with longstanding strands of thought in the empirical study of political behavior. From the classic formulation of Campbell *et al* (1960) that party identification raises a "perceptual screen through which the individual tends to see what is favorable to his partisan orientation," through more contemporary research detailing partisan polarization even in beliefs about basic political facts (e.g. Bartels 2002, Achen and Bartels 2006), decades-old ideas suggest that group membership causally shapes the way that individuals come to understand the political world in general and their own political interests in particular.

Taken together, the game-theoretic model and the experimental results presented here suggest a novel causal mechanism consistent with this literature’s findings. Members of partisan (and other) groups learn about aspects of the world relevant to group membership in part through communications from group leaders. If such leaders sometimes have incentives to misrepresent the world in order to facilitate group coordination, and if group members do not fully take these incentives into account when updating their beliefs about the world, then members of a given group may emerge with posterior beliefs that are systematically biased in the direction of the leader’s communications. Given that different groups have different leaders, whose interests may be served by sending different messages about the state of the world, such biases resulting from internal group dynamics could naturally be a source of *across-group* polarization in beliefs about the world.

## 2 The Leadership and Coordination Game

This section describes a model of interaction among three fully-rational agents: one (exogenous) Leader and two Followers,  $F_A$  and  $F_B$ .

Initially, the state of the world  $\omega \in \Omega = \{1, 2, 3\}$  is determined by a random draw from a probability distribution in which  $\omega = n$  with probability  $\rho_n > 0$  for all  $n \in \{1, 2, 3\}$ ,  $\sum_n \rho_n = 1$ . In addition, the Leader’s type  $\theta \in \{L_A, L_B\}$  is also determined by a random draw, from a probability distribution assigning probability  $q \in [0, 1]$  to  $\theta = L_A$  and complementary probability  $1 - q$  to  $\theta = L_B$ . The Leader learns both  $\omega$  and  $\theta$ , but the Followers know only the probability distributions defined by the  $\rho_n$  and  $q$ . All features of this structure of information are common knowledge.

The Leader begins by sending a common message  $m \in M = \{1, 2, 3\}$  to both Followers. This message can be thought of as a public cheap talk claim about what the state of the world is. Communication is costless for the Leader, and the contents of her<sup>2</sup> message may or may not correspond to the true state of the world (that is, it could either be true that  $m = \omega$  or that  $m \neq \omega$ ).

Upon receipt of this common message, each of the two Followers must simultaneously choose one

---

<sup>2</sup>For clarity of prose, throughout the Leader is referred to using female pronouns, while the Followers are referred to using male pronouns.

of two alternatives:  $A$  or  $B$ . Ultimately, each Follower receives utility that depends not only on both Followers' choices but also on the state of the world. Specifically, Followers' state-dependent preferences are as summarized in the payoff matrices below:

Payoff Matrix for State of the World  $\omega = 1$ .

.	$F_B: A$	$F_B: B$
$F_A: A$	1, 1	0, 0
$F_A: B$	0, 0	$\mu, \mu$

Payoff Matrix for State of the World  $\omega = 2$ .

.	$F_B: A$	$F_B: B$
$F_A: A$	1, $\mu$	0, 0
$F_A: B$	0, 0	$\mu, 1$

Payoff Matrix for State of the World  $\omega = 3$ .

.	$F_B: A$	$F_B: B$
$F_A: A$	$\mu, \mu$	0, 0
$F_A: B$	0, 0	1, 1

This payoff structure has two key features. First, note that in every state of the world, each Follower has strict best response  $A$  (resp.,  $B$ ) if the counterpart Follower chooses  $A$  (resp.,  $B$ ). That this is true in *every* state of the world implies that, in expectation, Followers must also have strict best response  $A$  (resp.,  $B$ ) at each of their information sets, given any strategy profile in which the counterpart Follower chooses  $A$  (resp.,  $B$ ), regardless of Followers' beliefs about  $\omega$ . This structure also implies that Followers must have mutual best responses in (non-degenerate) mixed strategies at each of their information sets, with specific mixing probabilities that do depend on Followers' beliefs. Second, the Followers' preferences over outcomes are fully aligned in some *but not all* states of the world. Because Followers are *ex ante* uncertain about the state of the world, they are therefore also *ex ante* uncertain about the underlying nature of their interaction with one another – that is, whether their interests are in truth fully aligned, or not.

In all states of the world, outcomes of “coordination failure” –  $(A, B)$  or  $(B, A)$  – yield utility  $\pi = 0$  for both Followers. However, Followers' utilities for “successfully coordinated” outcomes vary with  $\omega$ .

When  $\omega = 1$ ,  $F_A$  and  $F_B$  share an identical preference structure: a successful coordination outcome  $(A, A)$  yields utility  $\pi = 1$  for both Followers, while a successful coordination outcome  $(B, B)$  yields inferior utility  $\pi = \mu \in (0, 1)$  for both Followers. When  $\omega = 3$ ,  $F_A$  and  $F_B$  again share an identical preference structure, but one that differs from their preference structure under  $\omega = 1$ ; outcome  $(B, B)$  yields utility  $\pi = 1$  for both Followers, while outcome  $(A, A)$  yields inferior utility  $\pi = \mu$  for both. In contrast to these two states, in the final state of the world,  $\omega = 2$ ,  $F_A$  and  $F_B$  do *not* share an identical preference structure.  $F_A$  receives utility  $\pi = 1$  when the outcome is  $(A, A)$ , but only  $\pi = \mu$  when the outcome is  $(B, B)$ ; the reverse is true for  $F_B$ , who receives utility  $\pi = 1$  when the outcome is  $(B, B)$ , but only  $\pi = \mu$  when the outcome is  $(A, A)$ .<sup>3</sup>

The Leader herself gets utility from outcomes that depends not only on Followers' choices, but also on the state of the world and the Leader's type  $\theta \in \{L_A, L_B\}$ . Specifically, a Leader of type  $\theta = L_A$  has an identical preference structure to Follower  $F_A$ , for any outcome and in every state of the world; similarly, a Leader of type  $\theta = L_B$  has an identical preference structure to Follower  $F_B$ , for any outcome and in every state of the world. As mentioned above, the Leader knows her own type, but Followers know only the probability distribution characterized by  $q$ .

The structure of preferences described here reflects a simple, but intuitive, account of group life. In many settings, individuals can benefit when the groups of which they are members achieve successful coordination. Yet, members of a given group may agree under some circumstances, but disagree under others, about the *specific* coordination outcome they would *most* like to achieve. In the present framework, Followers  $F_A$  and  $F_B$  (and the Leader) share identical preferences as to *which* coordination outcome would individually be best in two of the three states of the world. These states,  $\omega = \{1, 3\}$ , will be referred to as *Agreement States*. In contrast, Followers  $F_A$  and  $F_B$  prefer different coordination outcomes in the other state of the world. This state,  $\omega = 2$ , will be referred to as the *Disagreement State*.

The potential either for agreement or for disagreement, depending on the circumstances, within

---

<sup>3</sup>This motivates the notation  $F_A$  and  $F_B$ , distinguishing the Followers based on the outcome each prefers in the state of the world where their preferences differ.

the context of an intertwined strategic destiny is quite typical of political and other social groups. Members of a given political party, for instance, may share a common interest in coordinating for the purposes of winning power. Yet, different party members may have differing underlying tendencies; for example, some may incline more towards a dovish foreign policy, while others may possess more hawkish inclinations. Such differing tendencies indicate the potential for disagreement under some circumstances, while distinct factions would nonetheless find themselves in complete agreement under others. In certain states of the world, more hawkish individuals may favor bold intervention, while more dovish ones would not. At the same time, in a state of the world following a direct attack against the country by a menacing threat, or in a state of the world where any present threat is considered by all sides to be minor and more amenable to negotiations, group members will instead share common preferences over potential courses of action.

### 3 Leader Messages and Follower Behavior: A Mechanism

The Leadership and Coordination Game described in the previous section is quite simple. Yet, a Leader’s message about the state of the world could potentially influence the behavior of fully-rational Followers through a number of distinct mechanisms. This section explicates the logic of one such mechanism that is plausible both in general but particularly in the context of the experimental scenario to be described in the next section. The section also offers equilibrium analysis within the context of this mechanism, ultimately providing a useful rational-choice benchmark for comparison with subjects’ behavior and belief formation in the laboratory context.

To begin, it will prove useful to define the following terminology:

**Definition 1. Self-Interested Messages.** Suppose that the state of the world is  $\omega$ , and that in this state the Leader’s most-preferred outcome is  $x \in \{(A, A), (B, B)\}$ . Define  $\omega^*$  to be that Agreement State in which  $x$  is the most-preferred outcome of both Followers. Then a message  $m^*$  is said to be the Leader’s *self-interested message* in state of the world  $\omega$  if  $m^* = \omega^*$ . ■

According to this definition, a Leader’s message is “self-interested” if it claims – truthfully or

untruthfully – that the outcome privately most-preferred by the Leader is the best outcome for everyone, Leader and Followers alike. Intuitively, to the extent that such a message may sometimes induce rational Followers to choose the Leader’s most-preferred outcome – via whatever mechanism – the Leader may sometimes receive a benefit from sending it.

Clearly, a Leader of either type ( $L_A$  or  $L_B$ ) has self-interested message  $m^* = 1$  when  $\omega = 1$ , and self-interested message  $m^* = 3$  when  $\omega = 3$ . In these Agreement States, a Leader who sends her self-interested message makes a *truthful* statement about the state of the world (that is, sends a message  $m = \omega$ ). However, when  $\omega = 2$ , a Leader of type  $L_A$  has self-interested message  $m^* = 1$ , while a Leader of type  $L_B$  has self-interested message  $m^* = 3$ . Thus, in the Disagreement State, a Leader (of either type) who sends her self-interested message makes a *false* statement about the state of the world (that is, sends a message  $m \neq \omega$ ). Of course, because Followers do not directly observe the state of the world, they do not have direct knowledge of whether a given message is true or false as a depiction of the world; additionally, because Followers do not directly observe the Leader’s type, they do not have direct knowledge of whether a given message was in fact a self-interested message.

### **Leadership-Correlated Equilibrium**

Given the structure of Followers’ state-dependent payoffs, Followers will always have mutual best responses in pure strategies ( $(A, A)$  and  $(B, B)$ ) as well as in (non-degenerate) mixed strategies, regardless of their posterior beliefs about the state of the world. As a result, the Leadership and Coordination Game has a multiplicity of Perfect Bayesian Equilibria. This observation suggests that it may be useful to specify in detail a specific mechanism through which a Leader’s message might plausibly be expected to help coordinate the behavior of rational Followers in equilibrium.

A natural mechanism through which Leaders’ communications may coordinate Followers’ actions is simply through providing a *focal point* around which Followers may rally. Consider a setting in which neither  $A$  nor  $B$  initially constitutes a focal alternative for Followers. In such a setting, in the absence of a message from a Leader, it seems natural to suppose that Followers will select mutual

best responses involving the play of non-degenerate mixed strategies. This state of affairs, of course, is inefficient because Followers will sometimes fail successfully to coordinate. Intuitively, a Leader’s message may affect Follower behavior by bestowing a focal property either on  $A$  or on  $B$ . It is natural to suppose that such a focal property may assist in the selection of an equilibrium in which Followers’ mutual best responses involve pure strategies – and perfect coordination.

It is useful formally to define a relationship between Leaders’ messages and the alternatives for Followers that are made focal by specific messages. Of course, it may be the case that some given message can bestow a focal property on some specific alternative, while another given message is not able to do so. Define  $M_F$  to be the set of messages, each of which has the ability to make one specific alternative focal, and define  $M_{NF}$  to be the set of messages (possibly empty) that do not have this ability. Clearly  $M = M_F \cup M_{NF}$ . Define a *focality function*  $f : M_F \rightarrow \{A, B\}$  as a mapping which links each message in  $M_F$  with the element of  $\{A, B\}$  upon which it bestows a focal property. The alternative made focal by any specific  $m \in M_F$  will accordingly be represented by  $f(m)$ .

With this notation, it is possible to offer the following definition:

**Definition 2.** A *Leadership-Correlated Equilibrium* is a Perfect Bayesian Equilibrium of the Leadership and Coordination Game that has the following properties:

- (i) When the Leader chooses any message  $m \in M_F$ , both Followers choose the same alternative  $f(m)$ ;
- (ii) When the Leader chooses any message  $m \in M_{NF}$ , each Follower randomizes his choice using the relevant mixing probabilities;<sup>4</sup>
- (iii) The Leader optimally chooses a message, given this profile of Follower responses. ■

The motivation behind this definition is straightforward. Followers have available to them three potential sets of mutual best responses (two in pure and one in mixed strategies); certain messages from the Leader may bestow a focal property on one or the other of Followers’ pure-strategy alternatives;

---

<sup>4</sup>In a Perfect Bayesian Equilibrium, Followers will hold posterior beliefs about the state of the world, given Bayes’ Rule and the strategies of the two Leader types; these posterior beliefs imply expected utilities (to Followers) for each potential outcome. Throughout, “relevant mixing probabilities” refers to the unique, non-degenerate mixed strategies that would comprise a mutual best response for  $F_A$  and  $F_B$ , given these in-equilibrium posterior beliefs.

when a Leader sends such a message, Followers can use it as a correlation device, thereby coordinating successfully on the newly-focal alternative; and a Leader has incentives to choose a message that induces Followers to coordinate on the feasible outcome she most prefers.

It is quite natural to suppose that different assumptions might be appropriate about *which* alternatives are made focal by *which* messages, depending on aspects of the specific setting, such as actors' past history of interactions with one another. No one assumption about focality can be appropriate to all settings of political communication.

Nonetheless, it will prove useful to explicate and derive the consequences of one especially natural set of assumptions. In state of the world  $\omega = 1$ ,  $(A, A)$  Pareto dominates all other outcomes. Recall that Followers are *ex ante* uncertain about the true state of the world, and continue to suppose that, in some specific setting, neither  $A$  nor  $B$  is initially a focal alternative for Followers. Consider Followers' potential reactions to a message  $m = 1$ . Intuitively, such a message may bestow a focal property on alternative  $A$  – because, after all, the outcome  $(A, A)$  is Pareto dominant in the state  $\omega = 1$  evoked by this message. By a parallel argument, a message  $m = 3$  may bestow a focal property on alternative  $B$ . In contrast, a Disagreement State message  $m = 2$  would not by the same argument bestow a focal property either on  $A$  or on  $B$ , because there is no Pareto dominant outcome in state  $\omega = 2$ ; a message  $m = 2$  cannot in this way break any perceived symmetry between  $A$  and  $B$ . These intuitions motivate the following assumption:

**Assumption 1.**  $M_F = \{1, 3\}$ , with  $f(1) = A$  and  $f(3) = B$ , while  $M_{NF} = \{2\}$ . ■

Under this assumption, it is straightforward to establish the following Proposition:

**Proposition 1.** *Under Assumption 1, it is the case that in every Leadership-Related Equilibrium of the Leadership and Coordination game: (1) a Leader of either type sends her self-interested message in every state of the world and (2) Followers coordinate on the Leader's preferred outcome in every state of the world.*

**Proof.** The proof, as well as a full specification of the relevant Perfect Bayesian Equilibria, is contained in the Appendix. ■

The intuition behind the Proposition is straightforward. In the context of Assumption 1, the Leader has at her disposal messages that potentially could make either  $A$  or  $B$  a focal alternative for Followers. By strategically choosing which message to send, and therefore which alternative to make focal, the Leader can guarantee that Followers will coordinate on her most-preferred outcome in a Leadership-Related Equilibrium.

It is worth reiterating that this model describes the behavior of fully-rational agents. Naturally, followers who are fully rational update their beliefs about the state of the world after observing the Leader's message, given Bayes' Rule and the equilibrium strategies of the two Leader types. Specifically, in the context of the Perfect Bayesian Equilibria relevant to Proposition 1, Followers will share posterior beliefs  $(\bar{\rho}_1, \bar{\rho}_2, \bar{\rho}_3) = (\frac{\rho_1}{\rho_1 + \rho_2 q}, \frac{\rho_2 q}{\rho_1 + \rho_2 q}, 0)$  upon receipt of a message  $m = 1$  and posterior beliefs  $(\bar{\rho}_1, \bar{\rho}_2, \bar{\rho}_3) = (0, \frac{\rho_2(1-q)}{\rho_3 + \rho_2(1-q)}, \frac{\rho_3}{\rho_3 + \rho_2(1-q)})$  upon receipt of a message  $m = 3$ .<sup>5</sup>

It will prove useful to define one additional piece of terminology:

**Definition 3.** A message  $m = n$  will be said to be *credible as a statement of fact* on the equilibrium path if Followers' posterior beliefs satisfy  $\bar{\rho}_n > \frac{1}{2}$ . ■

That is, a message is "credible as a statement of fact" if both Followers believe it to be more likely than not *ex post* that the message is an accurate depiction of the state of the world. Whether, in a particular context, a given message will be credible as a statement of fact depends on the Leader's strategy profile and the prior probabilities of the different states of the world. Given Proposition 1, which indicates that a Leader of either type always sends her self-interested message, this definition implies that the Agreement State messages  $m = 1$  and  $m = 3$  will be credible as statements of fact (or not) under the following conditions:

---

<sup>5</sup>Followers' posterior beliefs upon receiving an off-the-equilibrium path message  $m = 2$  are of course not constrained in this way.

	Conditions on $\rho$ 's	$m = 1$ credible as statement of fact?	$m = 3$ credible as statement of fact?
Case I	$\frac{\rho_1}{\rho_1 + \rho_2 q} > \frac{1}{2}$ and $\frac{\rho_3}{\rho_3 + \rho_2(1-q)} > \frac{1}{2}$	yes	yes
Case II	$\frac{\rho_1}{\rho_1 + \rho_2 q} < \frac{1}{2}$ and $\frac{\rho_3}{\rho_3 + \rho_2(1-q)} < \frac{1}{2}$	no	no
Case III	$\frac{\rho_1}{\rho_1 + \rho_2 q} < \frac{1}{2}$ and $\frac{\rho_3}{\rho_3 + \rho_2(1-q)} > \frac{1}{2}$	no	yes
Case IV	$\frac{\rho_1}{\rho_1 + \rho_2 q} > \frac{1}{2}$ and $\frac{\rho_3}{\rho_3 + \rho_2(1-q)} < \frac{1}{2}$	yes	no

Figure 1 offers a graphical depiction of these conditions for the case when  $q = \frac{1}{2}$ . In the Figure, the axes depict the prior probabilities  $\rho_1$  and  $\rho_2$ ; the value of  $\rho_3$  is implicit because  $\rho_3 = 1 - \rho_1 - \rho_2$ . In Case I, the prior probability  $\rho_2$  of the Disagreement State is not too large relative to the prior probabilities of each of the individual Agreement States. As such, upon receipt of an Agreement State message ( $m = 1$  or  $m = 3$ ), Followers will place relatively high posterior probability on the Agreement State message being an accurate depiction of the world. Thus, in Case I, either Agreement State message will be credible as a statement of fact. The opposite is true of Case II, for which  $\rho_2$  is not too *small* relative to the prior probabilities of each of the individual Agreement States; as such, upon receipt of an Agreement State message, Followers will place relatively low posterior probability on the Agreement State message being an accurate depiction of the world, and neither Agreement State message will be credible as a statement of fact. Finally, in Cases III and IV, the prior probability of *one* of the Agreement States is relatively high compared with that of the Disagreement State, but the prior probability of the other Agreement State is relatively low compared with that of the Disagreement State. Here, a message corresponding to the *ex ante* likelier Agreement State will be credible as a statement of fact, but a message corresponding to the *ex ante* less-likely Agreement State will not.

\* FIGURE 1 ABOUT HERE \*

While fully-rational followers update their beliefs about the state of the world in the way described here, it is worth re-emphasizing that it is irrelevant to the mechanism underlying Leadership-Related Equilibrium whether a Leader's message is credible as a statement of fact. That is, the ability of a message to bestow a focal property on some particular alternative is not a function of that message's credibility. In this specific sense, the mechanism underlying Leadership-Related Equilibrium is

effectively decoupled from Followers’ actual *beliefs* about the state of the world. Nonetheless, Follower behavior is fully rational in Leadership-Related Equilibria; Followers always choose mutual best responses, albeit in a way that is contingent on the specific common message they receive.

## 4 Experimental Instantiation

A laboratory experiment was conducted as a means of exploring how Leaders’ communications ultimately affect Followers’ behavior and beliefs about the state of the world in the context of the theoretical scenario described above. Five experimental sessions were conducted in a social science lab at a large American university. The 96 subjects, each of whom took part in one session only, interacted anonymously via networked computers; the experiments were programmed and conducted with the software z-Tree (Fischbacher 1999). Participants, almost all of whom were undergraduates from around the university, signed up via a web-based recruitment system that draws on a large, pre-existing pool of potential subjects. Subjects were not recruited from the author’s courses. After giving informed consent according to standard human subjects protocols, subjects received written instructions that were subsequently read aloud in order to promote understanding and induce common knowledge of the experimental scenario. Following the reading of the instructions, but before the experiment began, subjects took on an on-screen quiz testing their understanding of the instructions and giving them feedback as to whether their answers were correct. Virtually all quiz answers offered by subjects were correct.

### Matching and Basic Structure

At the beginning of each session, subjects were randomly assigned to a group of three people, consisting of a randomly-assigned “Group Speaker” and two “Group Members”; these role labels were thought to be more neutral than Leader and Follower, their analogues in the theoretical exposition.<sup>6</sup> Group and

---

<sup>6</sup>The Appendix contains a sample set of instructions to subjects, including an extensive series of screenshots showing the computer interface, that offers a complete depiction of the way the experiment was framed for participants. The description in this section places terminology from the experimental scenario in quotation marks where this differs from the theoretical exposition; for continuity, however, the results and analysis are presented in the same terms as the theoretical exposition.

role assignments remained fixed over 15 periods of interaction.

Each of these 15 periods consisted of two basic parts. In the first part, subjects within a group engaged in one play of the Leadership and Coordination stage game. In the second part, Followers answered a “Bonus Question,” which asked them to guess what the state of the world for that period actually had been. In both of these parts, subjects earned “tokens” that were ultimately converted into dollars at a known rate (15 tokens = US\$1) at the end of the experiment; a subject’s total earnings were equal to the sum of his or her payoffs for each of the 15 periods, plus a US\$5 show-up fee.

### **Leadership and Coordination Stage Game: Parameters**

The Leadership and Coordination stage game employed in the experiment had the same game-theoretic structure as the formal framework described in Section 2. The parameters  $q$  and  $\mu$ , relevant to actors’ payoffs, were held fixed during each experimental session; however, the prior probabilities  $\rho$ , the state of the world  $\omega$ , and the Leader’s type  $\theta$  all varied from period to period, in a way to be described shortly. All five sessions were conducted with  $q = 0.5$ . Two sessions (with 21 subjects each) were conducted with  $\mu = 0.8$ , while three sessions (with 21, 21, and 12 subjects respectively) were conducted with  $\mu = 0.2$ . As in the Section 2 framework, both  $q$  and  $\mu$  were common knowledge. Because no significant differences in Leader or Follower behavior were observed across these conditions, the analyses below pool the data from all five sessions. While  $\mu$  varied, under both conditions coordination on one’s most-preferred alternative yielded a payoff of 10 tokens (= US\$0.667), while a failure to coordinate at all yielded 0 tokens, for Leaders as well as for Followers.

The Leader’s type,  $L_A$  or  $L_B$ , was independently re-drawn for every group in every period from the distribution  $q = 0.5$ ; as such, regardless of past history, the Leader was as likely in any given period to share one of her Followers’ preferences as the other’s. This fact was common knowledge. As noted above, the “likelihood” (prior probability) values  $\rho$  also differed across periods, although each value  $\rho_1$ ,  $\rho_2$ , and  $\rho_3$  always took on a positive integer percentage point value (e.g. 1%, 2%, ..., 98%), with  $\rho_1 + \rho_2 + \rho_3 = 100\%$ . The likelihood triples  $(\rho_1, \rho_2, \rho_3)$  were uniform draws from the probability

simplex, except that the region of the simplex corresponding to Case II was modestly oversampled (using one round of rejection sampling; 27.3% of triples used came from the Case II region, though it covers only 16.7% of the simplex)<sup>7</sup>. These triples were randomly and independently drawn for every group, in every period, through this process, which was not described to subjects. The process ensured that subjects were exposed to a variety of strategic settings, including a substantial number of instances in which Leaders had clear incentives to misrepresent the state of the world. Finally, it was commonly known that the state of the world  $\omega$  was randomly drawn for every group, in every period, based on the distribution defined by the triple  $\rho$  that had been reported to all group members.

### **Leadership and Coordination Stage Game: Aspects of the Interface**

In addition, the experimental design also contained a number of features intended to maintain symmetry between coordination game alternatives and between Followers, both within a given period and across periods. This was important not only to maximize the fit between the experimental setting and the scenario that motivated the theoretical framework, but also to minimize the extent to which Followers could condition their coordination-game choices on previous periods' outcomes. First, the coordination game alternatives,  $A$  and  $B$  in the theoretical exposition, were referred to using different labels in every period. The labels for  $A$  and  $B$  in a given period were places, people, or things belonging to the same category (e.g., Cleveland and Cincinnati; Dandelion and Daffodil; Schooner and Sloop, etc.), with neither label obviously more salient than the other,<sup>8</sup> and each subject saw any given pair of labels only once. Second, the possible states of the world and the corresponding messages available to Leaders (in the theoretical exposition,  $\omega = \{1, 2, 3\}$  and  $m = \{1, 2, 3\}$  respectively) were referred to using the names of colors as labels; three color names were drawn at random from a fixed set of ten for every group and in every period.<sup>9</sup> The actual state of the world was referred to as “the true color” for the period or simply “the color” for the period. Third, the on-screen interface listed coordination

---

<sup>7</sup>And apart from rounding considerations.

<sup>8</sup>Or at least, so it was intended. Data on Followers' coordination game choices, presented in the next section, suggests that this was sufficiently the case.

<sup>9</sup>The colors were White, Gray, Black, Yellow, Orange, Red, Brown, Green, Blue, and Purple.

alternatives and states of the world in an order that was known to be randomized for every subject, in every period, so that neither of Followers' alternatives would become more salient than the other as a locus of coordination due to trivial details like screen order. And finally, to maintain symmetry between the Followers themselves, each Follower was referred to on his own screen as "you," while his counterpart Follower was referred to as "your counterpart." On Leaders' screens, the two Followers were referred to as "Group Member 1" and "Group Member 2" for clarity, but it was known to all subjects that these labels were randomly reassigned from one period to the next, and they were never in any period observed by either Follower.<sup>10</sup>

### **The "Bonus Question" for Followers**

In the second part of each period, Followers were asked to guess that period's actual state of the world ("What is the true color for this period?"). Followers entered Bonus Question responses after choosing coordination game alternatives, but before receiving any feedback about their choices. A Follower offering a correct response earned 10 tokens (= US\$0.667), while a Follower offering an incorrect response earned nothing. Thus, the Bonus Question served as an incentive-compatible mechanism for Followers to reveal the state of the world to which they assigned highest posterior probability. Notably, the 10 tokens earned for a correct Bonus Question response equalled the maximum possible payoff from a given play of the Leadership and Coordination stage game itself, suggesting that Followers were well-motivated by laboratory standards to learn, and accurately report, the state of the world.<sup>11</sup>

It was commonly known that Leaders' payoffs were not affected by Followers' Bonus Question responses. Leaders, who knew the state of the world in each period, were given a flat 5 extra tokens

---

<sup>10</sup>For further details about the interface, and the way in which the interface was described to subjects, see the instructions, including the sample screenshots shown to subjects.

<sup>11</sup>Typical psychology studies exhibiting "biases" in decision making do not offer subjects concrete incentives based on their responses; as a result, such studies are treated skeptically by many political scientists and economists. Any demonstration that some kind of "biased" behavior remains intact even when subjects' responses are rewarded with a financial incentive makes a stronger case for that bias as a robust phenomenon. Of course, it is reasonable to imagine that the presence of such motivation leads subjects to attend to the experimental task differently than they would in the absence of the motivation, which may pose a problem for inference or for an experiment's external validity in some contexts. However, foreshadowing the results, this feature of the design actually strengthens inference in the present experiment; subjects exhibit a particular kind of bias in the lab even though they have additional (monetary) incentives to be unbiased, relative to the incentives typically faced in parallel real-world contexts.

per period in lieu of answering a Bonus Question.

## Feedback to Subjects

At the end of each period, once Followers had given their Bonus Question responses, subjects received some limited feedback. Both Followers and Leaders were informed of the alternatives that had been chosen by each of the two Followers, so that all subjects knew whether coordination had been achieved. However, Followers were never informed what the true state of the world had been; as such, they did not receive feedback about their precise payoffs (either for game play or for their Bonus Question answers) from one period to the next. Leaders, who of course were informed about the state of the world, did learn their precise payoffs at the end of each period. Leaders were never informed about Followers' Bonus Question answers.

The limited feedback to Followers constituted an important part of the experimental design. Obviously, it was desirable to have Followers participate in a number of periods, over which they might be exposed to instances of different theoretical cases, gain familiarity with the experimental interface, and develop deeper insights into the dynamics of the experimental game. A demonstration of a “bias” in belief formation based on one period only would clearly be less convincing than a demonstration of a persistence of some “bias” over repeated interactions. Further, by interacting with a given Leader over a substantial number of experimental periods, Followers potentially could learn from patterns of Leader behavior over time; for example, it could in theory easily be inferred that a Leader who *never* sent a Disagreement State message over many periods must almost certainly have been lying about the state of the world at least some of the time. Allowing Followers to learn about Leader behavior in the context of such an ongoing relationship arguably parallels important dynamics in the real world, where leaders speak to group members about many issues and many facets of the world over time. More substantial feedback to Followers – for example, revealing the true state of the world at the end of every period – would short-circuit such dynamics by making Leaders' honesty, or lack of it, unrealistically transparent; although there are exceptions, citizens in real-world settings of uncertainty

relatively seldom have access to definitively credible, direct and immediate feedback unambiguously describing the true state of the world.

## Overall Earnings

In all, Leaders' average earnings were US\$18.22 in the  $\mu = 0.2$  sessions and US\$18.35 in the  $\mu = 0.8$  sessions. Followers' average earnings were US\$19.51 in the  $\mu = 0.2$  sessions and US\$20.14 in the  $\mu = 0.8$  sessions.

## 5 Experimental Results

### Followers' Actions

The theoretical analysis in Section 3 introduced the concept of Leadership-Related Equilibrium, and demonstrated that this concept makes sharp predictions about behavior, given the focal function described in Assumption 1. In order to explore the fit of these theoretical ideas to the experimental scenario, this section begins by exploring how Followers in the experiment actually made choices in the aftermath of Leaders' messages, an important test of the mechanism underlying Leadership-Related Equilibrium in general and Assumption 1 in particular.

**Experimental Result 1.** *When Leaders sent an Agreement State message, Followers successfully coordinated at very high rates on the outcome that would be Pareto-optimal in that Agreement State. When Leaders sent a Disagreement State message, Followers successfully coordinated at much lower rates.*

A key insight underlying Leadership-Related Equilibrium is that Followers may be able to use Leaders' messages as correlation devices, thereby achieving much higher rates of successful coordination than would be possible in the absence of such messages. The data in Table 1 indicate that this was indeed the case in the experiment, and that Follower behavior was closely in accordance with theoretical expectations.

\* TABLE 1 ABOUT HERE \*

Specifically, the discussion of Leadership-Related Equilibrium surmised that an Agreement State

message from a Leader may bestow a focal property on one of a Follower’s two alternatives – according to Assumption 1, the alternative on which coordination would be Pareto-optimal in that Agreement State. Strikingly, individual Followers indeed chose that “focal” alternative an overwhelming 96.9% (837/864) of the time after receiving an Agreement State message. This pattern of choices led to an overall coordination rate of 93.8% (405/432), conditional on an Agreement State message. 22 of the 32 experimental groups *never* failed to coordinate upon receipt of an Agreement State message; 6 groups accounted for fully 22 of the 27 failures (which occurred either three or four times each per group). These results indicate that subjects in the aggregate came close to the 100% coordination benchmark associated with Proposition 1 and Assumption 1, and that two-thirds of the experimental groups actually achieved 100% coordination rates upon receipt of an Agreement State message.

The practical usefulness of Agreement State messages in achieving coordination is clear when comparing this 93.8% coordination rate with the corresponding 52.1% (25/48) coordination rate achieved following Disagreement State messages; this difference is highly statistically significant ( $p < 0.0001$ , two-tailed). Consistent with Assumption 1, Disagreement State messages do not bestow a focal property on either of a Follower’s alternatives in the same way that Agreement State messages do. That this 52.1% coordination rate is statistically consistent with the 50% benchmark rate of purely random choice ( $p = 0.771$ , two-tailed) suggests that, as intended in the experimental design, Followers were unable systematically to coordinate based on the labels of the alternatives or other features of the experimental interface. Followers’ rates of coordination appear to have been more consistent with purely random choice than with the rates implied by Followers’ mutual best responses in mixed strategies.<sup>12</sup>

**Experimental Result 2.** *Coordination rates were very high when Leaders’ Agreement State mes-*

---

<sup>12</sup>The model did not make assumptions about Followers’ off-the-equilibrium-path beliefs upon receipt of a message  $m = 2$ . However, data presented shortly is suggestive of one natural assumption:  $\bar{p}_2 = 1$ . In the  $\mu = 0.8$  sessions, the coordination rate conditional on a Disagreement State message was 38.9% (7/18), consistent with random choice ( $p = 0.346$ , two-tailed) and with the coordination rate of 49.4% ( $p = 0.374$ , two-tailed) implied by  $\bar{p}_2 = 1$  and Followers’ corresponding mutual best responses in mixed strategies. In the  $\mu = 0.2$  sessions, the coordination rate conditional on a Disagreement State message was 60.0% (18/30), consistent with random choice ( $p = 0.273$ , two-tailed) but inconsistent with the coordination rate of 27.8% ( $p = 0.0001$ , two-tailed) implied by  $\bar{p}_2 = 1$  and Followers’ corresponding mutual best responses in mixed strategies. The 38.9% (7/18) and 60.0% (18/30) figures from the two treatments are statistically consistent with one another ( $p = 0.157$ , two-tailed).

*sages were not credible as statements of fact, but coordination rates were statistically significantly higher when Leaders' Agreement State messages were credible as statements of fact.*

At first glance, the coordination rate of 93.8% following Agreement State messages would seem to leave relatively little variation to analyze. However, a closer look at the data reveals that the likelihood of coordination does vary modestly with the credibility of Leaders' messages. Coordination rates were very high in Case I (95.5%, 190/199) and in Cases III and IV for messages indicating the likelier Agreement State (97.8%, 90/92); overall, successful coordination occurs 96.2% (280/291) of the time under these conditions, where Leaders' messages are credible as statements of fact. In contrast, coordination rates were somewhat lower in Case II (90.0%, 99/110) and in Cases III and IV for messages indicating the less-likely Agreement State (82.1%, 23/28); overall, successful coordination occurs 88.4% of the time (122/138) under these conditions, where Leaders' messages are *not* credible as statements of fact. The difference between these 96.2% and 88.4% figures is statistically significant ( $p = 0.002$ , two-tailed).

The very high rates at which coordination is achieved upon receipt of an Agreement State message are broadly consistent with the expectations of Leadership-Correlated Equilibrium and Assumption 1. However, the effectiveness of Agreement State messages in coordinating behavior does exhibit a mild dependence on the credibility of such messages, in a way that is not anticipated by Leadership-Correlated Equilibrium.

## **Leaders' Messages**

Another key insight from the theoretical framework is that Leaders sometimes have incentives strategically to misrepresent the world in order to facilitate coordination among Followers – specifically, coordination on the Leader's most-preferred outcome. The patterns of Follower behavior described in the previous subsection, largely consistent with theoretical expectations, indicate that Leaders faced such incentives in the experimental setting as well. The data indicates that subjects in the role of Leader overwhelmingly perceived, and responded to, these incentives in the context of the experiment.

**Experimental Result 3.** *When the true state of the world was an Agreement State, Leaders nearly always honestly reported the state of the world. When the true state of the world was instead the Disagreement State, Leaders overwhelmingly misrepresented the state of the world by claiming it to be an Agreement State.*

Consistent with insights derived in Section 3, Leaders were highly averse to sending Disagreement State messages, regardless of the true state of the world. In an Agreement State, of course, the Leader’s self-interested message coincides with the message that honestly reports the state of the world. During periods conducted in an Agreement State, Leaders’ messages honestly reported the state of the world an overwhelming 93.5% of the time (272/291); only 1.7% of the time (5/291) did they send a message corresponding to the Disagreement State. The remaining 4.8% of Leaders’ messages indicated the “other” Agreement State (the one that was not the true state of the world).

In the Disagreement State, in contrast, a Leader’s self-interested message does not coincide with the message that honestly reports the state of the world. During periods conducted in the Disagreement State, Leaders’ messages honestly reported the state of the world only 22.8% of the time (43/189). This figure is obviously statistically different from the 93.5% “honesty” figure for Agreement States, cited above ( $p < 0.0001$ , one-tailed).<sup>13</sup> Instead, in a substantial majority of cases (61.4%, 116/189), Leaders in the Disagreement State sent their self-interested message, indicating the Agreement State in which the Leader’s most-preferred option would be Pareto-optimal. The remaining 15.9% (30/189) of Leaders’ messages indicated the “other,” non-self-interested Agreement State. Notably, the tendency of Leaders to misrepresent Disagreement States as Agreement States increased somewhat over time as experimental sessions progressed.<sup>14</sup>

The 48 instances in which Leaders’ messages announced the Disagreement State were distributed very unequally across subjects in the role of Leader. 15 of the 32 experimental Leaders *never* sent such

---

<sup>13</sup>The analysis employs two-tailed tests except where there is a clear directional hypothesis based on theoretical expectations, as is the case here.

<sup>14</sup>In the Disagreement State, Leaders sent honest messages 29.9% (29/97) of the time in the first 7 periods but 15.2% (14/92) of the time in the last 8 periods; this difference in proportions is significant ( $p = 0.016$ , two-tailed). Leaders instead sent their self-interested message 53.6% (52/97) of the time in the first 7 periods but 69.6% (64/92) of the time in the last 8 periods ( $p = 0.024$ , two-tailed).

a message, while 5 Leaders sent such a message only once. Of the remaining 43 Disagreement State messages, fully 20 were due to only three Leaders, who each sent such a message at least six times; the other 23 were distributed among 9 different Leaders.

\* TABLE 2 ABOUT HERE \*

Table 2 disaggregates Leaders' messages by Case.<sup>15</sup> For the most part, Leader behavior is quite consistent across Cases. Four of the eight panels in Table 2 describe the messages sent by Leaders in Agreement States, disaggregated into four distinct situations. In two of these situations (Case I, Cases III and IV in the likelier Agreement State), an honest Agreement State message would be credible as a statement of fact. In the other two situations (Case II, Cases III and IV in the less-likely Agreement State), an honest Agreement State message would not be credible as a statement of fact. Across all of these four situations, Leaders send (self-interested) messages that honestly communicate the true state of the world between 88.6% and 100% of the time; none of these rates are significantly different (minimum  $p$ -value 0.217, two-tailed). That these rates are close to 100% and statistically invariant across Cases is in accordance with the Proposition 1 result for Leadership-Correlated Equilibrium.

The other four panels in Table 2 describe the messages sent by Leaders in the Disagreement State, again disaggregated into four distinct situations: Case I, Case II, Cases III and IV when the Leader's "preferred" Agreement State is the likelier Agreement State, and Cases III and IV when the Leader's "preferred" Agreement State is the less-likely Agreement State. Across these situations, Leaders send honest Disagreement State messages between 19.0% and 26.5% of the time; again, none of these rates are significantly different (minimum  $p$ -value 0.525, two-tailed). In Case II, neither Agreement State message would be credible as a statement of fact; in the other three situations, at least one Agreement State message *would* be credible as a statement of fact. The observed lack of variation across Cases suggests that Leaders were not systematically sending Disagreement State messages in an attempt to maintain "credibility" with Followers from one period to the next.<sup>16</sup> At the same time, the behavior of

---

<sup>15</sup>Aggregate totals slightly exceed the sum of Case-by-Case totals because a tiny number of draws from the probability simplex fell directly onto a boundary between Cases. Data from periods with these *ex ante* likelihoods is included in the aggregate totals, but is not tallied as belonging to any of the four Cases.

<sup>16</sup>In addition, regression analyses estimate that the rate at which Leaders send Disagreement State messages within

Leaders in Disagreement States does appear modestly to have differed across Cases in one other respect: the propensity to send (self-interested) messages corresponding to their “preferred” Agreement State as opposed to messages corresponding to the “other” Agreement State. In the first three relevant panels – corresponding to Case I, Case II, and Cases III and IV when the Leader’s “preferred” Agreement State is the likelier – Leaders send their self-interested message between 60.0% and 71.4% of the time, instead indicating the other Agreement State between 9.5% and 14.3% of the time. None of the comparisons across these situations suggests a statistically significant difference. In the fourth situation – Cases III and IV when the Leader’s “preferred” Agreement State is the less-likely Agreement State – Leaders send their self-interested message only 44.1% of the time, instead indicating the other Agreement State 29.4% of the time, behavior that does exhibit a significantly different pattern. In this fourth situation, unlike in the other three, the Leader’s self-interested message would *not* be credible as a statement of fact while the “other” Agreement State message *would be*. Taken together, these results suggest that Leaders are not greatly affected by details of the strategic setting when choosing whether or not to misrepresent the Disagreement State in communicating with Followers, but that Leaders who *do* choose to dissemble are influenced in some modest way by these details in choosing a specific misrepresentation to communicate with Followers.

The theoretical expectations encapsulated in Proposition 1 suggested that a Leader could be expected always to send her self-interested message within Leadership-Related Equilibria, regardless of the *ex ante* likelihoods of different states of the world. Overall, the evidence presented above suggests that Leaders send self-interested messages at high rates that respond at most very mildly to distinctions between Cases.

---

Case II is independent of  $\rho_2$ ; the propensity of Leaders to send Disagreement State messages does not increase even as the Disagreement State becomes overwhelmingly likely *ex ante*. A reasonable interpretation of this finding is that, in the aggregate, Leaders do not act as though they are concerned that Followers may cease to “trust” them. It is also worth re-emphasizing that many Leaders *never* sent a Disagreement State message, and that a large fraction of all Disagreement State messages came from a very small number of Leaders.

## Followers' Beliefs about the State of the World

The above experimental evidence indicates that Leaders' messages about the world were highly effective in coordinating Followers' actions, and that most Leaders acted on their incentives strategically to misreport the state of the world. These results establish that Assumption 1 was appropriate in the context of the laboratory scenario, and that subjects' behavior in the Leadership and Coordination Game was largely consistent with the expectations of Leadership-Related Equilibrium. The central question, however, remains. How do Followers form posterior beliefs about the state of the world in this simple experimental scenario? Do they form posterior beliefs in the same way that Bayesians would, fully taking account of Leaders' strategic incentives to dissemble? Or are Followers' beliefs about the world influenced by Leaders' messages in a way that is inconsistent with in-equilibrium Bayesian updating?

As noted above, the Bonus Question served as an incentive-compatible mechanism for eliciting the state of the world on which each Follower placed highest posterior probability. Table 3 summarizes Followers' "guesses" about the state of the world, disaggregated by the kind of message received as well as by Case.

\* TABLE 3 ABOUT HERE \*

**Experimental Result 4.** *When Leaders' Agreement State messages were credible as statements of fact, Followers almost always guessed that these messages accurately portrayed the state of the world.*

In Case I, as discussed above, Leaders' Agreement State messages are credible as statements of fact. As such, Bayesian-rational Followers who receive an Agreement State message in Case I would place highest posterior probability (exceeding  $\frac{1}{2}$ ) on that Agreement State. Followers' guesses are highly consistent with this prediction; in the data, Followers who receive an Agreement State message in Case I guess that the message is an accurate portrayal of the state of the world an overwhelming 90.2% (359/398) of the time. A similar result is observed for messages indicating the likelier Agreement State in Cases III and IV (86.4%, 159/184), which are also credible as statements of fact.

**Experimental Result 5.** *When Leaders' Agreement State messages were not credible as state-*

*ments of fact, Followers nonetheless guessed about 40% of the time that these messages accurately portrayed the state of the world.*

In Case II, however, Leaders' Agreement State messages are *not* credible as statements of fact. Indeed, Bayesian-rational Followers who receive an Agreement State message in Case II would place highest posterior probability (of at least  $\frac{1}{2}$ ) on the *Disagreement State*. In spite of this, the data indicates that Followers who received an Agreement State message in Case II guessed that the Disagreement State was the true state of the world only 56.8% (125/220) of the time. Fully 42.7% (94/220) of the time, Followers instead guessed that the Leader's message accurately communicated the state of the world. A similar result is observed for messages indicating the less-likely Agreement State in Cases III and IV, which are also not credible as statements of fact; upon receipt of such messages, Followers guessed that the Leader's message was an accurate depiction of the world 44.6% (25/56) of the time.

This result is striking. In the experimental scenario, Leaders' incentives to misrepresent a Disagreement State as an Agreement State are fairly transparent. Leaders only receive positive payoffs when Followers coordinate successfully, and Followers are able successfully to coordinate at extremely high rates only when an Agreement State message makes focal one or the other of Followers' alternatives. Indeed, Leaders' incentives to dissemble are arguably much more transparent in the experiment than in richer, more complicated real-world settings where analogous dynamics are not laid so bare. In spite of this, Followers often fail fully to account for Leaders' strategic incentives when updating their beliefs about the world in the aftermath of Leaders' messages. As a result, in the aggregate, Followers' beliefs about the state of the world are systematically biased in the direction of Leaders' messages.

The remainder of this section considers the robustness of, and the interpretation of, this key finding. A first indication that "overbelieving" Leaders' messages may be a robust behavioral phenomenon is that Followers exhibit little aggregate improvement in their guesses about the state of the world as they gain experience and listen to successive messages from a specific Leader over time. Indeed, in Case II, Followers who received a (non-credible) Agreement State message guessed that the state of the world was the Disagreement State 54.5% (48/88) of the time in the first 7 periods and 58.8% (77/131)

of the time in the last 8 periods; the difference between these proportions is statistically insignificant ( $p = 0.53$ , two-tailed).<sup>17</sup>

At the individual level, there was considerable variation in the extent to which different Followers accounted for Leaders' strategic incentives when updating their beliefs about the state of the world. Upon receiving a Case II Agreement State message, 23 Followers *always* guessed the Disagreement State to be the true state of the world, consistent with in-equilibrium Bayesian inference, while 24 Followers *never* guessed the Disagreement State to be the true state of the world. The remaining 13 Followers who were exposed to a Case II Agreement State message split their guesses between Agreement and Disagreement States in different periods.

The in-equilibrium Bayesian posterior beliefs in the theoretical section were calculated under the assumption that Leaders always send their self-interested message. While Leaders do exhibit a pronounced tendency to send self-interested messages, they of course did not do so all of the time. In principle, rational Followers could become aware that Leaders sometimes deviate from their equilibrium messages, and subsequently update their beliefs in a way that takes such deviations into account. For example, a Follower who receives a Disagreement State message could potentially infer that the Leader places some unmodeled value on being honest; if so, the "equilibrium" behavior that the Follower would expect of the Leader may differ from the earlier theoretical predictions – and because of this, the Follower may draw different inferences about the state of the world, given a specific message.

In probing this idea, a first step is to recall that deviations from the theoretically-predicted self-interested messages were distributed very unevenly across Leaders. A simple and natural way to explore the effects of "deviant" messages on Followers' beliefs about the world is simply to split the data on Followers' guesses into two subsamples – first, all guesses made by Followers who had *never* up to that point seen a Disagreement State message, and second, all other guesses, that were made after at least one such exposure. Interestingly, the observed difference across these subsamples was modest in magnitude and statistically insignificant – following a (non-credible) Case II Agreement State message,

---

<sup>17</sup>This and the remaining results for Case II are all comparable to results obtained for non-credible messages in Cases III and IV.

60.1% (83/138) of guesses in the first subsample indicated the Disagreement State, compared to 51.2% (42/82) in the second subsample ( $p = 0.198$ , two-tailed). Notably, the 60.1% figure from the subsample “uncontaminated” by Disagreement State messages is not meaningfully different from the overall 56.8% rate reported above.<sup>18</sup>

The above results are particularly striking because Followers actually had ample opportunity to learn about the nature of their Leader’s behavior over 15 periods of interaction within a fixed matching. In one sense, feedback to Followers was limited – they were not informed of the true state of the world at the end of each period, and could therefore not directly infer that a Leader had misrepresented the world in any *specific* period. However, consider those Followers who were associated with one of the 15 Leaders who *never* sent a Disagreement State message. By the middle of period 8, the halfway point of each group’s existence, each such Follower would have seen 8 consecutive Agreement State messages from his Leader. This fact alone offered considerable evidence that the Leader in such groups was highly unlikely to be sending “honest” messages in every period. For the median group with such a Leader, there was a mere 0.75% probability that the prior probabilities  $\rho$  seen by Followers each period would have generated Agreement States in all eight periods.<sup>19</sup> Yet, Followers in such groups continued, like other Followers, to “overbelieve” Case II Agreement State messages.

This split-sample analysis offers compelling evidence that Followers’ failures to account for Leaders’ strategic incentives cannot substantially be explained by the presence of Disagreement State messages unanticipated by the theory. An alternative method for addressing this question is to incorporate the empirical frequency with which Leaders send Disagreement State messages into Bayesian inference. In the data, Leaders who find themselves in a Disagreement State send messages acknowledging that fact 22.8% of the time. Under an assumption that Leaders in a Disagreement State randomly send Disagreement State messages at this rate, but send their self-interested message the other 77.2% of the time, Followers’ posterior belief  $\bar{\rho}_1(m = 1)$  would for example be:

---

<sup>18</sup>A similar result is obtained when cutting the data slightly differently, restricting the analysis only to those Followers whose Leader *never at any point* delivered a Disagreement State message. There, 61.9% (78/126) of all guesses indicate the Disagreement State, given a Case II Agreement State message.

<sup>19</sup>The corresponding figure over all 15 periods was an infinitesimal 0.0069%.

$$\overline{\rho_1}(m = 1) = \frac{\text{prob}(m=1|\omega=1)\text{prob}(\omega=1)}{\text{prob}(m=1|\omega=1)\text{prob}(\omega=1) + \text{prob}(m=1|\omega=2)\text{prob}(\omega=2)} = \frac{\rho_1}{\rho_1 + 0.772q\rho_2}$$

rather than the equilibrium value  $\frac{\rho_1}{\rho_1 + q\rho_2}$ . Within the context of such a framework, a message  $m = 1$  would be credible as a statement of fact under a wider variety of circumstances than was the case in the original analysis; effectively, the boundary of the Case II region depicted in Figure 1 is pushed upwards, redefining Case II as corresponding only to situations with more extreme high values of  $\rho_2$ . Yet, within this more extreme region, Followers still guess that the Disagreement State is the true state of the world only 62.2% (97/156) of the time, a rate statistically indistinguishable from the full sample ( $p = 0.294$ , two-tailed). A variety of alternative specifications exploring the posterior beliefs that rational Followers might possess given Leader behavior “in the data” all return similar results.

Taken together, these robustness checks add further strength to the conclusion that Followers fail fully to account for Leaders’ strategic incentives in choosing messages a substantial fraction of the time, even after repeated interactions with a given Leader. As a result, Followers’ beliefs about the state of the world are systematically biased away from Bayesian judgments, towards the messages about the world that they receive from Leaders.

Finally, it is worthwhile to note that Followers who receive a Disagreement State message almost always guess the state of the world to be the Disagreement State (94.7% of the time, 89/94).

## 6 Discussion and Conclusion

Collectively, these experimental results offer important insights into the dynamics of leadership and followership. In order to achieve many of their political objectives, leaders must find ways to coordinate the actions of their followers. Noting that members of political groups often learn about the political world through speech or other forms of communication from elites, this paper advanced a novel game-theoretic framework in which leaders can send messages about the state of the world to their followers. Through one mechanism or another, such messages can lead followers to coordinate their actions more successfully; but these messages also can and do influence followers’ beliefs about the state of the world. In the experiment, subjects in the role of leader did routinely choose to misrepresent the state

of the world to followers when it was in their interests to do so, and leaders' strategically-chosen messages about the world were found to be highly effective in coordinating group members' actions. The central finding, however, is that such messages appear strongly to influence followers' beliefs about the state of the world even when Bayesian-rational followers would not find the messages to be credible as statements of fact. Followers appear not fully to account for leaders' strategic incentives to misrepresent the state of the world in forming their posterior beliefs. This finding is suggestive of one mechanism through which factual beliefs about the world may come to be polarized across different groups.

Of course, the external validity of laboratory-experimental results to real-world processes is always a matter of interpretation. Yet, in a number of respects, various features of the laboratory environment would seem if anything to make it *less* rather than more likely that Followers would fail to account for Leaders' strategic incentives in formulating beliefs about the world. Subjects in the experiment were directly paid for the accuracy of their guesses about the state of the world, an immediate incentive for clear thinking in forming beliefs that has no ready analogue in mass politics; indeed, Followers could potentially earn as much from these guesses as from the coordination games themselves. Followers were randomly assigned to groups and interacted anonymously with fellow group members, a context presumably largely devoid of the kinds of emotional relationships that individuals may have with groups of which they are members, or with group leaders, in national, partisan, social-identity-group, or other real-world collective contexts. Subjects were given precise, objective information about the probabilities of different states of the world, a feature of the experimental protocol that made these prior probabilities far more accessible and substantially more salient than in nearly any real-world setting of interest. Clear inferences about the actual state of the world, given Leaders' messages, were doubtless easier to make given these features of the experiment than they would have been if such information had been less salient or more vague. And, finally, Leaders' incentives to dissemble by misrepresenting Disagreement States were quite transparent in the experiment, arguably far more so than in parallel real-world contexts.

The results presented here suggest that, under certain circumstances, Followers may systematically fail fully to account for Leaders’ strategic incentives to misrepresent the state of the world in forming posterior beliefs. This finding differs in tone from some other results in the literature, which suggest that, in other contexts, recipients of communications may be able accurately to account for such incentives in updating their beliefs (e.g., Lupia and McCubbins 1998). In the model presented here, the Leader’s underlying preferences are always at least partially aligned with those of Followers – and are, with some probability, fully aligned. In this context, Followers “overbelieve” a Leader’s messages in a way that receivers did not “overbelieve” senders’ messages when interests were known or believed likely to be strictly opposed in Lupia and McCubbins’ experiments.

The difference between these findings underscores the potential relevance of different features of the strategic and informational environment to the way in which individuals infer meaning from communications in practice, and suggests a variety of avenues for future research. Does the effectiveness of a leader’s communications in influencing follower beliefs and behavior depend systematically on the structure of uncertainty about the leader’s preferences? On the way in which the leader *became* leader in the first place? On the *kind* of group that is under consideration? On other features of the strategic environment, such as the presence or absence of external threats to the group? Such questions are central to the understanding of leadership and followership, and an understanding of the dynamics of groups more generally. Further theoretical and experimental work will be required to arrive at the answers.

## 7 Appendix

### Proof of Proposition 1.

The following notation will be used to specify a Perfect Bayesian Equilibrium in the Leadership and Coordination Game:  $(s_A(\omega = 1), s_A(\omega = 2), s_A(\omega = 3), s_B(\omega = 1), s_B(\omega = 2), s_B(\omega = 3); \sigma_A(m = 1), \sigma_A(m = 2), \sigma_A(m = 3); \sigma_B(m = 1), \sigma_B(m = 2), \sigma_B(m = 3) : \phi_A(m = 1), \phi_A(m = 2), \phi_A(m = 3), \chi_A(m = 1), \chi_A(m = 2), \chi_A(m = 3); \phi_B(m = 1), \phi_B(m = 2), \phi_B(m = 3), \chi_B(m = 1), \chi_B(m = 2), \chi_B(m = 3))$

2),  $\chi_B(m = 3)$ ), where  $s_j(\cdot)$  is a pure action for a Leader of type  $L_j$ ,  $\sigma_k(\cdot)$  is a mixed (possibly degenerate) action for a Follower of type  $F_k$ ,  $\phi_k(\cdot)$  is a vector  $(\overline{\rho_{1,k}(\cdot)}, \overline{\rho_{2,k}(\cdot)}, \overline{\rho_{3,k}(\cdot)})$  specifying the posterior beliefs of a Follower of type  $F_k$  about the state of the world, and  $\chi_k(\cdot)$  is the posterior belief of a Follower of type  $F_k$  that the Leader is of type  $\theta = L_A$ .

Any Perfect Bayesian Equilibrium satisfying Assumption 1 and the conditions for Leadership-Correlated Equilibrium in Definition 2 must satisfy  $\sigma_A(m = 1) = \sigma_B(m = 1) = A$  and  $\sigma_A(m = 3) = \sigma_B(m = 3) = B$  as well as  $\sigma_A(m = 2) = ((a)A, (1 - a)B)$  and  $\sigma_B(m = 2) = ((b)A, (1 - b)B)$ , where  $a$  and  $b$  are the non-degenerate mixing probabilities ensuring that  $\sigma_A(m = 2)$  and  $\sigma_B(m = 2)$  are mutual best responses for  $F_A$  and  $F_B$ , given their posterior beliefs upon receiving message  $m = 2$ .

Given this strategy profile for  $F_A$  and  $F_B$ , it is clearly a strict best response for either type of Leader to send her self-interested message in every state of the world; doing so guarantees the Leader utility 1 while deviating to the other Agreement State message yields utility  $\mu$  and deviating to  $m = 2$  yields utility strictly less than 1 because  $F_A$  and  $F_B$  would play non-degenerate mixed strategies.

If a Perfect Bayesian Equilibrium exists in which each type of Leader sends her self-interested message in every state of the world, by Bayes' Rule Followers' posterior beliefs along the equilibrium path must satisfy  $\phi_A(m = 1) = \phi_B(m = 1) = (\frac{\rho_1}{\rho_1 + \rho_2 q}, \frac{\rho_2 q}{\rho_1 + \rho_2 q}, 0)$  and  $\phi_A(m = 3) = \phi_B(m = 3) = (0, \frac{\rho_2(1-q)}{\rho_3 + \rho_2(1-q)}, \frac{\rho_3}{\rho_3 + \rho_2(1-q)})$  and  $\chi_A(m = 1) = \chi_B(m = 1) = \frac{(\rho_1 + \rho_2)q}{(\rho_1 + \rho_2)q + \rho_1(1-q)}$  and  $\chi_A(m = 3) = \chi_B(m = 3) = \frac{\rho_3 q}{\rho_3 q + (\rho_2 + \rho_3)(1-q)}$ . Denote  $F_A$ 's off-equilibrium path beliefs following receipt of message  $m = 2$  by  $\phi_A(m = 2) = (\psi_{1,A}, \psi_{2,A}, \psi_{3,A})$  and  $\chi_A(m = 2) = \xi_A$ , and  $F_B$ 's off-equilibrium path beliefs by  $\phi_B(m = 2) = (\psi_{1,B}, \psi_{2,B}, \psi_{3,B})$  and  $\chi_B(m = 2) = \xi_B$ . Then the values  $a$  and  $b$  above must satisfy  $a = \frac{\mu\psi_{1,B} + \psi_{2,B} + \psi_{3,B}}{1 + \mu}$  and  $b = \frac{\mu\psi_{1,A} + \mu\psi_{2,A} + \psi_{3,A}}{1 + \mu}$ .

The above can be summarized in the following profile using the above notation:

$$\begin{aligned} & (s_A(\omega = 1) = 1, s_A(\omega = 2) = 1, s_A(\omega = 3) = 3, s_B(\omega = 1) = 1, s_B(\omega = 2) = 3, s_B(\omega = 3) = \\ & 3; \sigma_A(m = 1) = A, \sigma_A(m = 2) = ((a)A, (1 - a)B), \sigma_A(m = 3) = B; \sigma_B(m = 1) = A, \sigma_B(m = 2) = \\ & ((b)A, (1 - b)B), \sigma_B(m = 3) = B : \phi_A(m = 1) = (\frac{\rho_1}{\rho_1 + \rho_2 q}, \frac{\rho_2 q}{\rho_1 + \rho_2 q}, 0), \phi_A(m = 2) = (\psi_{1,A}, \psi_{2,A}, \psi_{3,A}), \phi_A(m = \\ & 3) = (0, \frac{\rho_2(1-q)}{\rho_3 + \rho_2(1-q)}, \frac{\rho_3}{\rho_3 + \rho_2(1-q)}), \chi_A(m = 1) = \frac{(\rho_1 + \rho_2)q}{(\rho_1 + \rho_2)q + \rho_1(1-q)}, \chi_A(m = 2) = \xi_A, \chi_A(m = 3) = \end{aligned}$$

$$\frac{\rho_3 q}{\rho_3 q + (\rho_2 + \rho_3)(1-q)}; \phi_B(m = 1) = \left( \frac{\rho_1}{\rho_1 + \rho_2 q}, \frac{\rho_2 q}{\rho_1 + \rho_2 q}, 0 \right), \phi_B(m = 2) = (\psi_{1,B}, \psi_{2,B}, \psi_{3,B}), \phi_B(m = 3) = \left( 0, \frac{\rho_2(1-q)}{\rho_3 + \rho_2(1-q)}, \frac{\rho_3}{\rho_3 + \rho_2(1-q)} \right), \chi_B(m = 1) = \frac{(\rho_1 + \rho_2)q}{(\rho_1 + \rho_2)q + \rho_1(1-q)}, \chi_B(m = 2) = \xi_B, \chi_B(m = 3) = \frac{\rho_3 q}{\rho_3 q + (\rho_2 + \rho_3)(1-q)}$$

where the  $\xi$ 's and  $\psi$ 's are unconstrained except that  $\psi_{1,j} + \psi_{2,j} + \psi_{3,j} = 1$  for  $j \in \{A, B\}$ ;  $a$  and  $b$  are defined in terms of the  $\psi$ 's as above.

The Leadership and Coordination Game has no proper subgames. In these profiles both Leader types are playing strict best responses in every state of the world; Followers' beliefs are formed using Bayes' Rule on the equilibrium path; Followers have beliefs at their information sets off the equilibrium path; and each Follower is playing a best response, given his beliefs and other players' strategies. Thus, these profiles constitute Perfect Bayesian Equilibria. It was demonstrated above that these profiles also satisfy the other requirements of Definition 2, so they constitute Leadership-Related Equilibria as well. ■

## 8 References

- Achen, Chris and Larry Bartels. 2006. "It Feels Like We're Thinking: The Rationalizing Voter and Electoral Democracy." Princeton University Working Paper.
- Banks, Jeffrey S. and Randall L. Calvert. 1992. "A Battle-of-the-Sexes Game with Incomplete Information." *Games and Economic Behavior* 4: 347-372.
- Bartels, Larry M. 2002. "Beyond the Running Tally: Partisan Bias in Political Perceptions." *Political Behavior* 24(2): 117-150.
- Berelson, Bernard, Paul F. Lazarsfeld, and William N. McPhee. 1954. *Voting: A Study of Opinion Formation in a Presidential Campaign*. Chicago: University of Chicago Press.
- Calvert, Randall L. 1992. "Leadership and Its Basis in Problems of Social Coordination." *International Political Science Review* 13(1): 7-24.
- Calvert, Randall L. 1995. "The Rational Choice of Social Institutions: Cooperation, Coordination, and Communication," in *Modern Political Economy*, ed. by J.S. Banks and E.A. Hanushek. Cambridge UK: Cambridge University Press.

- Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press.
- Campbell, Angus, Philip Converse, Warren Miller, and Donald Stokes. 1960. *The American Voter*. Wiley: New York.
- Canes-Wrone, Brandice, Michael Herron, and Ken Shotts. 2001. "Leadership and Pandering: A Theory of Executive Policy-Making." *American Journal of Political Science* 45(3): 532-550.
- Dewan, Torun and David P. Myatt. 2007. "Leading the Party: Coordination, Direction and Communication." *American Political Science Review*.
- Dewan, Torun and David P. Myatt. 2008. "Qualities of Leadership: Communication, Direction and Obfuscation." *American Political Science Review*.
- Fischbacher, Urs. 1999. "z-Tree Zurich Toolbox for Readymade Economic Experiments – Experimenters Manual." Working Paper Nr. 21, Institute for Empirical Research in Economics, University of Zurich.
- Levi, Margaret. 2006. "Why We Need a New Theory of Government," *Perspectives on Politics* 4(1): 5-19.
- Lupia, Arthur and Mathew D. McCubbins. 1998. *The Democratic Dilemma*. Cambridge University Press: Cambridge UK.
- Schelling, Thomas C. 1960. *The Strategy of Conflict*. Harvard University Press: Cambridge, MA.
- Wilson, Rick K. and Carl M. Rhodes. 1997. "Leadership and Credibility in N-Person Coordination Games." *Journal of Conflict Resolution* 41(6): 767-791.
- Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. Cambridge University Press: Cambridge UK.

**Table 1. Coordination Rates: by Case.**

	Coordination Rate: Message Indicates Agreement State	Coordination Rate: Message Indicates Disagreement State
Overall	93.8% (405/432)	52.1% (25/48)
Case I	95.5% (190/199)	45.5% (5/11)
Case II	90.0% (99/110)	52.4% (11/21)
Cases III and IV	-	53.3% (8/15)
(When Message Indicates Likelier Agreement State)	97.8% (90/92)	-
(When Message Indicates Less-Likely Agreement State)	82.1% (23/28)	-

**Table 2. Leaders' Messages: by Case.**

**Case I.**

In an Agreement State

In a Disagreement State

preferred (and true) Agreement State	other Agreement State	Disagreement State	preferred Agreement State	other Agreement State	(true) Disagreement State
94.3% (165/175)	4.6% (8/175)	1.1% (2/175)	60.0% (21/35)	14.3% (5/35)	25.7% (9/35)

**Case II.**

In an Agreement State

In a Disagreement State

preferred (and true) Agreement State	other Agreement State	Disagreement State	preferred Agreement State	other Agreement State	(true) Disagreement State
88.6% (31/35)	11.4% (4/35)	0% (0/35)	64.6% (62/96)	13.5% (13/96)	21.9% (21/96)

**Cases III and IV.**

In the Likelier Agreement State

In a Disagreement State when pref'd A.S. is likelier

preferred (and true) Agreement State	other Agreement State	Disagreement State	preferred Agreement State	other Agreement State	(true) Disagreement State
94.4% (67/71)	2.8% (2/71)	2.8% (2/71)	71.4% (15/21)	9.5% (2/21)	19.0% (4/21)

In the Less-Likely Agreement State

In a Disagreement State when pref'd A.S. is less-likely

preferred (and true) Agreement State	other Agreement State	Disagreement State	preferred Agreement State	other Agreement State	(true) Disagreement State
100% (9/9)	0% (0/0)	0% (0/0)	44.1% (15/34)	29.4% (10/34)	26.5% (9/34)

**Table 3. Followers' Guesses About the State of the World: by Case.**

**Case I.**

Receive an Agreement State Message			Receive a Disagreement State Message	
Guess that Agreement State	Guess other Agreement State	Guess the Disagreement State	Guess an Agreement State	Guess the Disagreement State
90.2% (359/398)	5.0% (20/398)	4.8% (19/398)	13.6% (3/22)	86.4% (19/22)

**Case II.**

Receive an Agreement State Message			Receive a Disagreement State Message	
Guess that Agreement State	Guess other Agreement State	Guess the Disagreement State	Guess an Agreement State	Guess the Disagreement State
42.7% (94/220)	0.5% (1/220)	56.8% (125/220)	2.4% (1/42)	97.6% (41/42)

**Cases III and IV.**

Receive Likelier Agreement State Message			Receive Disagreement State Message	
Guess that Agreement State	Guess other Agreement State	Guess the Disagreement State	Guess an Agreement State	Guess the Disagreement State
86.4% (159/184)	0.0% (0/184)	13.6% (25/184)	3.3% (1/30)	96.7% (29/30)

Receive Less-Likely Agreement State Message

Guess that Agreement State	Guess other Agreement State	Guess the Disagreement State
44.6% (25/56)	19.6% (11/56)	35.7% (20/56)

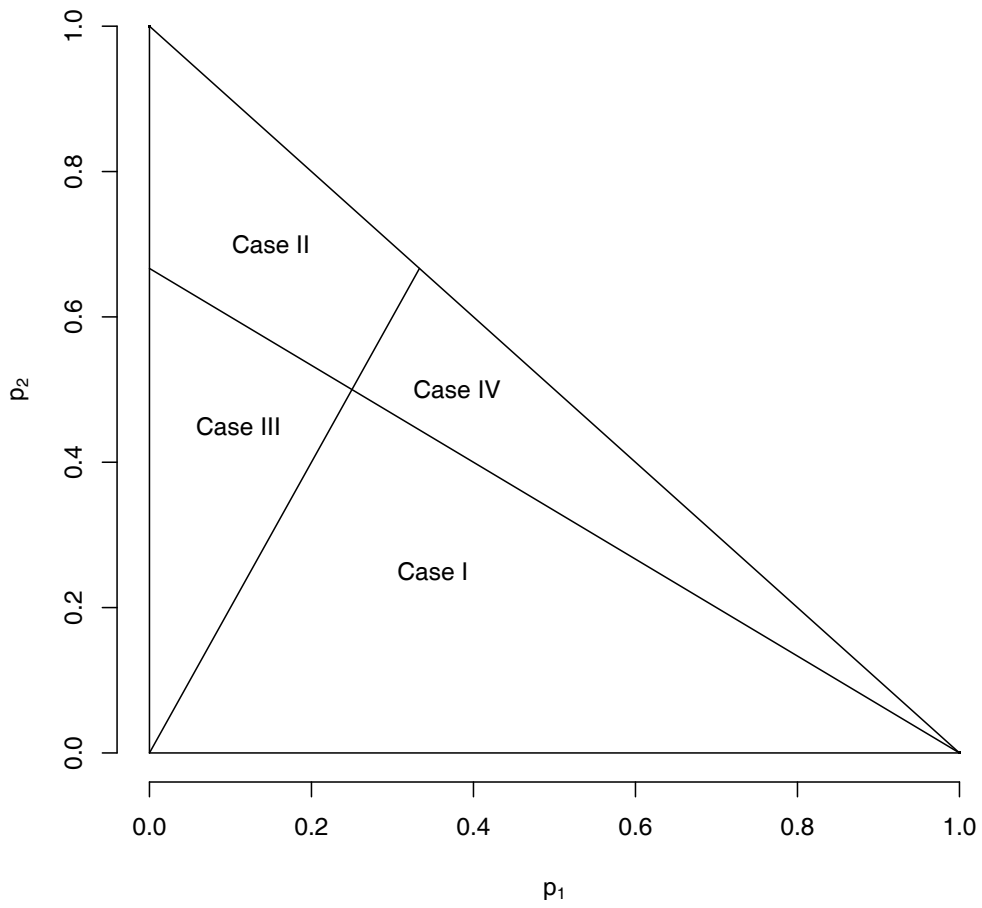


Figure 1: Graphical Depiction of the Conditions for Cases I-IV on the Probability Simplex, for  $q = 0.5$ .