

# Competition, Preference Uncertainty, and Jamming: A Strategic Communication Experiment\*

William Minozzi<sup>†</sup>     Jonathan Woon<sup>‡</sup>

May 22, 2012

## Abstract

Modern democracies are built on an edifice of competition and information asymmetry, yet citizens must remain uncertain about the preferences of those they rely upon. But competition and preference uncertainty enable the informed to jam, impeding communication. We describe a game-theoretic laboratory experiment in which subjects play an information transmission game with two senders who have private information about their preferences. Although we find support for many equilibrium predictions, we also observe that senders “overjam,” exaggerating even when they are predicted to tell the truth. Overjamming can be explained by a bounded rationality framework in which senders choose messages based on experience within the experiment, exaggerating more when they have observed their opponents exaggerate more previously. Our results suggest that jamming and the resulting polarization of competing messages may be persistent features of political discourse featuring competition and preference uncertainty.

---

\*We gratefully acknowledge comments and advice from Mark Fey, Sean Gailmard, Becky Morton, Michael Neblo, Laura Paler, John Patty, Maria Petrova, Craig Volden, Rick Wilson, and seminar audiences at Duke University and Washington University in St. Louis. Previous versions of this paper were presented at the Experiments Mini-Conference at the 2011 Southern Political Science Association meeting, the 2011 Midwest Political Science Association Meeting, and the 2011 American Political Science Association Meeting.

<sup>†</sup>Assistant Professor, Department of Political Science, 2189 Derby Hall, Ohio State University, Columbus, OH 43210 Phone: 614-247-7017, Email: minozzi.1@osu.edu

<sup>‡</sup>Assistant Professor, Department of Political Science, 4814 Wesley W. Posvar Hall, University of Pittsburgh, Pittsburgh, PA 15260 Phone: 412-648-7266, Email: woon@pitt.edu

Competition permeates democracies. Candidates for high office offer competing visions for national policy, lobbyists and legislators construct arguments for and against legislation, and adversarial courts rely on opposing advocates to inform their decisions. Although these highly specialized institutions differ significantly in their details, each involves informed and interested parties pressing rival ideas on key decisionmakers. But no matter how well known these informed parties are, each retains a penumbra of privacy about their fundamental motives, leaving everyone else uncertain about what they really want or believe.

Our understanding of competitive political communication is limited, despite its seeming ubiquity. In the abstract, *competition* between information providers has the potential to resolve controversy by encouraging the truth, but it also creates strategic opportunities to obfuscate (Minozzi 2011). Although an information provider might be counted on to tell the truth when it suits her purposes, another for whom the truth has unpalatable consequences might instead *jam* her opponent's message, leaving decisionmakers uncertain as to who told the truth. Because the informed parties retain private information about their own preferences, key decisionmakers will not be able to infer the truth from competing claims. Thus, competition and preference uncertainty combine to leave decisionmakers without the means to make informed judgments.

If competition—a linchpin of democracy—impedes the communication of information to those who need it, the implications for democratic governance will be significant. On the institutional side, legislatures are organized around delegating information-intensive tasks to committees (Kiewiet and McCubbins 1991), but these committees may be composed of opposing interests (Gilligan and Krehbiel 1989). Legislators design bureaucracies in part to generate useful information (Bawn 1995; Bendor, Taylor, and Van Gaalen 1985), but they delegate tasks to many overlapping agents (Landau 1969). In hearings, legislators frequently rely on experts (Esterling 2004), but these experts are often counterbalanced on opposing sides of an issue. Courts depend on advocates to articulate legal rationales and fact patterns (Bailey, Kamoie, and Maltzman 2005), but opposing sides of an issue are represented. An

informed public plays a key role in the legitimacy of judicial institutions (Carrubba 2009; Staton 2006; Vanberg 2001), but legitimacy can be contested and threatened by competing interests. Information cues can help voters learn from and be persuaded by elites (Lupia and McCubbins 1998), public opinion develops amidst a cacophony of conflicting messages (Zaller 1992). And representative government itself depends on a division of intellectual labor between office-holders and citizens, but competitive elections are the chief mechanism of accountability (Canes-Wrone, Herron, and Shotts 2001; Fox and Shotts 2009). Each of these environments is characterized by communicative competition.

Some argue that competition in the “marketplace of ideas” facilitates the communication of information. Indeed, more heterogeneous committees are thought to yield more credible information (Gilligan and Krehbiel 1989; Krehbiel 1991). Competition among lobbyists may lead to better decisions by legislators (Austen-Smith and Wright 1992), and competition among elites may engender a more fully informed public (Page and Shapiro 1992). But such competition may simply offer more chances to jam rather than to facilitate learning. Thus, it is crucial to learn how people behave when faced with this strategic situation.

We conduct experiments to investigate whether messages and decisions arising in a strategic environment with both competition and preference uncertainty conform to game theoretic predictions. Equilibrium analysis identifies two key properties of communication in this environment. First, truth-telling and jamming are conditional on the truth and senders’ privately known biases. Second, jamming messages are countervailing—inversely related to the truth. Together, these properties imply that equilibrium messages will be conditionally polarized. Our findings lend support to many of the equilibrium predictions, including a set of seemingly counterintuitive comparative static hypotheses, but we also identify key points of departure.

Surprisingly, senders in our competitive communication environment “overjam.” That is, senders who are predicted to tell the truth instead send messages that are false. This behavior results in messages that, in contrast to the equilibrium theory, are consistently polarized. Our analysis of the etiology of overjamming suggests that bounded rationality and limited

strategic sophistication play important parts. We find that overjamming is best explained by a framework in which subjects play best responses based on their experience within the experiment. Senders seem to counteract the exaggerations they would expect their opponents to make, exaggerating more when they have observed their opponents exaggerate more previously. Although behavior responds to strategic incentives, it is also boundedly rational because senders always have an incentive to exaggerate further, which is not consistent with any game theoretic equilibrium. Moreover, individuals exaggerate less than they would if they were to maximize payoffs or if they had correctly anticipated competing senders' messages.

Our experiment is related to two broad literatures on communication. A large body of experimental work in political psychology explores how and to what extent subjects' decisions can be manipulated via agenda setting, priming, and framing (e.g., Chong and Druckman 2007; Iyengar and Kinder 1987; Nelson, Clawson, and Oxley 1997). Although we have learned a great deal from this work, it cannot tell us how would-be manipulators might behave. This omission is surprising, not least because the terms themselves—agenda setting, priming, and framing—imply that someone is engaged in strategic action. One reason for this omission might be that psychological experiments are not typically designed to explore the extent of strategic political behavior.

In contrast, experimental economists have thoroughly studied strategic behavior in Crawford and Sobel's (1982) seminal cheap talk, sender-receiver game (e.g., Blume et al. 1998, 2001; Cai and Wang 2006; Crawford 1998; Dickhaut, McCabe, and Mukherji 1995; Gneezy 2005; Hurkens and Kartik 2009). And in political science, Lupia and McCubbins (1998) conduct experiments based on similar games in which the decisionmaker is uncertain about a (single) informed player's preferences. But none of these studies focuses on communicative competition with uncertainty about the preferences of the information providers, and our finding of overjamming stands in contrast to the finding that senders overcommunicate.<sup>1</sup>

---

<sup>1</sup> Boudreau and McCubbins (2008) have their subjects solve math problems with the help of “experts”—other subjects who have access the correct answers. These experts are privately informed of whether they

## Theory and Hypotheses

The strategic situation faced by competitive political elites and experts is far different from that facing an information provider in a non-competitive environment. Competition among information providers allows receivers to compare and contrast the messages they receive, but it also breeds novel opportunities for senders to obfuscate. These opportunities can change dramatically from setting to setting; arguments before the Supreme Court can look far different from presidential debates. However, the strategic aspects of communicative competition can be boiled down to their most elemental form using tools from game theory.

Consider the simplest political environment in which there can be communicative competition: one with two senders and one receiver.<sup>2</sup> At the outset, both senders observe a *target*  $T$ , which functions as the “truth” in the game. Each sender  $i$  also privately knows his *shift*  $S_i$ , which represents the direction and degree of preference divergence between  $i$  and the receiver.<sup>3</sup> The senders then simultaneously select messages  $m_i$  to send to the receiver, who then chooses an action  $c$ . The receiver prefers  $c$  to be as close as possible to  $T$  while each sender  $i$  prefers that  $c$  be as close as possible to  $T + S_i$ . In terms familiar from the spatial model of politics,  $T$  is the receiver’s ideal point while  $T + S_i$  is sender  $i$ ’s ideal point. In our discussion, we designate one sender as the *left sender* and one sender as the *right sender* (with shifts  $S_L$  and  $S_R$  and respective ideal points that are always to the left and right of the receiver’s). Importantly, the receiver knows that the senders are *opposed* but is uncertain whose ideal point is *closer* to hers. In the *Symmetric Baseline Condition*, both shifts are uniformly distributed.

---

will benefit if the problem-solver gets the answer right, or if she gets it wrong, and then they offer her suggested answers. While our setting shares some limited features with theirs, ours affords a much richer information structure and set of messages and, therefore, much more varied communication strategies. And our setting is spatial, meaning that our findings have straightforward applications to many fundamental formal models of politics.

<sup>2</sup> For a detailed formal analysis of this game, see Minozzi (2011). In our exposition, we will use male pronouns to refer to senders and female pronouns to refer to receiver.

<sup>3</sup> The distributions of  $T$ ,  $S_L$ , and  $S_R$  are common knowledge, and we present specific parameterizations when discussing the experimental procedures.

Equilibria of this game involve messages that are either *truthful* or that *jam* truthful messages, and in equilibrium, the receiver learns exactly what the target is if and only if both senders send truthful messages.<sup>4</sup> If the messages agree, the receiver infers the truth  $T$  and responds by choosing  $c = T$ . However, if the messages disagree, the receiver understands that at least one sender must have jammed, and she therefore cannot infer the true target. In that case, the receiver’s equilibrium response is to choose a default action  $c_0 = 0$ , which is the same as the receiver’s optimal action based only on her prior beliefs.<sup>5</sup>

Receivers’ message strategies specify both *when* and *how* to lie. If the receiver uses the strategy just described, each sender recognizes that the only possible equilibrium path actions are  $c = T$  and  $c = 0$ . Truth-telling might therefore lead the receiver to choose  $c = T$  (if the other sender also tells the truth), but jamming will always lead to conflicting messages and to the action  $c = 0$ . Thus, sender  $i$  prefers to tell the truth when her ideal point is closer to  $T$  than to 0. In contrast,  $i$  prefers to jam when her ideal point is closer to  $c_0$ . It follows from this calculation that each sender has a *jamming region*—a set of targets for which the sender will jam—and that the size and location of these jamming regions depend on the sender’s shift.<sup>6</sup> Our first hypothesis summarizes the equilibrium prediction that jamming and truth-telling are conditional and describes the properties of the jamming regions.

**Conditional Jamming Hypothesis.** *When the target is in a sender’s jamming region, she will send the jamming message; otherwise, she will tell the truth. A jamming region will be on the side of the default action opposite the sender’s shift, does not include extreme targets, and its size is increasing in the magnitude of the sender’s shift.*

When a sender jams, he must lie in such a way to ensure that when the messages disagree,

---

<sup>4</sup> “Equilibrium” refers to perfect Bayesian equilibrium with beliefs as specified in Minozzi (2011).

<sup>5</sup> Technically, other default actions can support other jamming equilibria, but  $c_0 = 0$  is a natural focal point that is supported by very simple off-the-path beliefs (e.g., where the receiver ignores both messages). Alternatively,  $c_0 = 0$  is the optimal action in a perturbed game where there is a small probability that both messages are purely random; in this case every possible message pair occurs on the path of play and Bayes’ Rule always applies.

<sup>6</sup> Formally, the left sender’s jamming region is the interval  $[0, 2|S_L|]$  and the right sender’s jamming region is the interval  $[-2|S_R|, 0]$ .

the receiver will in fact choose the default  $c = 0$ . To determine what this message must be, first note that the jamming regions described in the Conditional Jamming Hypothesis imply that at least one of the senders will always send a truthful message  $m = T$ . Thus, when the receiver observes conflicting messages, she knows that one of them must be the true state. A sender who jams therefore knows that his jamming message might be (falsely) interpreted as the true state and will exploit the receiver’s uncertainty about senders’ preferences to ensure that the receiver remains completely uncertain about which of the senders actually jammed. This logic implies that jamming messages must be *countervailing*—it must be on the opposite side of the default as the true target and the more extreme the target, the more extreme the jamming message.<sup>7</sup> Our second hypothesis summarizes the characteristics of jamming messages.

**Countervailing Jamming Messages Hypothesis.** *Jamming messages will be on the side of the default opposite the target, and as the target becomes more extreme, the jamming message will become more extreme as well.*

Given that these message strategies specify conditional truth-telling and countervailing jamming, the receiver’s strategy outlined above is indeed a best response. If messages agree, she knows that both senders told the truth. If the receiver observes messages that disagree, she does not know who has jammed and who has revealed the target because of the way that the jamming message is designed. The next hypothesis summarizes the theoretical prediction for receivers’ actions.

**Receiver Actions Hypothesis.** *When messages agree, the receiver will choose an action equal to the target; otherwise, the receiver will choose the default action.*

---

<sup>7</sup> In the Symmetric Baseline Condition, the exact jamming message as a function of the target is  $m_J(T) = -T$ . To see more precisely how the receiver remains uncertain, suppose the true state is  $T$  so that if the left sender jams, the jamming message is  $-T$  and the message pair is  $(-T, T)$ . But this is the same message pair that would be observed if the true state were  $-T$  and the right sender sends the jamming message  $-(-T) = T$ . Because both shift distributions are uniformly distributed, the sender infers that the true state is equally likely to be  $T$  as it is to be  $-T$ . As a consequence, the optimal action is the midpoint  $c = 0$ , which is precisely the default action.

Different default actions can support different equilibria, meaning that multiple equilibria may pose a problem for the experimental analysis of this game.<sup>8</sup> Nevertheless, there are several reasons to study this game. Most interesting sender-receiver games feature multiple equilibria, meaning that any study of strategic communication in a sufficiently rich setting is likely to face this problem. For example, the seminal Crawford-Sobel game, which has been applied to many areas of political science (Gilligan and Krehbiel 1987) and is well-studied experimentally (Cai and Wang 2006), has a range of partition equilibria. If we are to study any sort of communicative competition, the problem of multiple equilibria is likely to recur unless the message space is limited to no more than two possible messages. Furthermore, there are good reasons to focus on this equilibrium. First, the set of equilibria we study is structurally identical in every respect except for the default action. That is, most of our empirical predictions are the same regardless of the default action. Second, the default action we concentrate on,  $c_0 = 0$ , is a natural focal point. Not only is it the prior expected value of the target, it is the median message and the median action. Finally, the presence of these multiple equilibria biases against finding evidence of our empirical predictions, which makes any evidence we do find more compelling.

In addition to the equilibrium point predictions, we test several comparative static hypotheses. For example, the receiver may believe that senders' messages include small errors, in which case messages are unlikely to ever agree. In this case, the receiver will treat messages that are closer to each other as more indicative of the underlying target than messages that are far from each other. Thus, the receiver should be better able to guess the target when the difference between messages is small.

**Message Difference Hypothesis.** *In comparative terms, the receiver should be less able to guess the target when the difference between the senders' messages is larger.*

---

<sup>8</sup> There is also one more class of equilibria we do not consider: the babbling-like equilibria in which all players effectively ignore some subset of targets. But in these equilibria, senders must coordinate on a particular subset of targets to ignore, and this coordination seems to be a difficult strategic task in its own right, so we do not consider this possibility.



We also test a set of seemingly counterintuitive comparative static predictions about how sender behavior changes when we relax symmetry. To that end, we posit an *Asymmetric Condition* in which the left sender has a shift that is more likely to be further from 0, while the right sender’s shift distribution is unchanged. In this Asymmetric Condition, we call the left sender the *extremist*, and the right sender the *moderate*.<sup>9</sup>

Perhaps surprisingly, predicted messages are not only asymmetric, but the extremist is also predicted to send more moderate jamming messages in equilibrium. The reason is that jamming messages must still ensure that the receiver chooses the default when messages disagree. To understand why, suppose instead that messages are symmetric reflections of each other (as in the equilibrium of the Symmetric Baseline Condition). If the receiver believes the extremist is more likely to have lied than the moderate, instead of choosing the default she will choose an action that is on the Moderate’s side of the default (and closer to the Moderate’s ideal point). But then to offset this change in equilibrium, the extremist must moderate his messages. In turn, the moderate’s jamming messages must become more extreme, and the extent of the extremist’s moderation and the moderate’s extremity will satisfy the usual equilibrium requirements. Importantly, this change should occur even though the right sender has the same shift distribution in both conditions. It is the direct result of strategic interaction as each sender’s strategy depends as much on the opposing sender’s characteristics as his own.

**Extremist Moderation Hypothesis.** *The Extremist will send more moderate jamming messages in the Asymmetric Condition than in the Symmetric Baseline Condition.*

**Moderate Extremism Hypothesis.** *The Moderate will send more extreme jamming messages in the Asymmetric Condition than in the Symmetric Baseline Condition.*

These hypotheses offer a well-defined set of expectations for the behavior of senders and receivers in communicative competition. However, to accurately and credibly test them, we

---

<sup>9</sup> We formally define a sender to be more *moderate* if the distribution of  $|S_{Moderate}|$  is first-order stochastically dominated by the distribution of  $|S_{Extremist}|$  for the other sender.

need to know not only what senders say, but also what their true preferences are as well as what they believe to be true. This is extremely difficult using observational data, and, thus, we turn to the lab.

## Experimental Procedures

We conducted our experiments at the Pittsburgh Experimental Economics Laboratory using subjects recruited through the lab's centralized database. Most subjects were undergraduates at the University of Pittsburgh or Carnegie Mellon University, and no subjects were recruited from the authors' classes. Each subject participated in only one session.

Upon arriving at the lab, subjects gave informed consent and were seated at separate computer terminals. All interactions between subjects took place anonymously through the networked computers using software programmed and conducted using z-tree (Fischbacher 2007). Subjects received strict instructions not to communicate with one another in any way throughout the session. The instructions were presented on their computer screens and read aloud in an effort to induce common knowledge among the participants. Subjects received printed copies of the instructions, to which they were encouraged to refer as often as they needed, and were given a quiz about the instructions in order to ensure comprehension. The quizzes were administered through the computers so that subjects privately received immediate feedback about whether or not they answered questions correctly and explanations of the correct answers. Consistent with the lab's governance policy, no deception or false feedback was used in the experiment.

After the instructions and quiz, the software randomly assigned subjects to one of the roles in the game: A (left sender), B (right sender), or C (receiver). The instructions only referred to the roles as "A," "B," or "C" and made no reference to "senders," "receivers," "left," "right," "moderate," or "extremist." In our presentation and discussion, however, we continue to use these terms. Subjects proceeded to play between 24 and 32 rounds of the

game (depending on the session and condition), with fixed roles throughout the session.<sup>10</sup>

At the beginning of every round, subjects were randomly matched into groups of three, with one subject in each role in each group. Groups were selected with replacement so that it was possible to be matched with the same group in different rounds. To preclude reputation effects, subjects never knew the ID numbers of the other subjects in their group.

The targets  $T$  and shifts  $S_L$  and  $S_R$  were then drawn independently for each group. In all conditions of the experiment,  $T$  was drawn uniformly from the integers between  $-100$  and  $100$ , and the right sender's shift  $S_R$  was drawn uniformly from integers between  $0$  and  $50$ . The distribution of the left sender's shift  $S_L$  varied across treatment conditions. In the *Symmetric Baseline Condition*,  $S_L$  was drawn uniformly from integers between  $-50$  and  $0$ ; in the *Asymmetric Condition*, integers from  $-50$  to  $-25$  were three times as likely as integers from  $-24$  to  $0$  (with each element of each region equally likely). In the instructions and throughout the experiment, we referred to each player's ideal action as a "target." That is,  $T$  is referred to as "C's target,"  $T + S_L$  is "A's target," and  $T + S_R$  is "B's target."<sup>11</sup> Our use of nearly continuous distributions is in contrast with previous experiments on cheap talk games that typically involve a small state and action space (4 states in Dickhaut, McCabe, and Mukherji (1995) and 5 states in Cai and Wang (2006)). Without this innovation, we could not have accurately conveyed the notion of the spatial model to the subjects.

[Figure 1 about here.]

The experimental interface we used presents information to subjects textually as well as graphically (see Figure 1). The graphical display intuitively conveys the notion of spatial distance inherent in the utility functions.<sup>12</sup> We reasoned that this would allow subjects to focus their cognitive resources on thinking strategically rather than on computing payoffs.

---

<sup>10</sup>Table A-1 in the Appendix provides summary statistics on the sessions.

<sup>11</sup>In our presentation, we continue to refer to "targets" and "shifts." When we do so, the "target" is understood to be C's target.

<sup>12</sup>We thank Stephen Haptonstahl for providing the initial z-Tree code to produce the slider in Haptonstahl (2009), which we modified substantially.

Although the instructions describe the set of targets and shifts as integers, our visual display reinforces the notion that the distributions are to be treated as continuous and spatial.

In every round, each sender simultaneously observed the receiver’s target and his own target (but not the other sender’s target), and then chose a message. As shown in Figure 1, possible messages and actions are displayed on a horizontal axis. To send a message, senders use the mouse to drag a slider along this axis to a position that corresponds to the desired message (any position between  $-150$  and  $150$ ). The interface also displays the range of possible targets for the receiver, the realized target, the range of possible targets for the sender and for the opposing sender, and the sender’s own target, all of which is also presented textually at the top of the screen. The sender’s interface also features a payoff calculator (manipulated via a separate slider) that shows the sender’s and receiver’s payoffs for each possible action the receiver might choose. The receiver observed messages simultaneously after both senders had finished, and the interface displayed this information both graphically and textually. The receiver then dragged a slider to select an action (any position between  $-150$  and  $150$ ).

At the end of every round, subjects were informed of all of the results from the round for their group: both messages, the action, every player’s target, and every player’s payoff. Subjects also observed the results from all previous rounds they played, but they never observed the results for groups to which they did not belong. Payoffs for each round were denominated in “points,” with 100 points being the maximum possible points a player could earn in a round (if the receiver’s action matched their own target exactly). In terms of points, the receiver’s payoff function was  $100 - |c - T|$  and a sender’s payoff function was  $100 - |c - (T + S_i)|$ .

At the end of the experimental session, total points were converted to cash at the rate of \$1 for every 150 points. Subjects were paid the sum of their earnings plus a \$7 participation payment. We conducted four sessions of the Symmetric Baseline Condition (69 subjects) and two of the Asymmetric Condition (27 subjects). Each session involved between 12 and

18 subjects (4 to 6 groups), and each subject participated in only one session.

## Testing Equilibrium Predictions

**Sender Messages.** As a first pass through the data, Figure 2 provides a visual depiction of sender behavior in the Symmetric Baseline Condition by plotting Messages against Targets. For our baseline parameterization of the game, the Conditional Jamming and Countervailing Jamming Messages Hypotheses together imply that messages will hew to one of the two lines. When Targets are outside of a sender’s jamming region (which depends on the sender’s Shift, denoted by filled markers), Messages should fall along the up-sloping, unbroken line representing truthful messages. In contrast, when Targets are inside a sender’s jamming region (denoted by empty markers), Messages should fall along the down-sloping, dashed line representing countervailing jamming messages.

[Figure 2 about here.]

The simple visual test suggests that senders exhibit some of the behaviors predicted by equilibrium analysis, but significant differences emerge between the theory and the evidence. We find that Messages predicted to be truthful are positively related to Targets, but very few such Messages actually equal the truth. Many of these Messages are outside the bounds of the possible Targets, and these should clearly be interpreted as lies. Furthermore, there are few Messages that are even close to being truthful while there is a substantial, systematic divergence between left senders’ (dark circles) and right senders’ (gray diamonds) Messages not predicted by theory. The relationship between Targets and Messages predicted to be countervailing is less clear. While the bulk of Messages for Targets in the jamming regions fall in the upper-left and lower-right quadrants of the Figure 2, seemingly consistent with countervailing jamming behavior, it is possible that the inverse relationship is due instead to the divergence between left senders’ and right senders’ Messages.

To estimate these relationships more carefully, we fit a series of multilevel regression models with each sender’s *Message* as the dependent variable. We regress Message on

the receiver’s *Target*; an indicator *Jam*, which equals 1 if the target lies in the sender’s jamming region; the interaction  $Target \times Jam$ ; and a separate intercept for *Left Sender* and *Right Sender*. Our multilevel models also control for subject-, round-, and condition-level heterogeneity. We rescale Targets, Messages, and Actions, dividing each by 100, so that coefficients are all on the same scale with our indicator variables, and we report two-tailed tests when testing point predictions and one-tailed tests for comparative statics.<sup>13</sup>

[Figure 3 about here.]

Equilibrium analysis gives precise predictions about coefficients in the regression models. The Conditional Jamming Hypothesis predicts that when Targets are outside a sender’s jamming region ( $Jam = 0$ ), senders will tell the truth regardless of which side of the receiver’s ideal point their own ideal point is on; thus, the coefficient on Target should be 1 while the coefficients of Left Sender and Right Sender should both be 0. The Countervailing Jamming Messages Hypothesis predicts that when Targets are inside a sender’s jamming region ( $Jam = 1$ ), senders will jam so that Messages are inversely related to Targets; this implies a coefficient of  $-2$  on the interaction  $Target \times Jam$  (so that the sum of the main coefficient on Target and the interaction is  $-1$ ) and a coefficient on Jam itself that is 0.

The evidence supporting these hypotheses is mixed. The first column (Baseline) of the “Sender Messages” panel in Figure 3 shows the coefficients and standard error bars from this model, against the hypothesis tests represented by dashed lines.<sup>14</sup> Starting from the top, the coefficient on Target, 0.92, is not far from the predicted value of 1 (although we can reject the hypothesis that it equals 1,  $p < 0.01$ , two-tailed). The positive relationship between Messages and Targets outside of jamming regions is broadly consistent with part of the Conditional Jamming Hypothesis. However, the second panel of the first column in Figure 3 shows that the coefficient on the interaction  $Target \times Jam$ ,  $-0.43$ , is far from the hypothesized

<sup>13</sup>The Appendix includes tables for all results we discuss.

<sup>14</sup>See Model [3] from Appendix Table A-2 for details. The model also contains the uninteracted variable Jam, which is not significant.

value of  $-2$  ( $p < 0.001$ , two-tailed). Thus, we can clearly reject the strong form of the Countervailing Jamming Messages Hypothesis because sender messages are not inversely related to Targets inside the jamming regions. Nevertheless, the coefficient is negative and statistically significant, which means that the relationship between Messages and Targets is weaker within the jamming region. Thus, to some extent, messages are conditional on the values of Targets relative to senders' Shifts in a manner broadly consistent with the Conditional Jamming Hypothesis. Finally, the bottom two panels of the first column show that, in contrast to the predictions, the Left Sender intercept is significantly less than 0 ( $p < 0.01$ , two-tailed) meaning that their Messages are consistently to the left of the Target while the Right Sender intercept is significantly greater than 0 ( $p < 0.01$ , two-tailed) meaning that their Messages are consistently to the right of the Target. Thus, there is consistent—rather than conditional—polarization between senders' messages; senders seem to “overjam” by exaggerating even when they are predicted to tell the truth. This finding stands in stark contrast to the “overcommunication” commonly observed in non-competitive strategic information transmission experiments.

To investigate whether subjects may learn how to play the equilibrium as they gain experience, we extend the regression to include interactions with a variable for the *Round* of play, effectively allowing the coefficients to vary over time.<sup>15</sup> The second column (Learning) in the Sender Messages component of Figure 3 presents these results. Up-pointing triangles indicate results for the first round of play, and down-pointing triangles indicate results for the last. If subjects' behavior converges to equilibrium play over time, then we expect that the down-pointing triangles should be closer than the up-pointing triangles to their predicted values at the dashed lines.

We find no evidence that senders learn to play equilibrium strategies over time. In the first round, the coefficient on Target is indistinguishable from its equilibrium point prediction of 1, but then then it decreases significantly by the last round. There is negligible movement

---

<sup>15</sup>Round has been rescaled to lie between 0 and 1. See Model [4] from Appendix Table A-2 for details.

in the coefficient on  $\text{Target} \times \text{Jam}$ , which remains far from its predicted value of  $-2$ . Finally, the Left and Right Sender intercepts start closer to their predicted values of 0, and then move significantly away. Thus, over time, senders seem to overjam even more.<sup>16</sup>

Overall, intriguing patterns emerge from this analysis. There is a fairly tight relationship between messages and targets, even when the latter lie outside the jamming region. While jamming messages are decidedly *not* countervailing in practice, senders do temper the correlation between their messages and targets when the latter lie within the jamming region. But senders also seem to overjam by adding large plenary shifts to all their messages, irrespective of the jamming region, and these shifts increase over time.

**Receiver Actions.** Turning to receiver behavior, we find very limited support for the lone equilibrium point prediction. Receivers are predicted to choose exactly the default  $c_0 = 0$  only whenever senders’ messages diverge, however minimally. Using the receivers’ *Action* as the dependent variable, we can quickly reject this point prediction—in the 640 rounds played in our baseline sessions, senders *never* sent identical messages, yet receivers chose  $c = 0$  in only 19 cases!

We test the comparative static prediction in the Message Difference Hypothesis by regressing Action on *Message Difference*, the absolute difference between the senders’ messages, the interaction  $\text{Target} \times \text{Message Difference}$ , and an intercept. Here, we expect the coefficient on  $\text{Target} \times \text{Message Difference}$  to be negative. That is, as the distance between senders’ messages grows, the receiver should believe it to be less likely that mere errors caused the discrepancy and more likely to believe that one or both sender has jammed.

The first column of the “Receiver Actions” panel in Figure 3 displays the results (without

---

<sup>16</sup>We also conducted sessions of a High Payment Condition to test whether increasing the stakes of the game might stimulate more careful thought, thereby leading to behavior closer to the equilibrium predictions. In the High Payment Condition, the game remained identical except for the fact that we increased the cash value of the points in the game and reduced the number of rounds played to keep total payments in the same range as in the Baseline. We found very little difference between the High Payment Condition and the Symmetric Baseline Condition, which suggests that raising the stakes does not explain “overjamming.” For details of this analyses, see the Web Appendix.



interactions with Round).<sup>17</sup> As predicted, we find that the coefficient on Target  $\times$  Message Difference is negative and statistically significant ( $p < 0.01$ , one-tailed), meaning that the relationship between Actions and Targets is weaker when senders’ messages are farther apart.<sup>18</sup> The final column presents the results when we allow the coefficients to vary over time (by including interactions with Round). We find that receivers do not display much evidence of learning (right-most column of the “Receiver Actions” panel). The coefficient on Target  $\times$  Message Difference decreases by  $-0.11$  from the first round to the last, but the difference is not statistically significant ( $p < 0.16$ , one-tailed).

Although we clearly reject the possibility that receivers play equilibrium strategies (Receiver Actions Hypothesis), we do find evidence that supports the comparative statics predictions of the theory (Message Difference). In the next section, we test the set of seemingly counterintuitive comparative statics regarding senders’ messages when we relax the symmetry of senders’ preferences.

## Extremist Moderation and Moderate Extremism

Our main comparative statics predictions concern the effects of varying the distribution of senders’ preferences, which we test with data from the Symmetric Baseline and Asymmetric Conditions. In the latter, the left sender’s shift is more likely to be larger than the right sender’s, and the Extremist Moderation Hypothesis predicts that the left sender (the “Extremist”) will send jamming messages closer to the target in the Asymmetric Condition than in the Symmetric Baseline Condition. In contrast, the Moderate Extremism Hypothesis predicts that the right sender (the “Moderate”), whose preference distribution does not change across conditions, will send jamming messages farther from the target in the Asymmetric Condition than in the Symmetric Baseline Condition.

We test these hypotheses by regressing the *Distance* between a sender’s message and

---

<sup>17</sup>See Model [6] from Appendix Table A-3.

<sup>18</sup>In the High Payment Condition, receivers’ behavior tends away from the equilibrium prediction; the coefficient on Target  $\times$  Message Difference is larger (closer to 0) in the High Payment Condition.

the receiver’s target on a series of indicator variables—Jam, Left Sender, Right Sender, and an indicator for the *Asymmetric* Condition—and all their interactions. We also estimated a regression including interactions with Round to allow for learning. Here, we expect the findings associated with the comparative statics hypotheses to be more evident in the last round than in the first round. Figure 4 presents the results from these two regressions. Filled markers indicate the Average Distance when the target is not in the jamming region, and empty markers indicate that Average Distance when the target does lie in the jamming region. The circles present point estimates of Average Distance across rounds. The dotted lines indicate the difference between messages sent when the target is inside and outside of the jamming region.

[Figure 4 about here.]

We first note that in almost every comparison, Average Distance is higher inside the jamming region than outside the jamming region, providing additional support for the equilibrium Conditional Jamming Hypothesis. In the Symmetric Condition, it is 0.16 higher ( $p < 0.01$ ) for the right sender and 0.23 higher for the left sender ( $p < 0.01$ ), and in the Asymmetric Condition it is 0.21 higher ( $p < 0.01$ ) for the (moderate) right sender and 0.06 higher ( $p \approx 0.05$ ) for the (extremist) left sender.

Turning to the comparative statics, we find strong evidence in favor of the Extremist Moderation Hypothesis. Relative to the Symmetric Baseline Condition (empty circles in the upper white box), extremists clearly temper their jamming messages in the Asymmetric Condition (empty circles in the upper gray box). The distance between jamming and non-jamming messages is 0.18 larger in the Symmetric Condition than in the Asymmetric Condition ( $p < 0.01$ , one-tailed). Over time, the difference becomes more pronounced; by the last round, the difference between conditions rises to 0.45 ( $p < 0.01$ ).

Support for the Moderate Extremism Hypothesis is weaker, yet still supportive of the theory when learning is accounted for. Relative to the Symmetric Baseline (empty circles in the lower white box), the distance between jamming and non-jamming messages is 0.05 larger

in the Asymmetric Condition than in the Symmetric Condition in the Asymmetric Condition (empty circles in the lower gray box), but this difference is not significant ( $p \approx 0.24$ ). However, over time, the predicted pattern does emerge. By the last round, the difference between conditions rises to 0.24 ( $p < 0.02$ ).

Despite the support we find for an intricate set of seemingly counterintuitive comparative statics, evidence of the overjamming phenomenon again emerges. In Figure 4, each black circle and triangle is predicted to be equal to 0 (because messages are supposed to be truthful outside the jamming region). All exceed that mark. Moreover, overjamming seems to increase as play proceeds. This phenomenon cannot be explained by equilibrium theory, as the effects should decrease over time. Thus, we must consider other possible explanations.

## Limited Strategic Sophistication

The fundamental idea underlying equilibrium analysis is the *mutual consistency* of beliefs and actions. Each player is assumed to choose the best response given her beliefs about what others will do and those beliefs are also assumed to be consistent with what others actually do. In this section, we consider whether a form of bounded rationality that relaxes the mutual consistency assumption helps to explain overjamming. Specifically, we continue to assume that individuals take into account the actions of others, but depart from equilibrium analysis by allowing players to have beliefs that do not perfectly coincide with other players' actions. That is, we consider the possibility that individuals are limited in their strategic sophistication.<sup>19</sup>

If we assume that senders form beliefs about what other players do and then choose the message that maximizes their payoffs given those beliefs, our modified theoretical framework must specify how those beliefs might be formed. We consider two possibilities. First, we

---

<sup>19</sup> Although a substantial literature in economics investigates limited strategic sophistication—experimentally (e.g., Camerer, Ho, and Chong 2004; Costa-Gomes, Crawford, and Broseta 2001; Nagel 1995; Stahl and Wilson 1995) and theoretically (e.g., Crawford 2003)—such models are rare in political science. For example, Dickson, Hafer, and Landa (2008) classify subject behavior in their experiment into several boundedly rational types. For a theoretical application in political economy, see Binswanger and Prufer (2010).

apply a “level- $K$ ” model in which subjects form expectations based on iterated reasoning: some subjects are naive, minimally sophisticated subjects best respond to naive behavior, more sophisticated subjects best respond to minimally sophisticated behavior, and so on. This model is useful for generating predictions about subjects who confront the game for the first time and for investigating variation in strategic sophistication. Second, we develop a model of “experiential best responses,” in which each sender expects that his opponent’s message will be the average of the messages he has observed his opponents send previously. The latter model helps to understand how individuals might adapt their behavior from experience and to gauge how sophisticated they are in doing so. To be clear, our goal is not to offer a complete model of behavior. Rather, we use these frameworks to understand aspects of communicative competition that cannot be explained by the equilibrium theory, especially overjamming.

**Level- $K$  Reasoning.** In the “level- $K$ ” framework,  $K$  denotes the degree of sophistication a subject evinces.<sup>20</sup> Level-0 senders are non-strategic and use naive decision rules. Level-1 senders believe their opponents are level-0 and choose the appropriate best response. In general, level- $K$  senders best respond given the belief that their opponents are level- $(K - 1)$ . Thus,  $K$  refers to the number of steps of iterated reasoning.

Throughout, we maintain the assumption that all senders believe receivers will choose the action equal to the average of the two messages she observes. Given that receivers observe neither the target nor the senders’ shifts, this assumption is simple and intuitive, and therefore plausible.<sup>21</sup> We then focus the remainder of our analysis on the implications of this model for sender behavior.

The logic of our level- $K$  analysis is straightforward and reveals a clear pattern between

---

<sup>20</sup>Our model is closer to Nagel (1995) and Stahl and Wilson (1995) than to Camerer, Ho, and Chong (2004). The latter develop a cognitive hierarchy model that generalizes level- $K$  analysis by assuming that senders form beliefs about the distribution of other players’ strategies. Instead, we assume that senders form the simple belief that the opposing sender uses a particular strategy with probability one.

<sup>21</sup>It is also consistent with the level- $K$  framework. As long as the receiver believes both senders are of the same level  $K > 0$ , the receiver’s best response is the average of the two messages.

the level of sophistication  $K$  and a sender's best response.<sup>22</sup> Naive, level-0, senders ignore strategic considerations and report their own ideal points. Level-1 senders realize that their goal is to send a message such that the average (of their own message and the opposing level-0 player's message) is equal to their own (the level-1 sender's) ideal point. This implies that level-1 senders will exaggerate in the direction of their own shift: left senders exaggerate to the left in order to pull the average message to their own ideal point on the left and right senders exaggerate to the right to pull the average to their ideal point on the right. Higher level senders will exaggerate even further so as to counteract the exaggeration of lower level senders. Thus, the level- $K$  framework has the potential to explain the overjamming phenomenon.<sup>23</sup>

More precisely, the model implies that each sender should send a message equal to his own target ( $T + S$ ) plus an additional ***Exaggeration***. The degree of Exaggeration is the sum of  $|S|$ , the magnitude of the sender's shift, and a multiple of  $|E(S_{opp})|$ , the magnitude of the expected value of the opposing sender's shift, where higher multiples are associated with higher levels of sophistication.<sup>24</sup>

We use two methods to empirically assess the level- $K$  framework. In the first method, we classify each subject's level of sophistication using a two-step process. We compare observed messages to the predicted message for each level and classify a *message* as belonging to type- $K$  if the distance between the prediction and message is within an error band (about 12% of the message space).<sup>25</sup> We then classify a *subject* as being level- $K$  if at least 70% of the

---

<sup>22</sup>For simplicity of presentation, we do not formally derive this pattern in the text, instead relegating its derivation to the appendix.

<sup>23</sup>Our model differs from models in the economics literature that assume level-0 players choose randomly. In any level- $K$  model, assumptions must be made about what constitutes level-0 behavior. The key property, in our view, is that level-0 behavior must be naive and non-strategic but this does not necessarily imply that it must be random. Anchoring our analysis with random level-0 messages implies that level-1 players would sometimes send messages on the opposite side of the target from their shifts, but this alternative level- $K$  model is clearly rejected by our previous analysis; messages are strongly related to targets (contradicting random message behavior) and senders consistently overjam in the same direction as their ideal points (contradicting best responses to random opposing messages).

<sup>24</sup>In the Symmetric Baseline Condition, the expected value of the opposing sender's shift is  $|E(S_{opp})| = 25$ .

<sup>25</sup>Specifically, let  $M$  be an observed message and let  $m_K$  be the predicted message for type  $K \in \{0, 1, 2, 3\}$ . if  $|M - m_K|$  is less than 0.35 (about 12% of the message space), we consider the message as being consistent

subjects' messages are consistent with type- $K$  messages. If this criterion is satisfied for more than one level, we make strict classifications by selecting the level at which this consistency percentage is largest.<sup>26</sup>

The classification analysis reveals a distribution of subjects' levels of strategic sophistication. We classify 71% of subjects, most of whom possess some degree of strategic sophistication beyond naivete (i.e.,  $K > 0$ ). Specifically, 15% are classified as level 0, 54% as level-1, and 2% as level-2. No senders are classified as having sophistication greater than level-2. The level- $K$  framework therefore appears to organize the data reasonably well.<sup>27</sup>

Best responses in the level- $K$  framework also differ subtly yet fundamentally from those in the equilibrium analysis. The key distinction is that in the level- $K$  bounded rationality framework, Messages depend unconditionally on Shifts in addition to Targets, rather than conditionally on Targets being in the jamming region. The second method we use to assess the level- $K$  framework is therefore to regress the Message on Target, Shift, and intercepts for Left and Right Sender. The level- $K$  framework implies that the coefficient on Target should be 1, the coefficient on Shift should be 2, and the intercepts should depend on a subject's sophistication level.

[Figure 5 about here.]

The results in Figure 5 (circles in the first column) lend support to our interpretation that bounded rationality—in the form of limited strategic sophistication—contributes to overjamming. Starting at the top, the coefficient on Target is close to (albeit significantly different from) 1. The coefficient on Shift is also close to (but significantly less than) 1. Messages

---

with the type.

<sup>26</sup>This method is consistent with those used by Cai and Wang (2006) and Costa-Gomes, Crawford, and Broseta (2001). If we do not make strict classifications using the largest-percentage criterion, some subjects are assigned to multiple levels. We chose these classification thresholds because they maximize the percentage of subjects uniquely classified (54.3%).

<sup>27</sup>We also compared behavior in the competitive communication game with a one-shot Beauty Contest game but found that classifications in the two games differed markedly, likely because of significant differences in their strategic contexts. Whereas the Beauty Contest game implicates few obvious norms of good behavior, communicating immediately raises the prospect of truth-telling and lying.

are therefore unconditional functions of Shifts as predicted by the level- $K$  framework, but much less responsive than the framework predicts. Thus, it seems that, relative to payoff-maximizing levels of exaggeration established by the level- $K$  framework, there is systematic “understatement” in exaggerations. However, we also continue to see that the Left and Right Sender intercepts are significantly different from zero, meaning that the overjamming phenomenon may indeed owe to variation in strategic sophistication. Furthermore, when we add interactions with Round, the intercepts on Left Sender and Right Sender move further from 0 over time, which is consistent with the idea that senders become more sophisticated with experience.

The level- $K$  framework has provided considerable insight about overjamming beyond what could be explained by equilibrium analysis, although the explanation remains incomplete. Senders appear to engage in systematic exaggeration in a way that is generally consistent with limited strategic sophistication. Interestingly, the regressions also identify a second, as yet unexplained phenomenon. If, in fact, senders are engaged in best response play based on mutually inconsistent beliefs, then senders appear to do so in an understated manner—senders do not fully incorporate their shifts into their message strategies. We also find that experience matters, as messages become more exaggerated over time. Our next framework focuses explicitly on the role of experience.

**Experiential Best Responses.** In the level- $K$  framework, we assumed beliefs about what others will do are formed through a process of iterated reasoning (anchored by intuitive conjectures about naive behavior). But as individuals play the game, it is reasonable to assume instead that senders will adjust their beliefs to incorporate their observations about opponents’ behavior and thus play the best response to the empirical distribution of messages they have *experienced*.

The “experiential best response” framework is plausible for at least two reasons. First, subjects are reminded of their history at the end of each round; thus, they may simply be acting on the information we offer them. Second, we found above that senders exaggerate

more over time. Rather than becoming more sophisticated, it is possible that senders are instead simply responding to a (self-perpetuating) trend in the messages they observe. By focusing on the alternative framework, we attempt to disentangle these two explanations.

We now assume each sender believes his opponent to send a message equal to the opponent's ideal point plus the average Exaggeration he has observed.<sup>28</sup> The logic involved in deriving the experiential best response that follows from this belief is virtually identical to the derivation of best responses in the level- $K$  framework.<sup>29</sup> The more that a sender has observed his opponents' exaggerate in the past, the more the sender will himself exaggerate in order to pull the average message (i.e., what the sender expects to be the receiver's action) to his own ideal point. The experiential best response framework implies, like in the level- $K$  framework, that messages will be functions of Target and Shift. Unlike the level- $K$  framework, however, it predicts that the remaining extent of exaggeration will be completely determined by opponents' past exaggeration rather than by the distribution of opponents' shifts or level of strategic sophistication. That is, once the target, shift, and opponents' exaggeration are taken into account, there should be no additional exaggeration by either left or right senders.

To apply this framework to our data, we regress Message on Target, Shift, Left and Right Sender intercepts, and average Exaggeration. Because senders may have short or long memories, we use two different measures of average Exaggeration. First, we measure Exaggeration over the entire history a sender experienced, from round 1 up to the most previous round. Second, we measure Exaggeration as a moving average over the most recent five rounds. In each case, we expect the coefficient on Exaggeration to be -1, on Shift to be 2, on Target to be 1, and on Left Sender and Right Sender to both be 0.

---

<sup>28</sup>Formally, the expected value of an opponent's message is  $E(m_{opp}) = T + \bar{e}$ , where  $\bar{e}$  is the sample average Exaggeration of his opponent's messages. Thus, in round  $t$ ,  $\bar{e} = \sum_{\tau=1}^{t-1} (m_{opp}^{\tau} - T^{\tau})$ , where  $\tau$  indexes previous rounds.

<sup>29</sup>As in the level- $K$  framework, we assume that receivers choose the average of the senders' messages. This is broadly consistent with the sender behavior we observe; senders appear to choose actions that are weighted averages of the left and right senders' messages, although the weights are slightly less than 0.5 for both senders.



The regression coefficients reported in the second and third columns of Figure 5 suggest that the experiential best response framework provides a satisfactory account of overjamming. Using Exaggeration over the entire history of play, we see that its coefficient (indicated by circles in the second column, “Entire History”) is  $-0.83$  and that the predicted value of  $-1$  is within its confidence interval. Moreover, overjamming has now been accounted for, as zero is now in the confidence intervals of the separate left and right sender intercepts. These results are corroborated by the results when we include interactions with Round (triangles in the second column). The coefficient on Exaggeration moves from  $-0.65$  in the first round (up-pointing triangles) to  $-0.81$  in the last round (down-pointing triangles), and the intercepts (bottom two rows) do not move away from 0.

However, we also continue to find that the coefficient on Shift remains far less than 2. Thus, even when we account for experience, understatement in exaggeration persists. Furthermore, similar but weakened results emerge if we use the moving average measure of Exaggeration (third column, “Last 5 Rounds”). The predicted value of  $-1$  is no longer in the confidence interval of Exaggeration, and overjamming in the intercepts seems to reemerge by the last round. Standard model selection criteria indicate the latter model provides the better fit (e.g., deviance is much lower in the model where experience is based on the last five rounds rather than the entire history).

## Conclusion

We report on the first set of experiments to combine competition, preference uncertainty, and information transmission in a spatial setting. This strategic environment captures key features of many real-world political arenas in established democracies. In the realm of policy-making institutions, lobbyists, businesses, activists, and policy experts compete to influence legislation, regulatory rules, and court decisions. In the public arena, parties, candidates, and public intellectuals clash in attempts to sway public opinion.

Although many scholars have suggested that competition can help resolve information

transmission problems (Austen-Smith and Wright 1992; Gilligan and Krehbiel 1989; Krehbiel 1991; Page and Shapiro 1992), our results imply a more cautious view is warranted. We uncover systematic patterns of behavior that suggest enthusiasm for the benefits of competition should be tempered by the realization that there are potent strategic incentives for competing interests to engage in obfuscation and misdirection. The “marketplace of ideas” is replete with manipulative and false advertising.

We found that senders who are predicted to jam send messages that are further from the truth than those who are predicted not to, senders “overjam” by consistently exaggerating in the direction of their biases, and moderate senders send more extreme messages than their extremist opponents. These findings become more apparent the longer subjects play and are explained partly by the equilibrium jamming theory and partly by a theory of limited strategic sophistication. Importantly, competition does not fully mitigate senders’ strategic incentives to exaggerate. While receivers’ actions are correlated with the truth, there is nevertheless widespread, often substantial, deviation that increases as messages become more polarized. This is a consequence of jamming and overjamming.

Our analysis helps to understand the nature of policy debates across a wide variety of issues and institutional settings, especially those for which there is great uncertainty about the underlying science or policy consequences. For example, environmentalists often point to a scientific consensus about climate change and claim that it poses a catastrophic threat to human life while opponents argue that the economic costs of mitigating climate change are far too great and cast doubt on the credibility of the science. Advocates of genetically modified foods claim significant benefits including increased crop yields with the potential to end world hunger while opponents raise fears of “Frankenfoods” and the potential for adverse health effects. Energy companies claim that extracting natural gas via hydraulic fracturing is safe, creates jobs, and reduces reliance on foreign oil while critics argue that the “fracking” exploits rural communities, is dangerous to the environment, and poisons urban water supplies. In the domain of health policy, proponents claim that the Affordable Care

Act will tame the rise of health care costs while opponents believe it is will kill jobs and burden taxpayers. And in the realm of economic policy, liberals and conservatives constantly feud about the relative merits of fiscal stimulus versus austerity measures, each exaggerating the benefits and underestimating the costs of their respective favored policies.

The results suggest that for each of these policy issues, and many others, political discourse will consistently feature competing interests who might be instrumentally exaggerating the truth in pursuit of their own narrow policy goals. The consequence is a highly polarized political atmosphere in which policymakers and citizens can infer neither the truth itself nor which arguments in a debate are closer to the truth. Indeed, our analysis suggests that even if decisionmakers recognize that both sides exaggerate, the best response at first may simply be to split the difference. But then this gives advocates and critics incentives to exaggerate further, with little or no hope of ever reaching a consensus. Eventually decisionmakers may tune them out, leaving all parties worse off from the inability to communicate important policy-relevant information.

Although the normative implications of this study tend to be pessimistic, our experimental setting also provides opportunities to study conditions that encourage or discourage the truth to emerge in a competitive political environment. For example, priming norms of honesty, allowing senders and receivers to converse in ordinary language, or framing the game as a scientific investigation may decrease jamming and exaggeration while framing the game as a game of politics or lobbying may increase them. Greater truth-telling might also emerge from varying the sequence or structure of debate, such as allowing for rebuttals, cross-examination or the selection of specific questions by the decision-maker (e.g., limiting the range of responses). Or it might be more likely to emerge from environments in which interests can develop long-term reputations, in which there is variation in the expertise of competing interests, or variation in the goals of policy-makers. We hope that future research will fully explore the strategic incentives created by competition and preference uncertainty—two unavoidable features of democratic politics.

## References

- Austen-Smith, David, and John R. Wright. 1992. "Competitive Lobbying for a Legislator's Vote." *Social Choice and Welfare* 9: 229–257.
- Bailey, Michael, Brian Kamoie, and Forrest Maltzman. 2005. "Signals from the Tenth Justice: The Role of the Solicitor General at the Merits Stage." *American Journal of Political Science* 49: 72–85.
- Bawn, Kathleen. 1995. "Political Control Versus Expertise: Congressional Choices about Administrative Procedures." *American Political Science Review* 89(1): 62–73.
- Bendor, Jonathan., Serge Taylor, and Roland Van Gaalen. 1985. "Bureaucratic Expertise versus Legislative Authority: A Model of Deception and Monitoring in Budgeting." *The American Political Science Review* 79(4): 1041–1060.
- Binswanger, Johannes, and Jens Prufer. 2010. "Democracy, Populism, and (Un)bounded Rationality."
- Blume, Andreas, Douglas V. DeJong, Yong-Gwan Kim, and Geoffrey B. Sprinkle. 1998. "Experimental Evidence on the Evolution of Meaning of Messages in Sender-Receiver Games." *American Economic Review* 88: 1323–1340.
- Blume, Andreas, Douglas V. DeJong, Yong-Gwan Kim, and Geoffrey B. Sprinkle. 2001. "Evolution of Communication with Partial Common Interest." *Games and Economic Behavior* 37: 79–120.
- Boudreau, Cheryl, and Mathew D. McCubbins. 2008. "Nothing but the Truth? Experiments on Adversarial Competition, Expert Testimony, and Decision Making." *Journal of Empirical Legal Studies* 5(December): 751–789.
- Cai, Hongbin, and Joseph Tao-Yi Wang. 2006. "Overcommunication in Strategic Information Transmission Games." *Games and Economic Behavior* 56: 7–36.
- Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong. 2004. "A Cognitive Hierarchy Model of Games." *Quarterly Journal of Economics* 119(3): 861–898.
- Canes-Wrone, Brandice, Michael C. Herron, and Kenneth W. Shotts. 2001. "Leadership and Pandering: A Theory of Executive Policymaking." *American Journal of Political Science* 45(3): 532–50.
- Carrubba, Clifford J. 2009. "A Model of the Endogenous Development of Judicial Institutions in Federal and International Systems." *Journal of Politics* 71(1): 55–69.
- Chong, Dennis, and James N. Druckman. 2007. "Framing Public Opinion in Competitive Democracies." *American Political Science Review* 101: 637–655.
- Costa-Gomes, Miguel, Vincent P. Crawford, and Bruno Broseta. 2001. "Cognition and Behavior in Normal-Form Games: An Experimental Study." *Econometrica* 69: 1193–1235.

- Crawford, Vincent, and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50: 1431–1451.
- Crawford, Vincent P. 1998. "A Survey of Experiments on Communication via Cheap Talk." *American Economic Review* 78: 286–298.
- Crawford, Vincent P. 2003. "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions." *American Economic Review* 93: 133–149.
- Dickhaut, John W., Kevin A. McCabe, and Arijit Mukherji. 1995. "An Experimental Study of Strategic Information Transmission." *Economic Theory* 6: 389–403.
- Dickson, Eric S., Catherine Hafer, and Dimitri Landa. 2008. "Cognition and Strategy: A Deliberation Experiment." *The Journal of Politics* 70(04): 974–989.
- Esterling, Kevin M. 2004. *The Political Economy of Expertise*. Ann Arbor: University of Michigan Press.
- Fischbacher, Urs. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics* 10: 171–178.
- Fox, Justin, and Kenneth W. Shotts. 2009. "Delegates or Trustees? A Theory of Political Accountability." *Journal of Politics* 71(4): 1225–37.
- Gilligan, Thomas W., and Keith Krehbiel. 1987. "Collective Decisionmaking and Standing Committees: An Informational Rationale for Restrictive Amendment Procedures." *Journal of Law, Economics, and Organization* 3: 287–335.
- Gilligan, Thomas W., and Keith Krehbiel. 1989. "Asymmetric Information and Legislative Rules with a Heterogeneous Committee." *American Journal of Political Science* 33: 459–490.
- Gneezy, Uri. 2005. "Deception: The Role of Consequences." *American Economic Review* 95: 384–394.
- Haptonstahl, S.R. 2009. "Bargaining Under Uncertainty: a Strategic Statistical Model of the Ultimatum Game."
- Hurkens, Sjaak, and Navin Kartik. 2009. "Would I Lie to You? On Social Preferences and Lying Aversion." *Experimental Economics* 12: 180–192.
- Iyengar, Shanto, and Donald R. Kinder. 1987. *News that matters: Television and American opinion*. University of Chicago Press.
- Kiewiet, D. Roderick, and Mathew D. McCubbins. 1991. *The Logic of Delegation: Congressional Parties and the Appropriations Process*. Chicago: University of Chicago Press.
- Krehbiel, Keith. 1991. *Information and Legislative Organization*. Ann Arbor, MI: University of Michigan Press.

- Landau, Martin. 1969. "Redundancy, Rationality, and the Problem of Duplication and Overlap." *Public Administration Review* 29(4): 346–358.
- Lupia, Arthur, and Mathew D. McCubbins. 1998. *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge: Cambridge University Press.
- Minozzi, William. 2011. "A Jamming Theory of Politics." *Journal of Politics* 73(2): 301–315.
- Nagel, Rosemarie. 1995. "Unraveling in Guessing Games: An Experimental Study." *The American Economic Review* 85(5): 1313–1326.
- Nelson, Thomas E., Rosalee A. Clawson, and Zoe M. Oxley. 1997. "Media Framing of a Civil Liberties Conflict and Its Effect on Tolerance." *American Political Science Review* pp. 567–583.
- Page, Benjamin I., and Robert Y. Shapiro. 1992. *The Rational Public*. Chicago: University of Chicago Press.
- Stahl, D.O., and P.W. Wilson. 1995. "On Players' Models of Other Players: Theory and Experimental Evidence." *Games and Economic Behavior* 10(1): 218–254.
- Staton, Jeffrey K. 2006. "Constitutional Review and the Selective Promotion of Case Results." *American Journal of Political Science* 50: 98–112.
- Vanberg, Georg. 2001. "Legislative-Judicial Relations: A Game-Theoretic Approach to Constitutional Review." *American Journal of Political Science* 45: 346–361.
- Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.

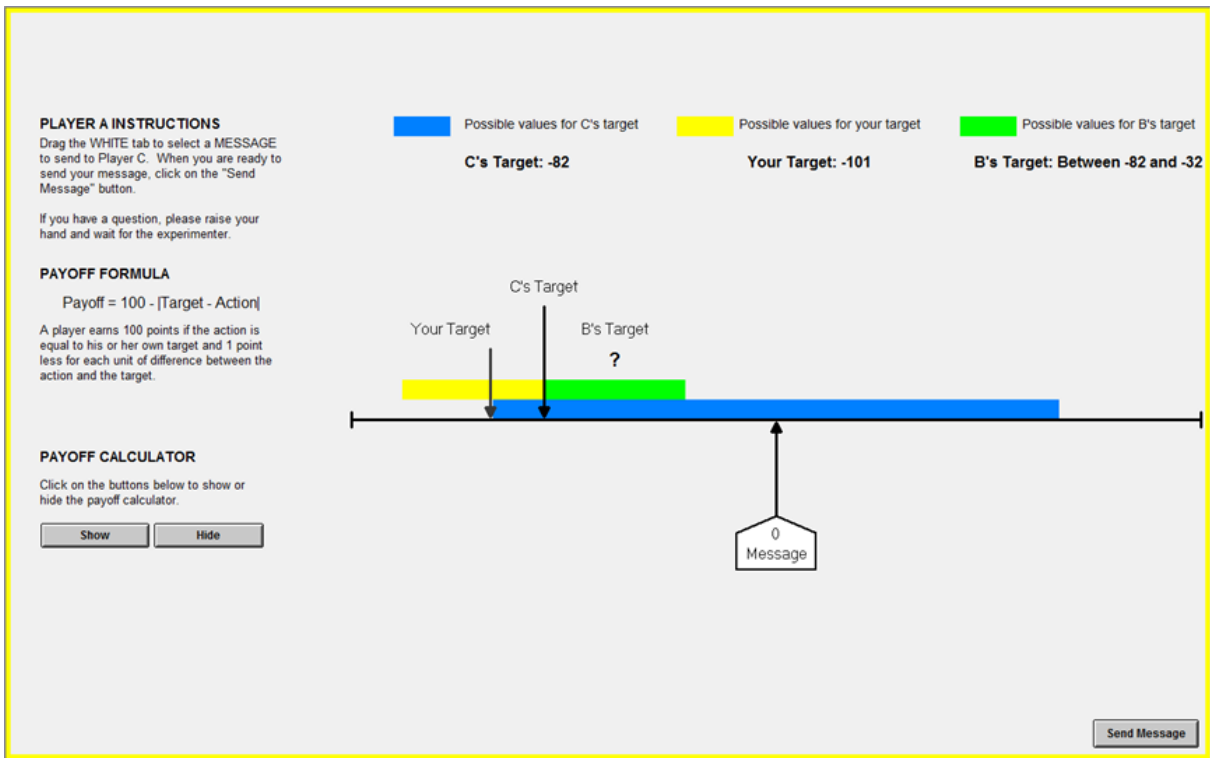


Figure 1: *Screenshot from the Experiment*

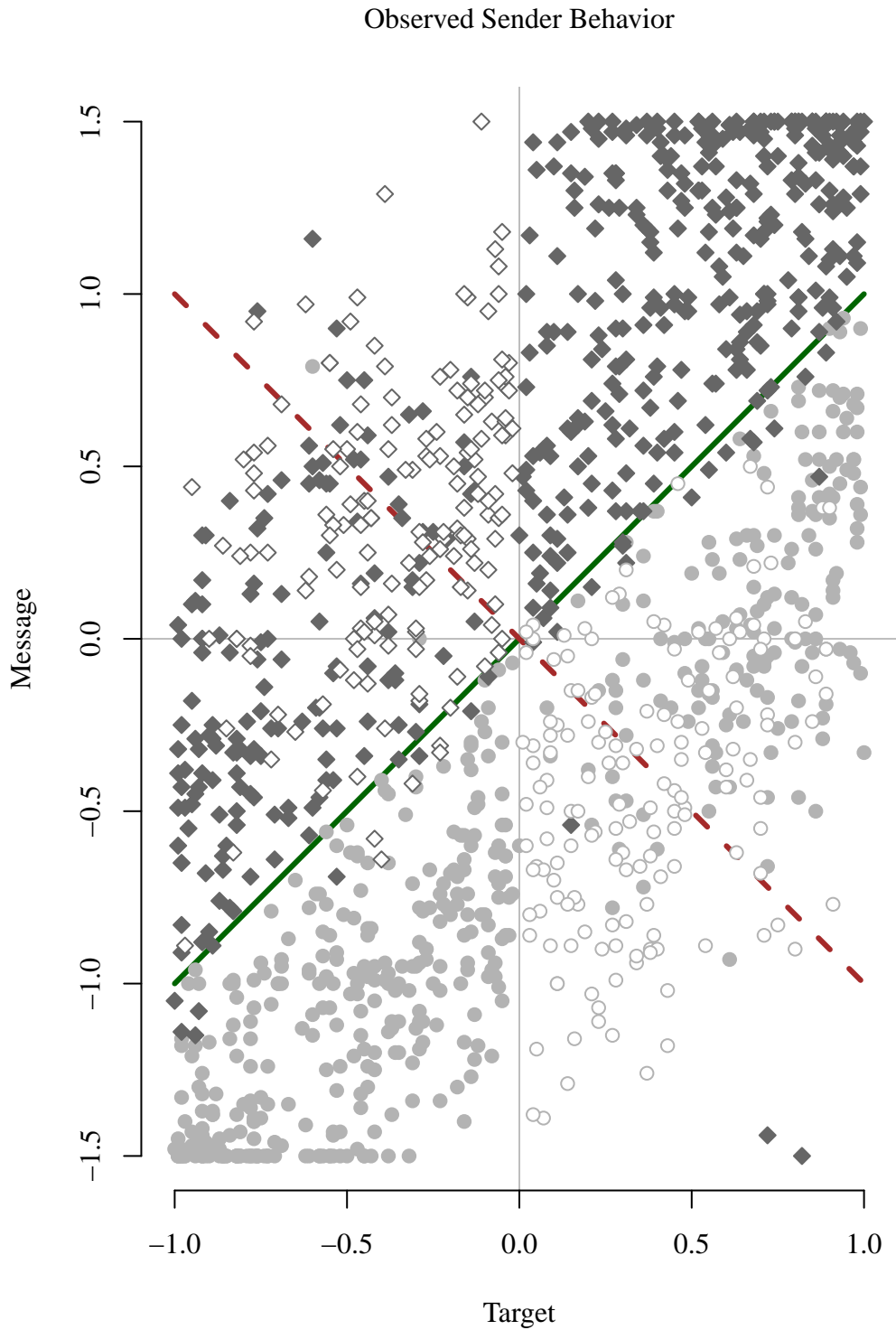


Figure 2: *Observed Sender Behavior*. Messages from Right Senders are notated with (darker) diamonds; messages from Left Senders by (lighter) circles. Filled markers represent Targets outside the jamming region, and empty markers represent Targets within the jamming region. The up-sloping unbroken line depicts the hypothesized relationship between Target and Message when the Target is outside the jamming region, and the down-sloping dashed line represents the hypothesized relationship when the Target is in the jamming region.



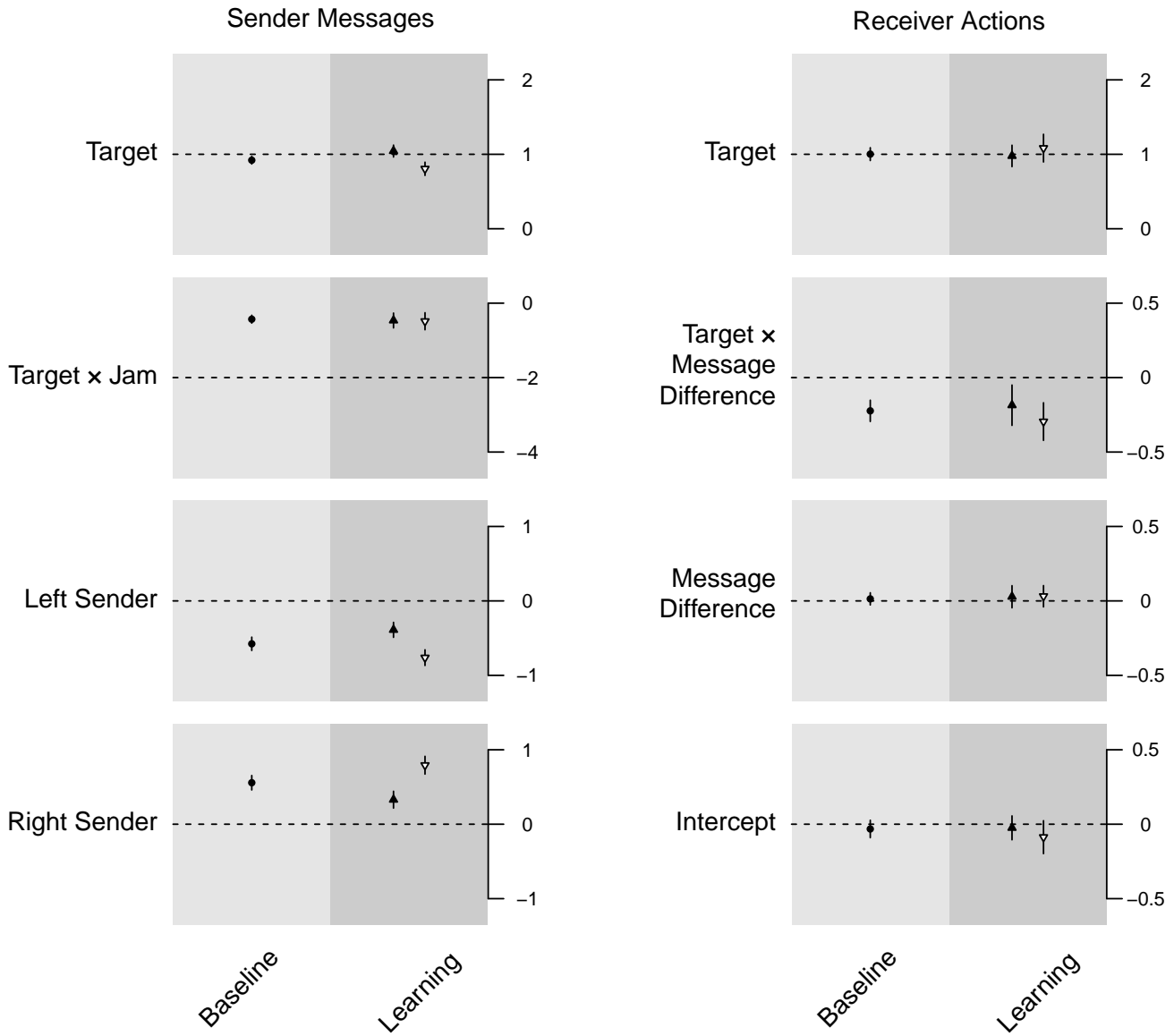


Figure 3: *Equilibrium Messages and Actions*. Each column displays coefficients and standard errors from a separate regression.

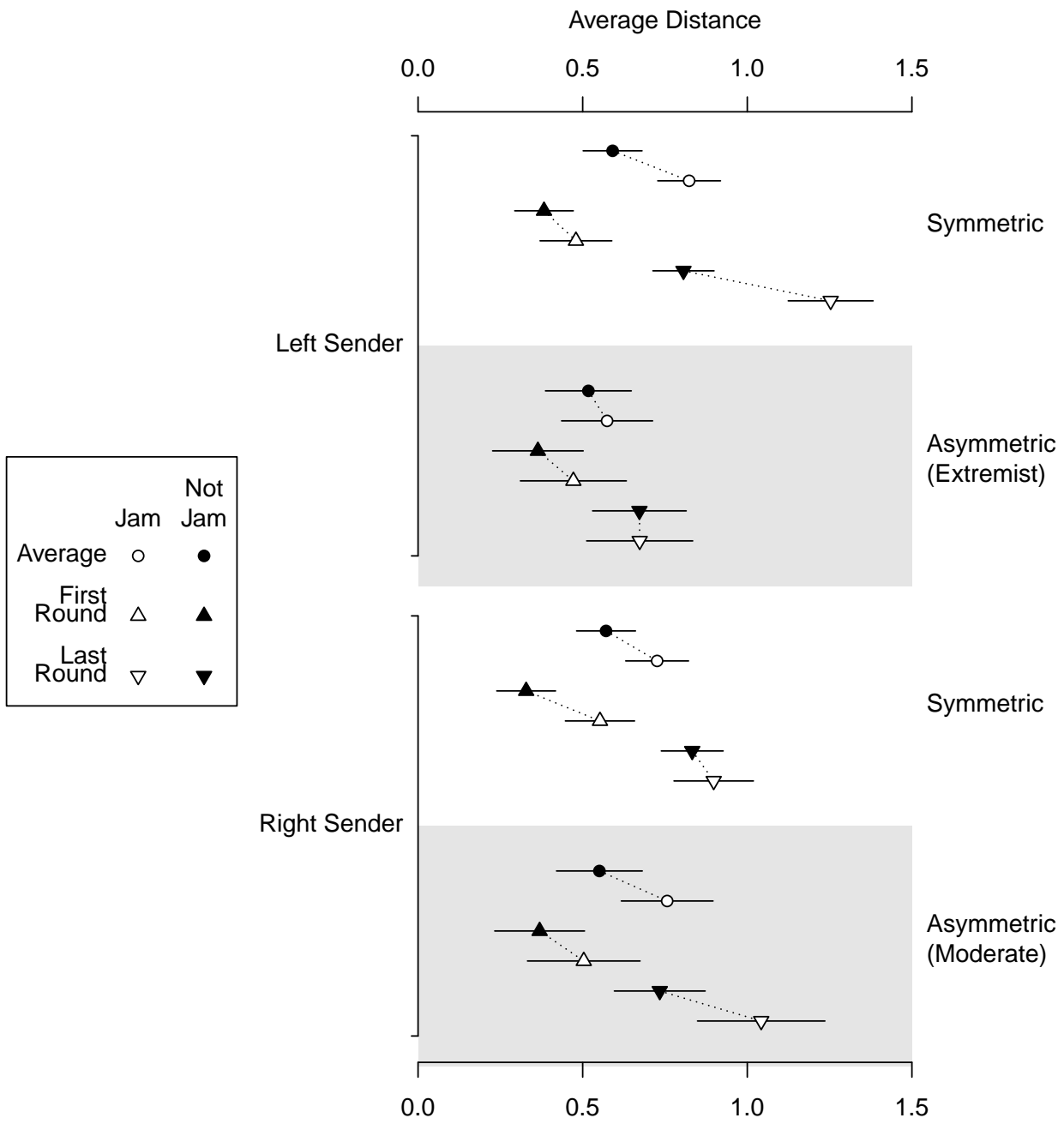


Figure 4: *Extremist Moderation and Moderate Extremism.*

### Limited Strategic Sophistication

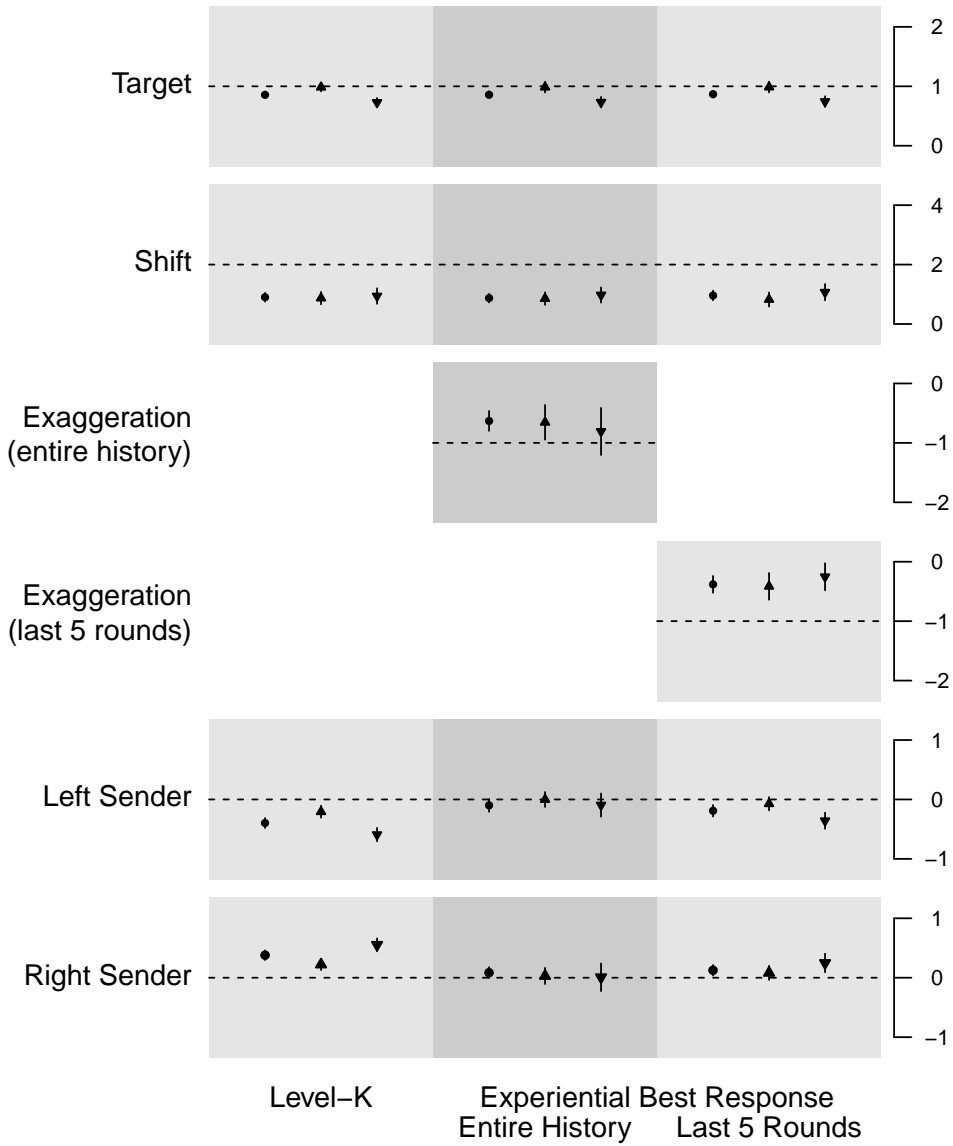


Figure 5: *Limited Strategic Sophistication*. Each column displays point estimates and 95% confidence intervals from regressions of Message. Circles refer to regressions without Round interactions; up-pointing triangles to first-round estimates from models with Round interactions; and down-pointing triangles to last-round estimates.

# Supplemental Web Appendix: Experimental Instructions

## Instructions

### General Information

This is an experiment in communication. The University of Pittsburgh has provided funds for this research. If you follow the instructions closely and make appropriate decisions, you may make a considerable amount of money. In addition to the \$7 participation payment, these earnings will be paid to you, in cash, at the end of the experiment.

During the experiment, all earnings will be denominated in points, which will be converted to cash at the rate of \$1 per 150 points. The exact amount you receive will be determined during the experiment and will depend on your decisions and the decisions of others. You will be paid your earnings privately, meaning that no other participant will find out how much you earn. Also, each participant has a printed copy of these instructions. You may refer to your printed instructions at any time during the experiment.

**If you have any questions during the experiment, please raise your hand and wait for an experimenter to come to you. Please do not talk, exclaim, or try to communicate with other participants during the experiment. Also, please ensure that your cell phones are turned off and put away for the duration of the experiment. Participants intentionally violating the rules will be asked to leave the experiment and may not be paid.**

### Roles, Rounds, and Matching

Each participant will be assigned to one of three roles: A, B, or C. Your role will be assigned before the first round and will remain fixed throughout the experiment.

In this experiment you will make decisions in a series of rounds, and there are a total of 32 rounds. **Each round is a separate decision task. Before every round, you will be randomly matched with two other participants.** In every group of three participants there will be one player in each role (one A player, one B player, and one C player).

**You will not know the identity of the other participants you are matched with in any round, and your earnings for each round depend only on your action in that round and the actions of the participants you are matched with in that round.**

### Targets

At the beginning of every round, the computer will randomly select a **target** for each player.

**Player C's target will be a number between -100 and 100.** Each number is equally likely to be C's target.

**Player A's target will be less than Player C's target.** The difference between A's target and C's target will be some amount between 0 and 50 units. Each amount is equally likely, and the exact amount will be selected at random in every round.

**Player B's target will be greater than Player C's target** by some amount between 0 and 50 units and each amount of difference is equally likely.

For example, suppose that the computer selects 25 as Player C's target. For Player A's target, the computer will randomly select a number from -25 to 25. Likewise, Player B's target will be a randomly selected number from 25 to 75.

It is important to note that Player A's target and Player B's target are randomly selected by the computer independently. That is, the value of Player A's target does not affect the value of Player B's target and vice versa.

Similarly, the computer will randomly determine each player's target at the beginning of the round so that the targets in one round are selected independently of the targets in another round.

## Sequence of Decisions

The sequence of decisions in every round is as follows:

1. Players A and B each find out the value of Player C's target and the value of their own target. (Note that Player A does not see Player B's target, nor does Player B see Player A's target.) Independently and simultaneously, Players A and B each select a **message** to send to Player C.
2. Player C sees the messages sent by Player A and Player B. Player C then chooses an **action** (any number between -150 and 150). (Note that Player C sees both messages but none of the targets.)

## Payoffs

Each player's payoff depends only on how close Player C's action is to his or her own target. More specifically, a player earns 100 points if the action is equal to his or her own target and 1 point less for each unit of difference between the action and the target. This is described by the following formula (where the straight lines indicate absolute value):

$$\text{Player's Payoff} = 100 - |\text{Player's Target} - \text{C's Action}|$$

Note that the messages sent by Player A and Player B are not part of the payoff formula.

To illustrate, consider a few examples. Suppose you are Player A, your target is 10 and Player C chooses the action 40. The difference between your target and the action is 30, so your payoff would be 70. If Player C's target is 25, then the difference between C's target and the action is 15, so C's payoff would be 85.

Now suppose instead that Player C chooses the action -40. If Player A's target is 20, then the difference between A's target and the action is 60 and A's payoff would be 40. If Player B's target is 80, then the difference between B's target and the action is 120, so B's payoff would be -20. If Player C's target is 45, then the difference between C's target and the action is 85, so C's payoff would be 15. (Note that it is possible for payoffs to be negative.)

## Sample Screens

We will now see what the screens look like for each type of player during the experiment.

This is the screen that will be seen only by Player A. There is a brief set of instructions in the upper left-hand corner. A description of the payoff formula is also shown on the left

side of the screen. The top of the screen shows several values: C's actual target, A's target (which is labeled "your target"), and the range of possible targets for B.

The targets are indicated graphically in the figure in the middle of the screen, which also indicates the possible range of values for each player's target. Player A chooses a message by dragging the white tab to any position along the horizontal black line. After moving the tab, it will indicate the value of the selected message.

Note that there is also a section on the left marked "payoff calculator." Click on the "Show" button to reveal an orange tab that can be used to calculate hypothetical payoffs for each possible action that Player C can take. If you move the orange tab to different positions, the bold text at the bottom of the screen changes to indicate what Player A's payoff and player C's payoff would be. Note that the payoff calculator does not show B's hypothetical payoff because you do not know the value of B's target. Note also that you can hide the payoff calculator by clicking on the "hide" button.

When Player A is ready to send the message, he or she will click on the "Send Message" button in the lower right-hand corner of the screen. Feel free to move the message tab and try out the payoff calculator. When you are ready to continue, click on the "Send Message" button.

This is the screen that only Player B will see. B players see this screen at the same time that the A players see their screens. It is pretty much the same as Player A's screen except that B's target is known while A's is not. When you are done looking at this screen, click on the "Send Message" button to continue.

After Player A and Player B send their messages, Player C will see this screen. In the upper-left corner there is again a brief set of instructions. The top of the screen shows the numerical values of the messages. The messages are also indicated graphically in the middle of the screen. To select an action, Player C moves the red tab to the desired location. As with the other tabs, it shows the numerical value of its location after it is moved. Note that Player C does not have a payoff calculator because the actual values of the targets are not known. Try moving the "Action" tab and the click on "Choose Action" button when you are ready to continue.

At the end of every round, you will see this screen, which shows you the results from the round—including the actual targets of every player, both messages, and the action chosen by Player C, and the payoffs earned by every player in your group. At the bottom of the screen, it will show the results of every previous round that you played.

## QUIZ INSTRUCTIONS.

To check your understanding of the decision tasks, please answer the questions below as best you can. **Note that your quiz answers do not affect your earnings**, and you may refer to your printed instructions as often as you like. When you are finished, feedback about the correct answers will be shown on the screen. You must attempt to answer all of the questions. If you have any further questions at this time, please raise your hand and the experimenter will come to you.

1. C's target can be any number from: [0 to 10, 0 to 100, -100 to 100, -150 to 150]
2. If C's target is -40, then A's target can be any number from: [-100 to 0, -90 to -40, -40 to 10, 40 to 90]

3. If C's target is 30, then B's target can be any number from: [-20 to 30, 0 to 50, 30 to 80, 50 to 100]
4. If you are Player C, your target is 85, and you choose the action 45, how many points will you receive? [15, 40, 60, 85]
5. If you are Player A, your target is -70, and Player C chooses the action 50, how many points will you receive? [-70, -20, 30, 50]
6. Suppose that you are Player B, your target is 10 and Player C's target is -15. If you send the message 10 and Player C chooses the action 0, how many points will you receive? [10, 15, 85, 90]
7. Suppose that you are Player C. Player A sent you the message -50 while Player B sent you the message 50. If you choose the action 30 and your actual target was 50, how many points will you receive? [20, 30, 70, 80]
8. In every round, will you be matched with same participants? [Yes, No]

## **Bonus Game**

[NOTE: This section was available only to subjects in the experimental sessions that included the Beauty Contest.]

Before we conclude the experiment, there will be a bonus game.

The total prize for winning the bonus game is \$6, and the rules are simple.

- In this game, you may choose any whole number from 0 to 100.
- The computer will calculate the average of all the numbers submitted.
- The winner is the person who chooses the number closest to  $2/3$  of the average.
- There can be more than one winner in the case of ties. If there is more than one winner, then the winners will split the prize equally.

If you have a question about the bonus game, please raise your hand.

# Supplemental Web Appendix: Tables

Table A-1: Summary Statistics for Experimental Sessions

Session	Condition	Subjects	Rounds	$T$	$S_L$	$S_R$	$m_L$	$m_R$	$c$
1	Symmetric Baseline	18	32	0.01 (0.58)	-0.26 (0.15)	0.25 (0.15)	-0.59 (0.59)	0.41 (0.67)	-0.08 (0.58)
2	Symmetric Baseline	15	32	0.00 (0.58)	-0.25 (0.15)	0.26 (0.15)	-0.73 (0.60)	0.78 (0.66)	0.03 (0.58)
3	High Payment (Symmetric)	15	12	-0.01 (0.57)	-0.24 (0.16)	0.27 (0.14)	-0.28 (0.55)	0.27 (0.58)	0.03 (0.57)
4	High Payment (Symmetric)	15	12	-0.02 (0.60)	-0.24 (0.14)	0.26 (0.15)	-0.42 (0.60)	0.39 (0.57)	-0.05 (0.60)
5	Asymmetric	12	28	-0.04 (0.57)	-0.31 (0.14)	0.26 (0.15)	-0.63 (0.61)	0.56 (0.62)	0.00 (0.57)
6	Asymmetric	15	32	0.01 (0.59)	-0.31 (0.13)	0.25 (0.14)	-0.32 (0.66)	0.57 (0.58)	0.08 (0.59)
7	Symmetric with Beauty Contest	18	24	0.04 (0.59)	-0.25 (0.15)	0.24 (0.14)	-0.57 (0.62)	0.63 (0.63)	0.02 (0.59)
8	Symmetric with Beauty Contest	18	24	0.05 (0.63)	-0.25 (0.14)	0.25 (0.15)	-0.60 (0.64)	0.57 (0.63)	0.05 (0.63)

Sample means and standard deviations are reported by session for target  $T$ , left sender's shift  $S_L$ , right sender's shift  $S_R$ , left sender's message  $m_L$ , right sender's message  $m_R$ , and receiver's action  $c$ . All parameters are have been rescaled by 1/100.



Table A-2: Senders in the Symmetric Baseline Condition

DV = Message	[1]	[2]	[3]	[4]
Target	0.93 (0.02)	0.94 (0.02)	0.92 (0.03)	1.04 (0.04)
Target $\times$ Jam	-0.44 (0.05)	-0.45 (0.07)	-0.43 (0.05)	-0.24 (0.07)
Jam	-0.04 (0.02)	-0.06 (0.04)	-0.05 (0.04)	0.08 (0.06)
Left Sender	–	–	-0.58 (0.05)	-0.39 (0.05)
Right Sender	–	–	0.56 (0.05)	0.33 (0.06)
Round $\times$ Target	–	–	–	-0.24 (0.07)
Round $\times$ Target $\times$ Jam	–	–	–	-0.02 (0.19)
Round $\times$ Jam	–	–	–	-0.28 (0.11)
Round $\times$ Left	–	–	–	-0.37 (0.07)
Round $\times$ Right	–	–	–	0.46 (0.08)
Intercept	-0.02 (0.09)	-0.01 (0.09)	–	–
<i>Subject <math>\sigma</math></i>				
Target	–	0.06	0.08	0.07
Target $\times$ Jam	–	0.08	0.11	0.10
Jam	–	0.04	0.07	0.07
Intercept	0.59	0.58	0.18	0.18
<i>Round <math>\sigma</math></i>				
Target	–	0.09	0.10	0.08
Target $\times$ Jam	–	0.26	0.10	0.14
Jam	–	0.19	0.15	0.15
Left Sender	–	–	0.14	0.08
Right Sender	–	–	0.18	0.11
Intercept	0.05	0.07	–	–
Residual $\sigma$	0.30	0.28	0.25	0.25

The reported models includes random intercepts and slopes by subject and round, as indicated. There are 1280 observations, 46 subjects, and 32 rounds. The Baseline column of the Senders panel in Figure 3 is based on Model [3], and the Learning column is based on Model [4].

Table A-3: Actions in the Symmetric Baseline Condition

DV = Action	[6]	[7]	[8]
Target	1.01 (0.04)	1.00 (0.04)	1.03 (0.08)
Message Difference	0.01 (0.02)	0.01 (0.02)	0.03 (0.04)
Target $\times$ Message Difference	-0.23 (0.03)	-0.22 (0.04)	-0.21 (0.08)
Round	-	-	-0.04 (0.10)
Round $\times$ Target	-	-	-0.10 (0.15)
Round $\times$ Message Difference	-	-	-0.01 (0.07)
Round $\times$ Target $\times$ Message Difference	-	-	0.01 (0.12)
Intercept	-	-	-0.02 (0.05)
Subject $\sigma$			
Intercept	0.04	0.04	0.03
Round $\sigma$			
Target	-	0.06	0.08
Message Difference	-	0.05	0.05
Intercept	0.02	0.08	0.08
Residual $\sigma$	0.24	0.24	0.24

The reported models includes random intercepts and slopes by subject and round, as indicated. There are 640 observations, 23 subjects, and 32 rounds. The Baseline column of the Receivers panel in Figure 3 is based on Model [5], and the Learning column is based on Model [7].

Table A-4: Distance and Learning in Symmetric and Asymmetric Environments

DV = Distance	[10]	[11]
<i>Right Sender</i>		
SRN	0.57 (0.05)	0.33 (0.05)
SRJ	0.73 (0.05)	0.55 (0.05)
Round $\times$ SRN	–	0.50 (0.05)
Round $\times$ SRJ	–	0.35 (0.08)
<i>Moderate</i>		
ARN	0.55 (0.07)	0.37 (0.07)
ARJ	0.76 (0.07)	0.50 (0.09)
Round $\times$ ARN	–	0.36 (0.06)
Round $\times$ ARJ	–	0.54 (0.13)
<i>Left Sender</i>		
SLN	0.59 (0.05)	0.38 (0.05)
SLJ	0.82 (0.05)	0.48 (0.06)
Round $\times$ SLN	–	0.42 (0.05)
Round $\times$ SLJ	–	0.77 (0.09)
<i>Extremist</i>		
ALN	0.52 (0.07)	0.36 (0.07)
ALJ	0.57 (0.07)	0.47 (0.08)
Round $\times$ ALN	–	0.31 (0.07)
Round $\times$ ALJ	–	0.20 (0.10)
Subject Intercept $\sigma$	0.18	0.18
Round Intercept $\sigma$	0.13	0.04
Residual $\sigma$	0.25	0.24

Regressors are indicators for each block, which are notated by **S**ymmetric or **A**symmetric Condition, **L**eft or **R**ight Sender, and **J**amming Region or **N**ot.

Table A-5: Limited Strategic Sophistication

DV = Message	[12]	[13]	[14]	[15]	[16]	[17]
Target	0.86 (0.02)	0.98 (0.04)	0.86 (0.03)	0.99 (0.05)	0.87 (0.03)	0.99 (0.05)
Shift	0.90 (0.08)	0.87 (0.11)	0.91 (0.07)	0.85 (0.11)	0.96 (0.08)	0.83 (0.12)
Exaggeration (Entire History)	–	–	-0.83 (0.10)	-0.64 (0.15)	–	–
Exaggeration (Last 5 Rounds)	–	–	–	–	-0.38 (0.07)	-0.41 (0.11)
Left Sender	-0.40 (0.04)	-0.21 (0.05)	0.00 (0.05)	-0.01 (0.07)	-0.19 (0.05)	-0.07 (0.06)
Right Sender	0.38 (0.04)	0.22 (0.05)	-0.03 (0.05)	0.04 (0.07)	0.13 (0.05)	0.08 (0.06)
Round × Target	–	-0.26 (0.06)	–	-0.25 (0.08)	–	-0.24 (0.08)
Round × Shift	–	0.07 (0.19)	–	0.13 (0.20)	–	0.25 (0.21)
Round × Exag. (Entire History)	–	–	–	-0.17 (0.27)	–	–
Round × Exag. (Last 5 Rounds)	–	–	–	–	–	0.16 (0.18)
Round × Left	–	-0.39 (0.08)	–	-0.09 (0.11)	–	-0.29 (0.08)
Round × Right	–	0.33 (0.07)	–	-0.03 (0.15)	–	0.17 (0.10)
<i>Subject <math>\sigma</math></i>						
Target	0.07	0.07	0.09	0.09	0.09	0.09
Shift	0.38	0.39	0.37	0.38	0.38	0.39
Exaggeration	–	–	0.35	0.37	0.28	0.28
Intercept	0.15	0.15	0.16	0.18	0.17	0.17
<i>Round <math>\sigma</math></i>						
Target	0.10	0.08	0.12	0.10	0.12	0.10
Shift	0.04	0.02	0.09	0.08	0.21	0.15
Exaggeration	–	–	0.31	0.30	0.24	0.22
Left Sender	0.14	0.09	0.06	0.06	0.08	0.04
Right Sender	0.11	0.05	0.12	0.12	0.09	0.08
Residual $\sigma$	0.24	0.24	0.22	0.22	0.22	0.21
Deviance	268	204	98	59	74	29
$n$ Observations	1280	1280	1234	1234	1234	1234
$n$ Subjects	46	46	46	46	46	46
$n$ Rounds	32	32	31	31	31	31

These models are the basis for Figure 5.

# Supplemental Web Appendix: Additional Analysis

## Payment Sensitivity

To investigate the Payment Sensitivity Hypothesis, we ran additional sessions under a *High Payment Condition*, in which the information structure of the Symmetric Baseline Condition was replicated, with two changes: the cash conversion rate was increased to \$1 for every 40 points (so that every round in the High Payment Condition was worth \$2.50), and subjects played only 12 rounds.

To test the possibility that overjamming resulted from low stakes, we amend the original regression models to include interactions with an indicator for **High Payment**, and use data from both the Symmetric Baseline and the High Payment Conditions.<sup>30</sup> The second column (Payoff Sensitivity) of the Sender Messages panel in Figure 3 presents the results. In the figure, the black squares indicate results for the Symmetric Baseline Condition, while the white squares indicate results for the High Payment Condition. The Payment Sensitivity Hypothesis predicts that the empty squares will be closer to their predicted values than the filled squares. For example, the empty square by Target should be closer to 1 than the filled square.

The results conform to those we hypothesized in only a few cases. The coefficient on Target actually moves down a bit under the High Payment Condition, against the predicted direction. Similarly, the coefficient on Target  $\times$  Jam moves up, again opposite to the predicted direction. However, under the High Payment condition, the Jam coefficient does not change significantly, and the intercepts on Left and Right Sender move closer to their predicted equilibrium values of 0. Left Sender becomes 0.13 (0.09) larger ( $p \approx 0.07$ , one-tailed), while Right Sender becomes 0.14 (0.09) smaller ( $p \approx 0.06$ , one-tailed). Stakes alone do not seem sufficient to encourage senders to play best responses.

## Derivation of Best Responses in Level- $K$ Model

In our competitive communication game, there are two plausible level-0 strategies that senders might employ. Senders might be *naive truthful* types ( $t_0$ ) who report the truthful message  $m_{t_0} = T$ . Alternatively, senders might be *naive selfish* types ( $s_0$ ) who instead report their own targets,  $m_{s_0} = T + S_j$ . The latter type of sender attempts to maximize his utility but does not consider how his opponent's strategy affects the receiver's action.

If a subject has sophistication  $K > 0$ , his reasoning process must ultimately be based on one of these two level-0 types. Suppose first that the naive truthful type anchors the iterated reasoning process. Type  $t_1$  denotes the level-1 subject who believes he is playing a truthful opponent. A subject of this type believes that the receiver will choose  $c = \frac{1}{2}(T + m_{t_1})$ , and so his best response is to choose  $m_{t_1} = T + 2S$ . To see this, recall that the sender wants to induce the receiver to choose an action equal to his own target,  $T + S$ . Thus,  $E(c) = \frac{1}{2}E(T + m_{t_1}) = T + S$  if and only if  $m_{t_1} = T + 2S$ . This argument is equally valid for left and right senders, regardless of the sign of  $S$ . At the next level of sophistication, type  $t_2$  believes he faces a type  $t_1$  opponent. Type  $t_2$  believes that  $c = \frac{1}{2}(T + 2S_{opp} + m_{t_2})$ , where  $S_{opp}$  is the opponent's shift. Although he does not know his opponent's shift, each

---

<sup>30</sup>To ensure that the difference in number of rounds played in the conditions does not confound the results, these models use only the first 12 rounds of data. See Model [5] from Appendix Table B-1 for details.

Table B-1: Sender Payment Sensitivity

DV = Message	Coef.	[5] SE
Target	0.97	0.09
Target $\times$ Jam	-0.32	0.15
Jam	0.03	0.06
Left Sender	-0.44	0.06
Right Sender	0.43	0.06
High Pay $\times$ Target	-0.09	0.13
High Pay $\times$ Target $\times$ Jam	-0.03	0.23
High Pay $\times$ Jam	0.03	0.09
High Pay $\times$ Left	0.13	0.09
High Pay $\times$ Right	-0.14	0.09
Subject $\sigma$		
Intercept	0.17	
Target	0.10	
Round $\sigma$		
Intercept	0.04	
Target	0.03	
Block $\sigma$		
Intercept	0.005	
Target	0.12	
Residual $\sigma$	0.24	
$n$ Observations	792	
$n$ Subjects	66	
$n$ Rounds	12	
$n$ Blocks	8	

In the High Payment Condition, each round was worth four times as much as a round in the Symmetric Baseline Condition. The reported model includes random intercepts and slope on Target by subject, round, and block (i.e., condition, sender type, jam or not jam). This model is the basis for the middle column in the “Senders’ Messages” panel of Figure 3.

Table B-2: Receiver Payment Sensitivity

DV = Action	Coef.	SE
Target	1.05	0.10
Message Difference	0.004	0.042
Target $\times$ Message Difference	-0.29	0.09
Intercept	-0.01	0.04
High Pay $\times$ Target	-0.28	0.17
High Pay $\times$ Message Difference	-0.23	0.09
High Pay $\times$ Target $\times$ Message Difference	0.14	0.19
High Pay	0.18	0.08
Subject $\sigma$		
Intercept	0.09	
Target	0.29	
Message Difference	0.09	
Target $\times$ Message Difference	0.26	
Round $\sigma$		
Intercept	0.00	
Residual $\sigma$	0.26	
$n$ Observations	396	
$n$ Subjects	33	
$n$ Rounds	12	

In the High Payment Condition, each round was worth four times as much as a round in the Symmetric Baseline Condition. The reported model includes random intercepts and slope on Target by subject and round. This model is the basis for the middle column in the “Receivers’ Actions” panel of Figure 3.

sender does know its sign and distribution. The best response is to choose the message that will ensure  $E(c) = T + S$ , which is  $m_{t2} = T + 2S - 2 * E(S_{opp})$ . For example, the left sender knows that  $S_{opp}$  is distributed uniformly between 0 and 50; therefore, his best response is  $m_{t2} = T + 2S_L - 50$ . Similar reasoning implies that a right sender with type  $t2$  has best response  $m_{t2} = T + 2S_R + 50$ . Furthermore, continuing this pattern of reasoning indicates that at the next level,  $m_{t3} = T + 2S_L - 100$  for left senders and  $m_{t3} = T + 2S_R + 100$  for right senders.

Rather than taking the naive truthful type as the anchor of the iterative reasoning process, suppose instead that a the naive selfish type is the base. Very similar reasoning yields the following conclusions. A left sender with type  $s1$  will have the best response  $m_{s1} = T + 2S_L - 25$ ; a right sender with type  $s1$  has  $m_{s1} = T + 2S_L + 25$ . At the next level,  $m_{s2} = T + 2S_L - 75$  for left senders and  $m_{s2} = T + 2S_R + 75$  for right senders.

In general, our simple model of limited strategic sophistication implies that message strategies will be a linear combination of  $T$ ,  $S$ , and a constant. Messages that reflect a Level-1 or higher degree of strategic sophistication take the general form  $m = \alpha + T + 2S$  where  $\alpha$  is generally some multiple of  $E(|S|) = 25$ , and  $\alpha \leq 0$  for left senders and  $\alpha \geq 0$  for right senders. In contrast, naive strategies are either less responsive or unresponsive to the shift parameter and do not involve a constant term.

Table B-3: Receivers' Limited Strategic Sophistication

Independent Variable		
Average Message	0.85 (0.02)	-0.87 (0.03)
Intercept	0.001 (0.010)	0.002 (0.015)
Round $\times$ Average Message	-	-0.03 (0.05)
Round	-	-0.002 (0.024)
$n$ Observations	1280	1280
$n$ Subjects	46	46
$n$ Rounds	32	32

Notes: Data are from the Symmetric Baseline Condition.

The dependent variable is Message. Coefficients and standard errors are from a multilevel model that includes a random intercept and slopes on Target, Shift, and Right Sender by subject.  $*p < 0.05$ , two-tailed.

Round has been rescaled to vary from 0 to 1.



## Beauty Contest

To vet this classification, we augmented two of our experimental sessions with a Beauty Contest, the game for which the level- $K$  model was originally developed (Nagel 1995). Subjects first played the strategic communication game and then played a “Bonus Game”. In the “Bonus Game,” subjects guess a number between 0 and 100. The subject who chooses the number closest to  $2/3$  of the average guess is paid an extra \$6. In equilibrium, everyone should guess 0, but this strategy is based on many iterations of strategic reasoning. Thus, the more iterations of reasoning one applies, the lower one’s guess will be. Using the data from the Beauty Contest, we identify best responses for each level  $K$  and classify a subject as level- $K$  if her guess is within a threshold of the level- $K$  best response.<sup>31</sup>

Surprisingly, the level- $K$  framework classifies subjects in our strategic communication game and the Beauty Contest very differently. Only 15 of the 24 subjects were classified in both games, and, of those, only 3 were classified at the same level. To ensure that this mismatch does not owe to overly stringent thresholds, we reclassified senders as being type  $K$  if at least 50% of the subjects’ messages are consistent with type  $K$  messages; the result does not change. The mismatch may owe to the different sorts of sophistication induced by the two games, e.g., social norms activated by communication. Whereas the Beauty Contest implicates few obvious norms of good behavior, communicating immediately raises the prospect of truth-telling and lying.

### Inferred Expectations.

In our last framework, we shift perspectives. Rather than positing possible beliefs and deriving best responses, we assume that senders have beliefs, and, however idiosyncratic those beliefs may be, that they choose the best response conditional on those beliefs. According to this framework, a sender who expects that his opponent will send the message  $E(m_{opp})$  will maximize his payoffs by choosing  $M = 2(T + S) - E(m_{opp})$ . Thus, as we observe  $T$ ,  $S$ , and  $M$ , we can infer that the sender must have had expectation  $E(m_{opp}) = 2(T + S) - M$ .

Given these inferred expectations, we can investigate two separate questions. First, how well do these inferred expectations match the messages sent by senders’ opponents? And second, do inferred expectations match opponents’ messages more closely given more experience? To answer these questions, Figure B-6 presents a scatterplot of Opponent’s Message against Inferred Expectation. Up-pointing triangles are points from the first 16 rounds of play, and down-pointing triangles are from the remaining rounds. If inferred expectations matched opponents’ messages perfectly, they would hew to the dashed 45° line. Of course, senders’ opponents have private information about their own Shifts, and therefore, we should not expect perfect matches. However, if senders’ expectations are correct on average, points should be gathered around the dashed line.

Left senders (filled triangles) appear to systematically underestimate their Opponents’ Messages, while right senders (empty triangles) overestimate those messages. These directions are consistent with the idea that senders believe they are one step more sophisticated

---

<sup>31</sup>Following Nagel (1995), the best response for level- $K$  is  $50(\frac{2}{3})K$ , and the tolerance thresholds are given by  $[50(\frac{2}{3})^{K-.5}, 50(\frac{2}{3})^{K+.5}]$ . All subjects who guess a number more than 50 are unclassified; 17 of 24 subjects were successfully classified. We studied several alternative thresholds, none of which substantively altered the results.

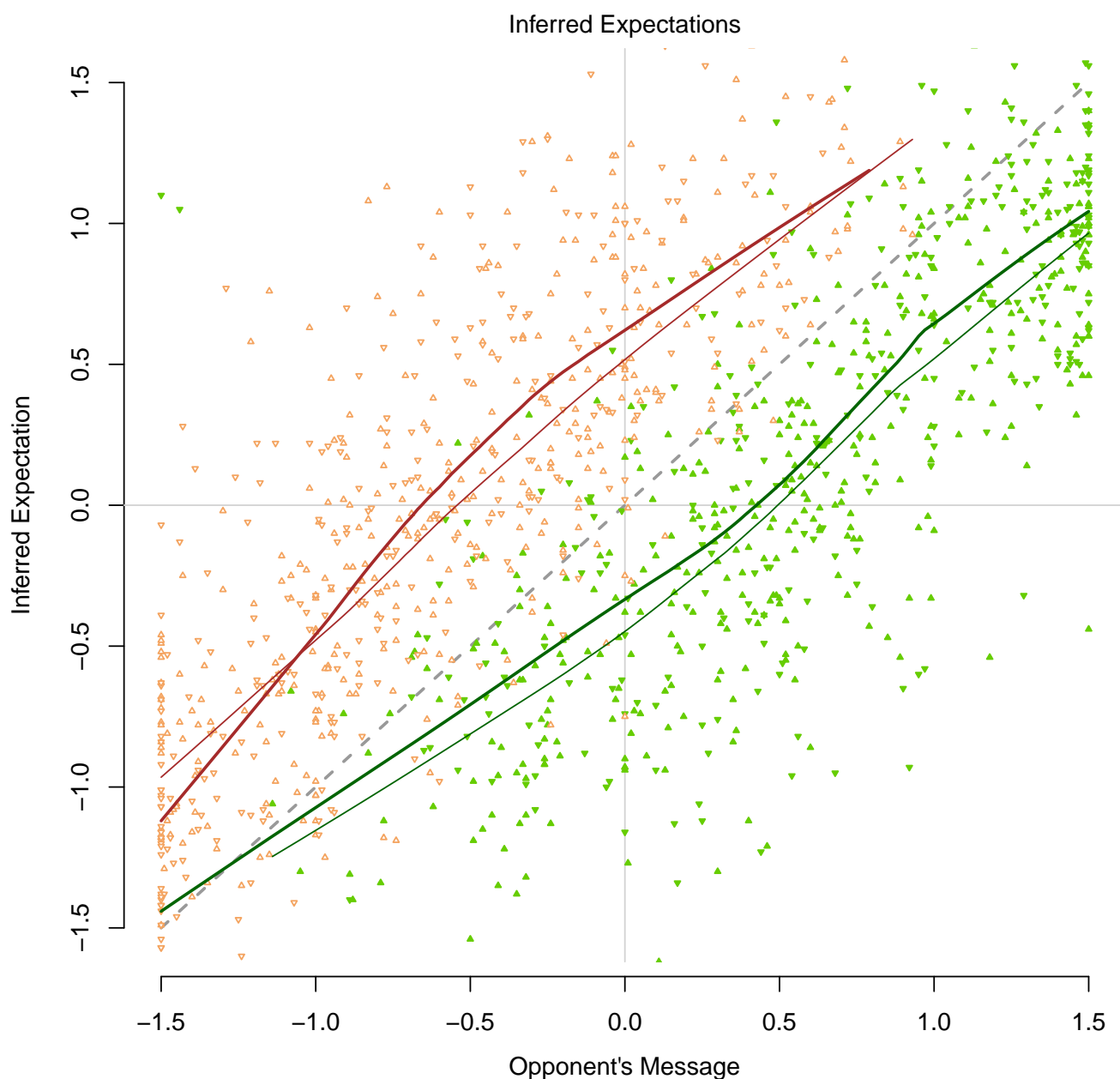


Figure B-6: *Inferred Expectations*. Given observed behavior, we infer what senders expected their opponents to do, assuming best response play. Here, senders appear to underestimate their opponents' exaggerations. The dashed line represents a perfect match between inferred expectations and opponents' messages; the points mark observed data. Separate nonparametric regression lines are fitted and displayed for left senders (below the dashed line) and right senders (above the dashed line) for the first 16 rounds (up-pointing triangles, thin lines) and the remaining rounds (down-pointing triangles, thick lines). Although there is a positive relationship between the two, both are clearly distinct from the dashed line. Moreover, left senders seem to improve their inferences over time, while right senders do not.

than their opponents, even though they are not. To further test this possibility, examine how the differences inferred expectations and opponents' messages change throughout the game. To that end, Figure B-6 displays nonparametric regression lines by sender type and round (thin lines for the first 16 rounds, thick lines for the remaining rounds). For left senders (the lower lines), it appears that the thick line is closer to the dashed line, meaning that left senders may be developing more accurate beliefs as the game progresses. However, for right senders (the upper lines), the two lines cross, meaning that there is no clear improvement in beliefs over time.