

Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games

By JOSEPH TAO-YI WANG, MICHAEL SPEZIO, AND COLIN F. CAMERER*

We report experiments on sender-receiver games with an incentive for senders to exaggerate. Subjects “overcommunicate”—messages are more informative of the true state than they should be, in equilibrium. Eyetracking shows that senders look at payoffs in a way that is consistent with a level-k model. A combination of sender messages and lookup patterns predicts the true state about twice as often as predicted by equilibrium. Using these measures to infer the state would enable receiver subjects to hypothetically earn 16–21 percent more than they actually do, an economic value of 60 percent of the maximum increment. (JEL C72, C91, D82, Z13)

“Why do almost all people tell the truth in ordinary everyday life?—Certainly not because a god has forbidden them to lie. The reason is, firstly because it is easier; for lying demands invention, dissimulation, and a good memory.”

—Friedrich Nietzsche, *Human, All Too Human*, II.54, 1878/1996

During the tech-stock bubble, Wall Street security analysts were alleged to inflate recommendations about the future earnings prospects of firms in order to win investment banking relationships with those firms.¹ Specifically, analysts of Merrill Lynch used a five-point rating system (1 = Buy to 5 = Sell) to predict how the stock would perform. They usually gave two 1–5 ratings for short run (0–12 months) and long run (more than 12 months) performance separately. Henry Blodget, Merrill Lynch's famously optimistic analyst, “did not rate any Internet stock a 4 or 5” during the bubble period (1999 to 2001). In one case, the online direct marketing firm LifeMindars, Inc. (LFMN), Blodget first reported a rating of 2-1 (short run “accumulate”—long run “buy”) when Merrill Lynch was pursuing an investment banking relationship with LFMN. Then, the stock price gradually fell from \$22.69 to the \$3–\$5 range. While publicly maintaining his initial 2-1 rating, Blodget privately e-mailed fellow analysts that “LFMN is at \$4. I can't

*Wang: Department of Economics, National Taiwan University, 21 Hsu-Chow Road, Taipei 100, Taiwan (e-mail: josephw@ntu.edu.tw); Spezio: Division of Humanities and Social Sciences, California Institute of Technology, MC228-77, 1200 E. California Blvd., Pasadena, CA 91125 and Department of Psychology, Scripps College, 1030 Columbia Avenue, Claremont, CA 91711 (e-mail: mlspezio@hss.caltech.edu); Camerer: Division of Humanities and Social Sciences, California Institute of Technology, MC228-77, 1200 E. California Blvd., Pasadena, CA 91125 (e-mail: camerer@hss.caltech.edu). Research support was provided by grants from the Betty and Gordon Moore Foundation and the Human Frontiers of Society Program (HFSP). Alex Brown, Noah Myung, Eileen Chou, Min Jeong Kang, Dan Knoepfle, and Chun-Ting Chen provided excellent research assistance. Thanks to comments from Vincent Crawford, Moran Cerf, Robert Östling, Navin Kartik, David Eil, Carlos R. Lever, two anonymous referees, and the audience of the ESA 2005 Tucson Meeting, the tenth BDRM Conference, the 2007 AEA Annual Meeting, the Yale SOM Whitebox Student Conference on Behavioral Science, and the 26th Arne Ryde Symposium of Communication in Games and Experiment.

¹For a detailed description of the tech-stock bubble, see Michael J. Brennan (2004). For evidence regarding analyst recommendations affected by conflicts of interest, see Hsiou-wei Lin and Maureen F. McNichols (1998) and Roni Michaely and Kent L. Womack (1999).

believe what a POS [piece of shit] that thing is." He was later banned from the security industry for life and fined millions of dollars.²

This case is an example of a sender-receiver game with divergent preferences (sometimes called a "cheap talk" or "strategic information transmission" game). Sender-receiver games are simple models of economic situations in which one agent has an incentive to exaggerate the truth to another agent. The central issues in these games are how well uninformed players infer the private information from the actions of players who are better informed, and what informed players do, anticipating the behavioral inference of the uninformed players. Vincent P. Crawford and Joel Sobel (1982) showed that in such games, the sender's incentive to exaggerate when his preferences differ from the receiver's precludes equilibria in which communication is perfectly informative. Instead, all equilibria are noisy, and the larger the difference between the sender's and receiver's preferences, the noisier is the most informative equilibrium. But as in any equilibrium model, there is no systematic deception: The receiver's beliefs conditional on the sender's message are unbiased estimates of the true state. Although previous experimental work starting with John W. Dickhaut, Kevin A. McCabe, and Arijit Mukherji (1995), Andreas Blume et al. (1998, 2001), and cumulating with Hongbin Cai and Wang (2006) in general confirms Crawford and Sobel's comparative statics prediction regarding the noisiness of communication, experiments also tend to disconfirm their model's prediction that with divergent preferences, senders never tell the truth except by accident, and that receivers are never systematically deceived. Crawford (2003) and Navin Kartik, Macro Ottaviani, and Francesco Squintani (2007) show that both "overcommunication" (senders telling the truth more than equilibrium predicts) and systematic deception can be explained by a class of nonequilibrium models of strategic thinking called level- k models. Cai and Wang (2006) show that such a model can describe the data from their sender-receiver experiments and could help to explain the overcommunication their subjects exhibit. The present paper builds on these results and attempts to investigate the cause behind the behavior patterns in such games. Understanding these behavioral patterns better should aid in the design of institutions to foster more accurate transmission of information when preferences diverge.

Incentives for exaggeration are common. Besides the Blodget case mentioned above, similar dramatic accounting frauds in the last few years, such as Enron, Worldcom, and Tyco, might have been caused by the incentives of managers (and perhaps their accounting firms) to inflate earnings prospects (Brennan 2004, and Brian J. Hall and Kevin J. Murphy 2003). For instance, Enron executives told shareholders at meetings that earnings prospects were rosy, at the same time as the executives were selling their own shares, leading to indictments and trials in 2006.³ In universities, grade inflation and well-polished recommendation letters help schools promote their graduates (Henry Rosovsky and Matthew Hartley 2002). Other examples of incentives for strategic information transmission include government-expert relationships in policymaking, doctor-patient relationships in health care choices, teacher cheating on student tests (Brian A. Jacob and Steven D. Levitt 2003) and the floor-committee relationship in Congress.

This paper reports experiments on a sender-receiver game. In the game, a sender learns the true state (a number S) and sends a costless message M to a receiver who then chooses an action

² See Complaint in *Securities and Exchange Commission v. Henry M. Blodget*, (2003) Civ. 2947 (WHP) (S.D.N.Y. No. 18115, paragraph 11–12 and 70–72, April 23; Securities and Exchange Commission *Order Against Henry M. Blodget* (2003), No. 3-11322, October 31; and United States District Court Final Judgement on *Securities and Exchange Commission v. Henry M. Blodget* (2003) Civ. 2947 (WHP) (S.D.N.Y.).

³ According to an SEC complaint filed in court, Kenneth Lay, Enron's then chairman and CEO, said "We will hit our numbers" and "My personal belief is that Enron stock is an incredible bargain at current prices" in an employee online forum on September 26, 2001. However, in the prior two months he was actually making net sales of over \$20 million in Enron stock (back to Enron). See Second Amended Complaint in *Securities and Exchange Commission v. Richard A. Causey, Jeffrey K. Skilling, and Kenneth L. Lay*, (2004) Civil Action No. H-04-0284 (Harmon) (S.D. Tx.) No. 18776, paragraph 81–82, July 8.

A. Payoffs depend only on S and A so the message M is “cheap talk.” The receiver prefers to choose an action that matches the state, but the sender wants the receiver to choose an action closer to $S + b$, where b is a known bias parameter. The value of b is varied across rounds. When $b = 0$ senders prefer that receivers choose S , so they almost always just announce S (i.e., $M = S$), and receivers believe them and choose $A = M$. When $b > 0$ senders would prefer to exaggerate and announce $M > S$ if they thought receivers would believe them. Since subjects choose 1–5, the numbers in our game are coincidentally the same as those used by Merrill Lynch. Indeed, when $b > 0$, we find that our subjects hardly ever report the number 1 (in only 8 percent of 208 rounds), much as Blodget never rated a stock 4 or 5 (the equivalent of 1–2 in our game). However, our game is presented in abstract terms without reference to stock analysts or deception. This could make subjects feel less guilty when “deceiving” others in the experiment.

Besides measuring choices in these games, our experiment uses video-based “eyetracking” to measure what payoffs or game parameters sender subjects are looking at (see Appendix: Methods). Eyetracking software records where players are looking on a computer screen every four milliseconds. These data are a useful supplement to econometric analysis of choices, when decision rules which produce similar choices make distinctive predictions about what information is needed to execute these rules. Previous “eyetracking” studies used a “Mouselab” system in which moving a cursor into a box opens the box’s contents and are more accurately described as “mouse-tracking.” See Camerer et al. (1993), Miguel Costa-Gomes, Crawford, and Broseta (2001), Eric Johnson et al. (2002), Costa-Gomes and Crawford (2006), Xavier Gabaix et al. (2006), and Crawford (2008).⁴

The eyetracking apparatus also measures how much subjects’ pupils “dilate” (expand in width and area). Pupils dilate under stress, cognitive difficulty, arousal, and pain.⁵ Pupillary responses have also been measured in the lie-detection literature for many years (see, for example, F. K. Berrien and G. H. Huntington 1943; Ira Heilveil 1976; Michel P. Janisse 1973; M. T. Bradley and Janisse 1979, 1981; Janisse and Bradley 1980; R. E. Lubow and Ofer Fein 1996; and Daphne P. Dionisio et al. 2001). These studies suggest that pupil dilation might be used to infer deceptive behavior because senders find deception stressful or cognitively difficult.

The choices, eyetracking, and pupil dilation measures generate four basic findings:

1. Overcommunication in sender-receiver game is consistent with $L0$, $L1$, $L2$, and equilibrium (EQ) sender behavior produced by a level- k model of the sender-receiver game in which $L0$ sender behavior is anchored at truth telling.
2. Eyetracking data provide the following support for the level- k model:
 - a. *Attention to structure and own payoffs:* Sender subjects pay attention to important parameters (*state and bias*) of the sender-receiver game. This indicates subjects are

⁴ One small handicap of the Mouselab system is that the experimenter cannot be certain the subject is actually looking at (and processing) the contents of the open box. Our system measures the eye fixation so we can tell if the subject’s eye is wandering, and pupil dilation is measured at the same time (which Mouselab cannot do). Nevertheless, Mouselab systems can be installed cheaply in many computers to measure lookups of many agents at the same time, which could prove useful in running efficient subjects and studying attention simultaneously in complex markets with many agents.

⁵ For pupillary responses to arousal, see Roberta A. Hicks, Tom Reaney, and Lynn Hill (1967), Ray Bull and Gabriel Shead (1979), and Darren Aboyouanand and James M. Dabbs Jr. (1998). For pupillary responses to cognitive difficulty, see Jackson Beatty (1982) and Bram C. Goldwater (1972). For pupillary responses to pain, see C. Richard Chapman et al. (1999) and Shunichi Oka, Chapman, and Robert C. Jacobson (2000). Min Jeong Kang et al (2009) show that pupils dilate in anticipation of finding out the answers to trivia questions about which they are curious. (Their self-reported curiosity is also shown by fMRI to activate the ventral striatum, a brain region involved in anticipated reward or “prediction error” and learning; and curiosity also enhances later memory for mistaken answers.)

thinking carefully about the basic structure of the game, even if they are not following equilibrium theory. Sender subjects also look at their own payoffs more than their opponents'.

- b. *Truth bias*: Sender subjects focus too much on the true state payoff row. This bias is consistent with a failure to “think in the opponent’s shoes” as in Meghana Bhatt and Colin F. Camerer (2005).
- c. *Individual level- k lookup patterns*: Sender subjects focus on the payoffs corresponding to the action $A = S$ (*L0 reasoning*), $A = S + b$ (*L1 reasoning*), ..., up to the corresponding level- k reasoning for each individual subject based on his or her level- k type. This indicates particular level- k type subjects do generally exhibit the corresponding lookup patterns.
3. Right before and after the message is sent, senders’ pupils dilate more when their deception is larger in magnitude. This suggests subjects feel guilty for deceiving (as in Uri Gneezy 2005), or deception is cognitively difficult (as the level- k model assumes).⁶
4. Prediction: Based on the eyetracking results, we can try to predict the true state observed by the sender using lookup data and messages. This prediction exercise suggests it is possible to increase the receiver’s payoff (beyond what was earned in the experiments) by 16–21 percent, resulting in an economic value of 60 percent of the maximum achievable increase.

Finally, this study shows the possible relevance of psychology and neuroscience to economics. Douglas Bernheim (2008) suggests that neuroeconomics will be successful if it can show how new non-choice data can solve a prediction or normative problem that could not be solved by standard choice data. One such problem is how to extract private information from choices. In the standard model, private information is, by definition, not directly observable to outsiders (such as receivers in our game); it can only be inferred assuming a particular model of behavior (e.g., inferring private values from auction bids). If eyetracking, pupil dilation, fMRI, or other biological measures enable one to infer more about private information than by using only choices, those “new” data—new to economists, that is—have some added value for something economists care about. Our data satisfy this criterion because lookups and pupil dilation enhance prediction of the true state beyond the predictions derived simply from observed messages (choice) and equilibrium theory.

This is the first study in experimental economics to use a combination of video-based eyetracking and pupil dilation, and is, of course, exploratory and is therefore hardly conclusive. But the eyetracking and pupil dilation results by themselves suggest that the implicit assumption in equilibrium theories of “cheap talk” in games with communication—namely, that deception has no (cognitive or emotional) cost—is not completely right. This provides the foundation for alternative theories such as costly talk (as in Kartik, Ottaviani, and Squintani 2007; Ying Chen 2007; and Kartik 2009) or the level- k model (as in Crawford 2003; Cai and Wang 2006). The Nietzsche passage quoted above describes the cognitive load of deception and is explored in Jennifer Maria

⁶ Note that although the pupil dilation results are consistent with both the guilt and cognitive difficulty explanations, the lookup results are more consistent with the cognitive difficulty story of overcommunication, since different lookup patterns each suggest a specific (level- k) reasoning process that has a particular level of cognitive difficulty. It is not obvious how guilt alone (or variations in guilt) can produce this result.

Nuñez et al. (2005). Mark Twain also famously quipped, “If you tell the truth, you don’t have to remember anything,” indicating the memory cost of deception.⁷ The corollary principle is that if subjects want to misrepresent the state to fool receivers, they have to figure out precisely how to do so (and whether receivers will be fooled). This process is not simple and seems to leave a psychological signature in looking patterns and pupil dilation. Future theories could build in an implicit cost to lying (which might also vary across subjects and with experience) and construct richer economic theories about when deception is expected to be widespread or rare.

I. The Sender-Receiver Game

In each round of the experiments, subjects play a game of strategic information transmission, involving cheap talk (Crawford and Sobel 1982). One player always acts as the sender, and the other as the receiver. The sender’s eye movements and pupil dilation are measured with a head-mounted Eyelink II eyetracker (see Appendix: Methods). At the beginning of the round, the sender is informed about the true state of the world, which is described as a “secret” number S uniformly drawn from the state space $\mathbf{S} = \{1, 2, 3, 4, 5\}$, and is informed about the bias b , which is either 0, 1, or 2 with known probabilities. The receiver knows the bias b , but not the realization of the state S . Both players are informed in instructions about the basic structure of the game.

The sender then sends a message to the receiver, from the set of messages $\mathbf{M} = \{1, 2, 3, 4, 5\}$.⁸ After receiving a message from the sender, the receiver chooses an action from the action space $\mathbf{A} = \{1, 2, 3, 4, 5\}$. The true state and the receiver’s action determine the two players’ payoffs in points according to $u_R = 110 - 20|S - A|^{1.4}$, and $u_S = 110 - 20|S + b - A|^{1.4}$, where u_R and u_S are the payoffs for the receiver and the sender, respectively. Note that the receiver earns the most money if her action matches the true state (since her payoff falls with the absolute difference between A and S). The sender prefers the receiver to choose an action equal to $S + b$. This payoff structure is made known to both senders and receivers. Figure S1 in the supplemental Appendix shows the screen display for $b = 1$ and $S = 4$.

When the bias is large ($b = 2$), the most informative equilibrium has the sender send an uninformative message, while the receiver ignores the message and chooses $A = 3$ based on her prior beliefs (babbling equilibrium). When $b = 1$, the most informative equilibrium requires the senders to send messages $\{1\}$ when $S = 1$, and send $\{2, 3, 4, 5\}$ when S is 2–5. When $b = 1$ the receivers should choose action $A = 1$ when seeing $M = \{1\}$, and $A = 3$ or 4 when seeing $M = \{2, 3, 4, 5\}$.⁹ When $b = 0$, truth telling by choosing $M = S$ (and receivers choosing $A = M$) is the most informative equilibrium.

On the other hand, following Crawford (2003) and Cai and Wang (2006), the level- k model for the sender-receiver game starts with $L0$ senders (who has the lowest level of sophistication) simply telling the truth, and $L0$ receivers best responding to $L0$ senders by following the message. Moving up the hierarchy, $L1$ senders best respond to the $L0$ receivers by inflating the message (stating their preferred states), and $L1$ receivers best respond to $L1$ senders by discounting the

⁷ Quotation taken from Mark Twain’s Notebook, 1894. In fact, Daniel Kahneman and Beatty (1966) showed how more difficult memory tasks induced larger pupillary response. Hence, memory load could also be a channel for deception to affect pupil dilation.

⁸ Following Cai and Wang (2006), we use the message, “The number I received is X ” to eliminate possible misinterpretation of the message (which contributes to the multiple equilibria problem typical in these types of games resulting from the need to assign meaning to messages).

⁹ Thanks to David Eil for clarifying the equilibrium analysis. Due to discreteness, there is another knife-edge equilibrium with $b = 1$ that produces higher information transmission: Senders send messages $M = \{1, 2\}$ and $\{3, 4, 5\}$, while receivers choose $A = 2$ and 4. However, this equilibrium is not robust since senders who see $S = 2$ are exactly indifferent between sending $M = \{1, 2\}$ and $M = \{3, 4, 5\}$. Moreover, the main results of the paper do not change even if we consider this equilibrium (then $\text{Corr}(S, A) = 0.791$, and $u_R = 94.56$).

TABLE 1—BEHAVIORAL PREDICTIONS OF THE LEVEL- k MODEL

Sender message (condition on state)						Receiver action (condition on message)					
State	1	2	3	4	5	Message	1	2	3	4	5
$b = 0$											
<i>L0/EQ</i> sender	1	2	3	4	5	<i>L0/EQ</i> receiver	1	2	3	4	5
$b = 1$											
<i>L0</i> sender	1	2	3	4	5	<i>L0</i> receiver	1	2	3	4	5
<i>L1</i> sender	2	3	4	5	5	<i>L1</i> receiver	1	1	2	3	4
<i>L2</i> sender	3	4	5	5	5	<i>L2</i> receiver	1	1	1	2	4
<i>EQ</i> sender	4	5	5	5	5	<i>EQ</i> receiver	1	1	1	1	4
<i>SOPH</i> sender	3	4	5	5	5	<i>SOPH</i> receiver	1	2	2	2	4
$b = 2$											
<i>L0</i> sender	1	2	3	4	5	<i>L0</i> receiver	1	2	3	4	5
<i>L1</i> sender	3	4	5	5	5	<i>L1</i> receiver	1	1	1	2	4
<i>L2</i> sender	4	5	5	5	5	<i>L2</i> receiver	1	1	1	1	4
<i>EQ</i> sender	5	5	5	5	5	<i>EQ</i> receiver	1	1	1	1	3
<i>SOPH</i> sender	5	5	5	5	5	<i>SOPH</i> receiver	2	2	2	2	3

Notes: *L0* senders are truthful and *L0* receivers best respond to *L0* senders by following the message. *L1* senders best respond to *L0* receivers, while *L1* receivers best respond to *L1* senders, and so on. Note that when $b = 2$, due to discreteness both *L2* and *EQ* ($= L3$) senders best respond to *L1* receivers.

message. Such procedure is continued until we reach the most informative equilibrium prediction.¹⁰ In addition, we include a sophisticated type (*SOPH*) which best responds to the empirical distribution of opponent's behavior. This represents the highest level of strategic thinking, knowing the exact heterogeneity of opponent types and behavior. Table 1 provides the list of different level- k types for $b = 0, 1$, and 2.¹¹ Note that in our data, *SOPH* senders act like *L2* senders when $b = 1$ and act like *EQ* ($= L3$) senders when $b = 2$, both a best response to *L1* receivers.

Under both the equilibrium and level- k models, the comparative statics are similar: Information transmission decreases as the bias increases, though the level- k model still allows transmission even when the bias is so big that the equilibrium model predicts babbling (zero transmission). Informativeness is measured by the correlation between actions and the true states, and by receiver payoffs (more informative equilibria have higher expected payoffs). In addition, we assume a literal interpretation of messages and measure the "informativeness" of senders' messages by the correlation between the true states and the messages M . How "trusting" the receivers are can be measured by the correlation between the messages M they receive and the actions A they take.¹² These comparative statics predictions were tested by Dickhaut, McCabe, and Mukherji (1995), Blume et al. (1998, 2001), and Cai and Wang (2006). Overcommunication—messages are more informative

¹⁰ Note that the level- k model itself provides an equilibrium selection criterion—it selects the most informative equilibrium where senders report the upper bound of the interval of true states. This pins down both the amount of information transmitted and the language used in the message sent. In general, level- k models will provide more precision (given a particular parameter value specification) than equilibrium concepts when there are multiple equilibria. Also note that, due to signal jamming, higher level types do not simply add (or subtract) multiples of the bias. This is particularly true when approaching the upper bound of the state space. For example, when $b = 1$, *L2* senders who see $S = 3 - 5$ will all send $M = 5$ (since higher messages are not feasible), and *L2* receivers, knowing the true state is equally likely to be 3, 4, or 5, would choose $A = 4$, instead of 3 ($= 5 - 1 \times 2$).

¹¹ Cai and Wang (2006) constructed a level- k model for the case where the most informative equilibrium is babbling. Here we extend it to other biases. Also, we use the econometric methods developed by Costa-Gomes and Crawford (2006) to estimate individual types.

¹² Such a natural language interpretation is justified by Blume et al. (2001) findings that equilibrium messages tend to be consistent with their natural language meanings and is used in Cai and Wang (2006). Moreover, many behavioral theories of lying, such as Crawford (2003) and Kartik, Ottaviani, and Squintani (2007), also lead to this sort of natural language interpretation since naïve receivers would take the message at face value.

of the true state than they should be, in equilibrium—are typically found in these studies, and Cai and Wang (2006) suggest two bounded rationality explanations: the level- k model and quantal response equilibrium.

II. The Experiment

To be sure that subjects learn, and to collect a lot of trials to pool across, the same game is played 45 times among the two players with random choices of bias b (and random states) in each round. Because we could only eyetrack one or two subjects at a time, only the senders were hooked up to the mobile Eyelink eyetracker (although studying receivers' eye fixations would be useful in future work). We randomly matched six subjects into pairs using a stranger-matching protocol, with different receivers in each round (with no immediate rematching with the same receiver), and eyetracked two of the senders in each group. Values of $b = 0, 1, 2$ were used with known probabilities (0.2, 0.4, 0.4) since we are less interested in the no-bias ($b = 0$) case than in the bias ($b > 0$) cases. We also added some noise (integers -4 to $+4$ with equal probability, i.i.d. across payoff cells) to each payoff to minimize memory effects. Since the noise is small, the equilibrium remains the same. To further eliminate any memory effect, the bias parameter was not shown to the eyetracked senders on the screen, although it was mentioned in the instructions. Instead, subjects were forced to look at the payoff table to infer it. Thus, this set of experiments is called the “hidden bias–stranger” design. The results reported below focus entirely on the eye fixations and pupil dilation of the eyetracked senders, and the message choices of all senders and action choices of receivers.¹³

Subjects' choices are compared to the most informative equilibrium in the one-shot game.¹⁴ We also use predictions from a level- k model (Table 1) to estimate individual sender types with a quantal response-like “spike-logit” error structure, using the econometric analysis developed by Costa-Gomes and Crawford (2006). In particular, we assume each sender subject exactly follows a certain level- k type and plays t^k (the “spike” of probability) with probability $(1 - \varepsilon)$. With probability ε , they make mistakes following a logit error density $d^k(m, \lambda | s) = \exp[\lambda \Pi^k(m | s)] / \sum_{\mu \neq t^k} \exp[\lambda \Pi^k(\mu | s)]$, in which $\Pi^k(m | s) = \sum_{\alpha=1}^5 \pi(s, \alpha) f^k(\alpha | m)$ is the expected payoff of sending message m when the true state is s . $\pi(s, a)$ is the payoff for true state s and receiver action a , and $f^k(\alpha | m)$ is level- k sender's belief about receiver's actions (seeing each message). The likelihood for observing a level- k subject i play $m^i = \{m_g^i\}_{g \in G}$ in the set of games G (making mistakes in subset $N^{ik} \subset G$, $n^{ik} = |N^{ik}|$) is therefore $d^k(m^i, \varepsilon, \lambda) = (1 - \varepsilon)^{|G| - n^{ik}} \varepsilon^{n^{ik}} \times \prod_{g \in N^{ik}} d^k(m_g^i, \lambda | s(g))$. The level- k type distribution is $p = (p^1, \dots, p^K)$.

For each individual subject, we estimate the parameters $(p, \varepsilon, \lambda)$ that maximizes empirical log-likelihood $L^i(p, \varepsilon, \lambda | m^i) = \ln[\sum_{k=1}^K p^k d^k(m^i, \varepsilon, \lambda)]$. Note that p will be estimated to have $p^k = 1$ for some k , so estimation results for a subject could be written as $(k, \varepsilon, \lambda)$.

We also ran an earlier set of experiments using a partner protocol in which a pair of subjects played repeatedly in a fixed-role protocol where $b = 0, 1, 2$ with known equal probability. The bias parameters were always revealed (sender subjects indeed look at them), and there was no payoff perturbation. This is a simpler design to implement logistically, requiring only one eyetracked subject and his/her (open box) opponent but creates potential repeated-game effects.

¹³ Two of the 12 eyetracked subjects experienced technical difficulty during the experiment, and their data were excluded (along with the corresponding receiver subjects' choices).

¹⁴ We do not consider a possible dynamic equilibrium that might sustain higher information transmission levels. This is not a problem for $b = 0$ or 2. When $b = 2$, babbling is the only equilibrium in the one shot game, and backward induction yields the babbling equilibrium for all finitely repeated games; when $b = 0$, the one shot game equilibrium already has full information transmission, and there is no room for improvement. Also note that overcommunication is the most striking when $b = 2$. Random rematching also limits repeated-game effects.

We refer to this as the “display bias–partner” design. Results of this design are briefly discussed in comparison to that of the “hidden bias–stranger” design in Section IIID. Corresponding tables and figures are in the supplemental Appendix.

Subjects were 60 Caltech students recruited from the Social Science Experimental Laboratory subject pool. Six sessions of six subjects were randomly matched in the “hidden bias–stranger” design, and 12 pairs were run in the “display bias–partner” design. They earned between \$12 and \$27 in addition to a \$5–15 show-up fee. For the “hidden bias–stranger” sessions, we used different randomly predrawn parameters for each of the six sessions. But in the “display bias–partner” design we used the same set of randomly drawn biases and states for nine of the 12 pairs, and used two other sets of parameters for the remaining three pairs to see if there were any effects for using the same parameters.

While 60 subjects might appear to be a small sample size,¹⁵ most experimental studies with larger samples have many fewer choices per subject. The eyetracked subjects played 45 games and made a very large number of eye fixations; so we recorded a lot of data for each subject and could often draw confident statistical conclusions from these sample sizes.

III. Results

A. Comparative Statics and Behavior

What do Players Choose?—Figures 1–3 display the three dimensions of the raw choice data—states, messages, and actions—for the three bias levels $b = 0, 1$, and 2 . Each figure is a 5-by-5 display. The true states 1–5 correspond to the five rows. The sender messages 1–5 correspond to the five columns. Within each stage-message cell, there is a pie chart. The area of the pie chart in each cell is scaled by the number of occurrences for the corresponding state and message; i.e., the most common state-message pairs have the largest pies. Hence, the rows indicate senders’ behavior with respect to different states and the columns represents the “informativeness” of each message, determined by the distribution of states conditional on each particular message. Several diagonal lines connect predicted messages for various level- k types. Each pie chart also shows the distribution of actions chosen by the receiver for that state and message, using a gray scale ranging from white (action 1) to black (action 5). The average receiver action is the number inside the pie.

For example, when $b = 0$, and there is no conflict of interest, large pie charts are concentrated on the diagonal ($L0/EQ$ sender behavior), which is a visual way of showing that the senders almost always send a message corresponding to the true state. Moreover, these pie charts mostly contain the same color ranging from light (lower actions) to dark (higher actions) as the message number increases across columns, showing that the receivers’ actions are typically equal to the message. The distribution of state frequencies conditional on each message (i.e., down each column) almost degenerates into mass points of the true states, indicating nearly full information transmission. This corresponds to the (most informative) truth-telling equilibrium predicted by equilibrium theory, as well as the $L0/EQ$ type in the level- k model.

When $b = 1$, and there is an incentive to bias the message upward, the results are different. There is a large tendency for deception, which is evident from having some large pie charts off the diagonal. This departure is lopsided—only the upper diagonal of Figure 2 is populated with large pie charts.¹⁶ That is, for a given state, the most common messages are the state itself or

¹⁵ We successfully eyetracked 22 of the 60 subjects, which is considered a *large* sample size for psychophysical studies involving eyetracking.

¹⁶ Note that this one-sided deception can potentially backfire since if seeing a message 1 indicates the true state is 1, the state is less likely to be 1 when other messages were sent.

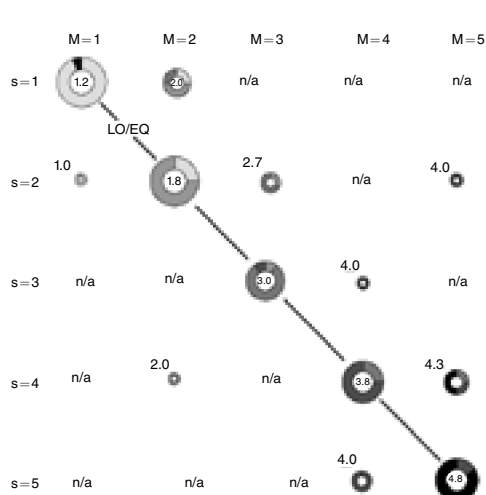


FIGURE 1. RAW DATA PIE CHARTS ($b = 0$)
(HIDDEN BIAS-STRANGER)

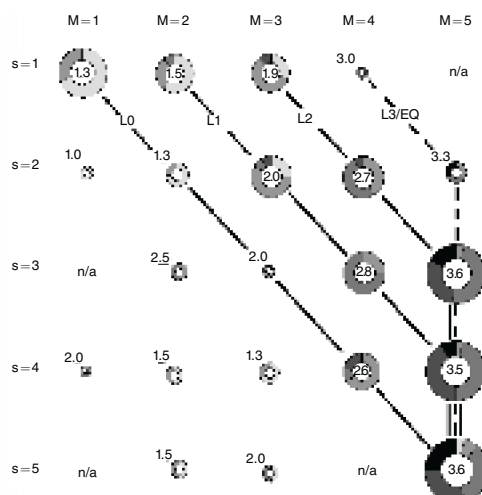


FIGURE 2. RAW DATA PIE CHART ($b = 1$)
(HIDDEN BIAS-STRANGER)

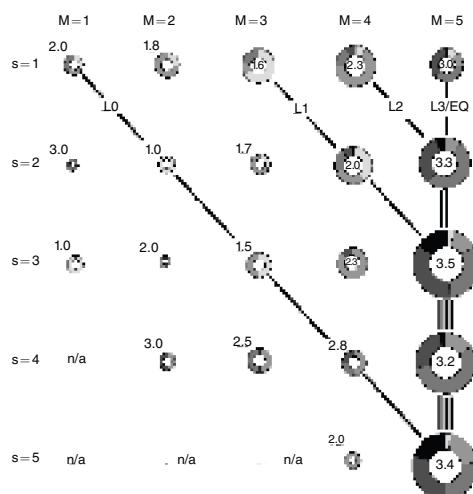


FIGURE 3. RAW DATA PIE CHART ($b = 2$)
(HIDDEN BIAS-STRANGER)

Notes: The true states are in rows, and senders' messages are in columns. Each cell contains the average action taken by the receivers and a pie chart breakdown of the actions. Actions are presented in a gray scale, ranging from white (action 1) to black (action 5). The size of the pie chart is proportional to the number of occurrences for the corresponding state and message.

higher messages (not lower messages). Furthermore, the largest pie charts of each row are mainly on the line one column or two columns to the right of the diagonal (i.e., states $S + 1$ and $S + 2$), consistent with $L1$ and $L2$ sender behavior. Within the upper diagonal, the pie chart gets darker and darker going down and right, showing how the receivers correctly increase their actions as the state and message increase. Since the conditional distribution of states (columns in Figure 2) shift from a mass point on the true state (as in Figure 1) to a distribution skewed toward state

TABLE 2—INFORMATION TRANSMISSION: CORRELATIONS BETWEEN STATES S , MESSAGES M , AND ACTIONS A

Bias	Eyetracked	$r(S,M)$		$r(M,A)$		$r(S,A)$		Predicted $r(S,A)$
0	Yes	0.92	} 0.93	0.90	} 0.92	0.86	} 0.86	1.00
	No	0.94		0.94		0.88		
1	Yes	0.68	} 0.64	0.73	} 0.71	0.53	} 0.49	0.65
	No	0.51		0.61		0.35		
2	Yes	0.41	} 0.34	0.52	} 0.58	0.34	} 0.32	0.00
	No	0.23		0.63		0.28		

Notes: In the hidden bias–stranger design, some senders' eye movements were recorded ("eyetracked") and others were not ("open box"). This comparison provides a useful test of whether obtrusively tracking a subject's eye fixations affects their behavior.

3 to 5, some information is transmitted. However, this distribution is not consistent with the $\{1\} - \{2, 3, 4, 5\}$ partition equilibrium, which requires that distributions of messages and actions for the bottom four rows (states 2–5) should all look the same.¹⁷

Finally, when $b = 2$, equilibrium theory predicts a babbling equilibrium. If subjects were playing this equilibrium, the pie-chart distributions in each column would be roughly the same (up to random sampling error of state frequencies), and the shading distributions on each pie chart would be the same. In fact, there is still a substantial amount of information transmitted, since the columns in Figure 3 do not show the same uniform distribution of state frequencies. However, many senders still sent message 5, especially for states 2–5. And a substantial amount of receivers did choose action 3, as predicted in the babbling equilibrium. This is consistent with the level- k model, since $L1$, $L2$, and EQ senders all send message 5 for states 3 to 5.

What are the Comparative Static Results?—Table 2 shows the actual amount of information transmitted, measured by the correlation between states S , actions A , and messages M . The key comparative static prediction of Crawford and Sobel (1982) holds in the data: As the bias b increases, the information transmitted decreases, measured either by the correlation $r(S, A)$ between state S and action A , or by receiver payoffs. But note that even when the bias is so large ($b = 2$) that theory predicts babbling (i.e., no correlation between S , A , and M), the correlations are still around 0.3. There are also very small learning effects: correlations and payoffs rise across trials for $b = 0$ and fall for $b > 0$ reflecting weak convergence toward equilibrium (see supplementary Appendix, Table S1). Payoffs also decline with bias b , as predicted by theory (Table 3). Data from both senders who are eyetracked and senders with "open boxes" (no eyetracking) are reported separately as a check on whether eyetracking, per se, changes behavior. There is no discernible effect of being eyetracked versus seeing all parameters ("open boxes").

When the bias b is large, information transmission is higher (measured by correlations among S , M , and A), and payoffs are higher, than predicted by equilibrium theory. These data replicate the "overcommunication" (too much truth telling) reported in Cai and Wang (2006).

¹⁷ If subjects were playing according to the partition equilibrium, column 1 should have probability 1 on state 1, and zero probability elsewhere, indicating the state being in partition $\{1\}$, while columns 2 to 5 should all have equal probability distributions (say, a mass point at 5 or $\frac{1}{4}$ each) on states 2 through 5, and zero elsewhere (indicating the state being in partition $\{2, 3, 4, 5\}$).

TABLE 3—SENDER AND RECEIVER'S PAYOFFS

Bias	Eyetracked	u_s (SD)	(combined)	u_R (SD)	(combined)	Pred. u_R (SD)
0	Yes	101.13 (18.68)	101.30 ^a (17.28)	100.85 (19.28)	101.27 ^a (17.69)	110.00 (0.00)
	No	101.89 (14.89)		102.07 (15.23)		
1	Yes	71.81 (39.56)	73.28 (37.46)	87.88 (28.63)	86.88 (27.59)	91.40 (19.39)
	No	75.44 (35.11)		84.44 (25.62)		
2	Yes	43.39 (52.17)	43.31 (52.79)	80.78 (27.17)	80.55 (27.57)	80.80 (20.76)
	No	43.21 (53.37)		80.21 (29.11)		

Note: ^a Payoffs are not exactly the same due to the random noise added and certain groups excluded.

Can Individual Players be Classified as Level- k Types?—Based on all behavioral data, we classify individual sender subjects into various types according to Table 1, assuming subjects remain the same type across different biases using the “spike-logit” estimation as in Costa-Gomes and Crawford (2006). The results are shown in Table 4. Subjects are classified as types (percentages) $L0$ – $L2$ (18 percent, 25 percent, and 25 percent), $SOPH$ (14 percent), and EQ (18 percent), with good compliance (above 60 percent, except for one).¹⁸ Individual level classifications therefore confirm that subjects are mostly choosing according to stable level- k types, as hinted by the aggregate choice data. Comparing the classification results with that of Cai and Wang (2006), we see a similar pattern (having few $L0$, mostly $L1$ beyond), although they use a more primitive way to conduct the classification.

B. Lookup Patterns

There are several goals in observing lookup patterns: First, we want to know what the aggregate lookup patterns are during the decision process. This indicates the subjects’ attention to different information and provides the basis for theorizing about subjects’ decision-making process. Moreover, since the level- k model relaxes the assumption that people hold consistent beliefs about others, beliefs about others’ beliefs, and so on, we expect the lookup patterns to indicate this. Finally, since the level- k model predictions explain individual behavior, it is natural to ask whether additional lookup data can provide more direct evidence supporting the level- k model than choices alone. In particular, we would like to ask whether individual subjects who are classified into different level- k types produce different lookup patterns matching their types.

The lookup results are organized according to the above goals as follows:

1. *Attention to Structure:* In reporting aggregate lookup counts and time spent on different parts of the screen, we expect to see different level- k subjects paying differential attention to important parameters of the sender-receiver game, such as state, bias, and payoffs.

¹⁸ Using only trials such that $b = 1, 2$ yields the exact same classification. Using a logit structure (instead of spike-logit) on $b = 1, 2$ also yields a similar distribution, in which only two subjects are classified differently: Subject #3 ($2 - 1$) is classified as $SOPH$ instead of $L1$, and subject #5 ($3 - 2$) is classified as $L2$ instead of $SOPH$. See Table S12. Note that $SOPH$ and $L2$ are almost identical, and from the lookup results below (Table 7), subject #3 has a lookup score more similar to $SOPH$ than $L1$. Finally, using a logit structure on all data adds three more $SOPH$ types ($2 - 2$, $4 - 3$, and $5 - 2$), all from “neighboring” types which often coincide with $SOPH$ (EQ , EQ , and $L2$, respectively).

TABLE 4—LEVEL- k CLASSIFICATION RESULTS

Session	ID	log L	k	Exact	lambda	Treatment
1	1	-46.23	<i>SOPH</i>	0.64	0.06	eyetracked subject #1
1	2	-25.99	<i>L1</i>	0.87	0.00	eyetracked subject #2
1	3	-15.98	<i>L2</i>	0.91	0.44	open box
2	1	-37.32	<i>L1</i>	0.60	0.52	eyetracked subject #3
2	2	-37.34	<i>EQ</i>	0.73	0.52	open box (eyetracked to round 20)
2	3	-25.70	<i>SOPH</i>	0.83	0.07	open box
3	1	-68.84	n/a	0.13	0.01	eyetracked subject #4
3	2	-17.71	<i>SOPH</i>	0.89	0.12	eyetracked subject #5
3	3	-54.73	<i>EQ</i>	0.60	0.03	open box
4	1	-50.86	<i>L1</i>	0.51	0.04	eyetracked subject #6
4	3	-25.22	<i>EQ</i>	0.82	0.48	open box
5	1	-22.26	<i>L1</i>	0.89	0.02	eyetracked subject #7
5	2	-35.77	<i>L2</i>	0.78	0.03	eyetracked subject #8
5	3	-25.17	<i>EQ</i>	0.87	0.04	open box
6	1	-16.27	<i>L2</i>	0.91	0.43	eyetracked subject #9
6	2	-42.02	<i>SOPH</i>	0.62	0.13	eyetracked subject #10
6	3	-52.17	<i>L0</i>	0.62	0.01	open box

2. *Truth Bias*: The level- k model assumes subjects best respond to perceived beliefs about their opponents' behavior, which are inconsistent with what opponents actually do.¹⁹ If senders cannot think like receivers (who do not know the true state), they would put too much attention on the payoff row corresponding to the true state, instead of treating all states equally. Hence, excessive attention to payoffs corresponding to the true state demonstrates a "curse of knowledge" and could be an attentional marker of these incorrect beliefs.

3. *Individual Level- k Type Lookup Patterns*: The level- k model assumes an anchoring $L0$ behavior of truth telling. Higher types go through beliefs about lower types until they reach their own level- k type. If this decision process is reflected in the lookup patterns, attention should be paid to payoffs corresponding to the action $A = S(L0)$, $A = S + b(L1)$, and so on, up to the corresponding level- k type for each individual subject. For example, when bias is 2, a $L2$ sender under state 2 would look at the payoffs corresponding to state 2 and action 2 (the $L0$ outcome if the message is taken literally), action 4 (the $L1$ outcome if the message is taken literally), and action 5 (the $L2$ outcome if the message is taken literally). In Table 1, this corresponds to the first three elements ($L0$ – $L2$ Senders) of the second column ($S = 2$) in the bottom panel ($b = 2$). Thus, in addition to the lookups required to figure out the bias parameter,²⁰ a level- k type sender (with truth bias) would follow the prediction of the level- k model (first $(k + 1)$ elements of column S in Table 1) up to his own level.

What are Senders Paying Attention to?—Table 5 shows the average lookup time (excluding fixations shorter than 50msec) for various numbers on the screen which are parameters of the game.²¹ Senders clearly are thinking carefully about the game because they look up the state for 0.86 seconds total (which is 3.2 fixations, about 270msec per fixation). The low time per lookup

¹⁹ If all subjects are *SOPH* who correctly best respond to others, *SOPH* behavior should coincide with equilibrium (*EQ*) behavior.

²⁰ In the hidden bias–stranger design, subjects must at least look at two payoffs to determine the bias: $A = S$ and $A = S + b$. Potentially, this S could be any state but should correspond to the true state due to the truth bias. In this case, the lookups would coincide with lookups linked to $L0$ and $L1$ thinking.

²¹ The number of separate fixations is very highly correlated with the lookup time—in no cell is the average time per fixation less than 250msec or greater than 300msec—so the number of fixations can be approximated well by dividing

TABLE 5—AVERAGE SENDER LOOKUP TIMES (IN SECONDS) ACROSS GAME PARAMETERS

Bias b	Response time		State	Sender payoffs	Receiver payoffs	Sender-to- receiver ratio
	Periods 1–15	Periods 31–45				
0	9.78	7.24	0.83	2.93	1.71	1.72
1	11.77	8.76	0.81	3.80	2.66	1.43
2	16.84	8.99	0.91	4.67	3.26	1.43
all	13.47	8.52	0.86	3.99	2.72	1.47

is a reminder that the eye glances around very rapidly, making frequent quick fixations, as is typical of other tasks including reading text passages.

Senders also look at their own payoffs longer. In particular, subjects look at their own (sender) payoffs at least 40 percent more than receiver payoffs. This difference is surprising since senders need to look carefully at receiver payoffs in order to determine the bias. Note that the ratio of lookup time for sender and receiver payoffs is the same for a small bias ($b = 1$) and large bias ($b = 2$). For $b = 2$, which creates the most scope for guilt to constrain deception, we divide senders into those who looked more often at receiver payoffs, and those who looked less often (relative to the median sender-receiver looking ratio). Importantly, the high receiver-lookup group is actually more deceptive than the low group, which is inconsistent with the guilt hypothesis that the more one cares about others' payoffs and looks at them, the less one should deceive. For the high group, the correlation between states and messages is 0.55, and the average *LIE_SIZE* ($|M - S|$) is 0.88; for the low group, the correlation is 0.69, and the average *LIE_SIZE* is 0.71.

Note that there is a reduction in total looking times across trials, about 35 percent less in later periods (31–45) than in earlier periods (1–15) (see Table 5), and this reduction is similar across bias levels and treatments (Table S7).

Do Senders have a “Curse of Knowledge”?—Table 6 shows that subjects look about five times longer at payoffs in the rows corresponding to the true state than they look at payoffs in rows corresponding to each of the four other states. When the bias is 0 this fixation on the actual state is understandable (and subjects typically choose message $M = S$), but the disproportionate attention to actual state payoffs is comparable when there is a bias of $b = 1$ or 2. This result indicates that subjects do not “think in others’ shoes” and cannot fully think like a receiver (who does not know the true state). Note that Table 6 suggests lookups might have statistical power to detect the actual state (i.e., to detect lies in which the message M deviates from the true state S). That is, a receiver who had online sender looking statistics might be able to predict what the actual state was rather reliably. This possibility is explored in Section IV.

Do Senders Follow Level- k Predictions of Lookups?—Tables 5 and 6 show there is a strong bias for senders to look more at payoffs from the state they know to be the true one. More detailed information about looking patterns across state-action pairs is conveyed by the icon graph in figures 4–7 (developed by Johnson et al. 2002). For brevity we show only data from

the Table 5 figures by 270 msec. Both fixations and lookup time are reported in the supplemental Appendix (Table S10 and S11).

TABLE 6—AVERAGE LOOKUP TIME PER ROW DEPENDING ON THE STATE

Bias b	True state rows	Other state rows	True-to-other ratio
0	2.76	0.47	5.89
1	3.88	0.64	6.02
2	4.29	0.91	4.70
overall	3.83	0.72	5.33

trials with positive biases for subjects classified as $L1$ and $L2$ (aggregate data are in the supplemental Appendix, Figures S6–S9).

Each box in figures 4–7 represents the attention paid to the payoff corresponding to a different state-action combination. Panels A and B represent attention to the sender payoff boxes and the receiver payoff boxes, respectively. The width of the box is a linear function of the average number of fixations on that box. The height of the box is a linear function of the average total looking time in that box. Boxes which are wide and tall were looked at repeatedly (wide) and for a longer time (tall). The vertical bars in the first columns represent the sum of looking time across each row. Longer bars represent longer time for all state-action boxes in that state. The “ruler” in the upper right shows the scale of looking time and number of fixations that can be used to “measure” each box.

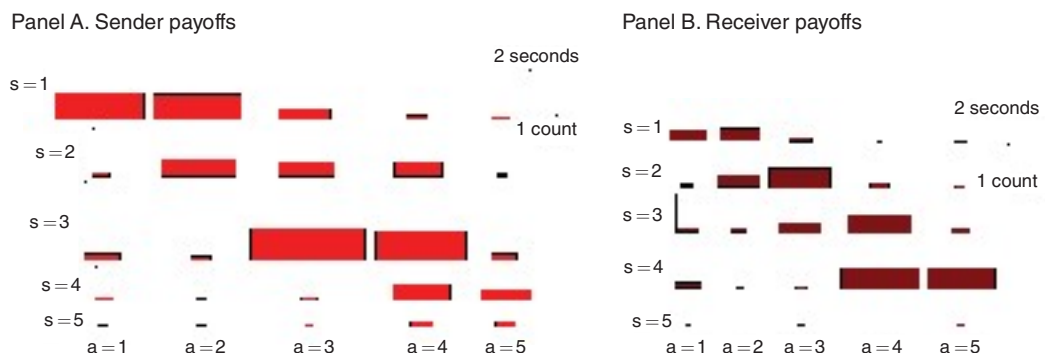
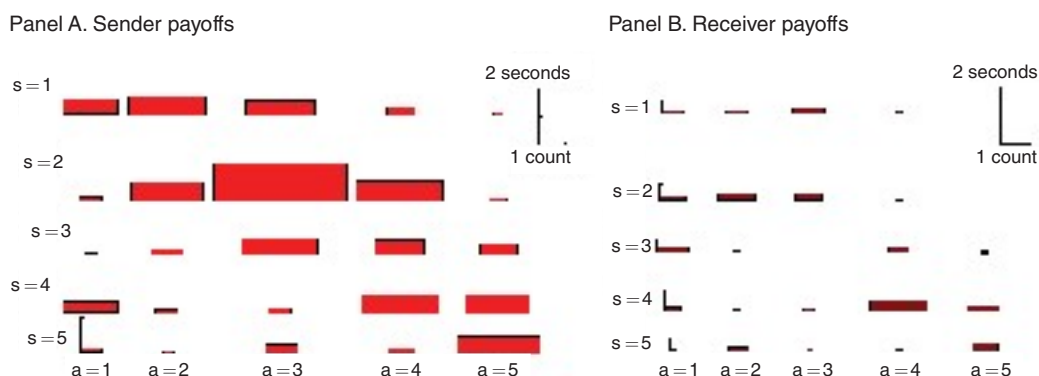
Figure 4 shows the icon graph for bias $b = 1$ when the subject is classified as $L1$.²² The first thing to notice is that subjects spend much more time looking at their own payoffs (Figure 4, panel A) than the payoffs of receivers (Figure 4, panel B), as the Table 5 statistics show. Subjects’ lookups are also more frequent and longer for actions that are equal to the actual state S or $S + 1$. This corresponds to the first two rows ($L0$ and $L1$ senders) of the top panel ($b = 1$) in Table 1, as well as the lookups to determine the bias.

Figure 5 shows the lookup icon graphs for bias $b = 1$ when the subject is classified as $L2$. Senders again look at their own payoffs more often than their opponents’ payoffs. When the state S is 1 – 3 they tend to look at their payoffs from actions corresponding to S , $S + 1$, and $S + 2$. This is consistent with the first three rows ($L0$, $L1$, and $L2$ senders) of the top panel ($b = 1$) in Table 1. However, when the state is 4 or 5 this pattern crumbles as states $S + 2$ and $S + 1$ do not exist; then lookup patterns resemble $L1$ lookups.

Similar patterns arise when $b = 2$ as well. Figures 6 and 7 show the lookup icon graphs for bias $b = 2$ when subjects are classified as $L1$ and $L2$, respectively. As the level- k model predicts, subjects look at payoffs corresponding to the first two or three rows ($L0 - L1$ or $L0 - L2$ senders) of the bottom panel ($b = 2$) in Table 1.

If we calculate the linear measure of predictive success (Reinhard Selten 1991), a subject who is classified as a certain level- k type almost always has the highest score for the corresponding lookups of the same type. In particular, let x equal the “hit rate,” the proportion of lookups in a period that fall in the target cells, and let a equal the proportional area of the target cells. Then the linear measure (LM score) equals $x - a$, the proportional hit rate minus the proportional area. This measure controls for the size of the predicted lookup area and takes a value of zero when subjects randomly scan the entire screen. Table 7 presents each subject’s LM score for various types. Among all the six subjects classified as $L1$ and $L2$ subjects, only one (subject #8) has another type’s LM score

²² When the bias $b = 0$ the looking data are very clear: Subjects look almost exclusively at their own payoffs corresponding to the actual state S and corresponding receiver action A , and they look at the receiver payoffs from the same $S - A$ pair less often than they look at their own payoffs ($1/2$ in the display bias-partner design and $2/3$ in the hidden bias-stranger). See Figure S5.

FIGURE 4: LOOKUP ICON GRAPH FOR $b = 1$, HIDDEN BIAS-STRANGER, TYPE = $L1$ FIGURE 5: LOOKUP ICON GRAPH FOR $b = 1$, HIDDEN BIAS-STRANGER, TYPE = $L2$

Notes: Each row reports the lookup counts and time for the “true state row” corresponding to the given true state. The width of each box is scaled by the number of lookups and the height by the length of lookups (scaled by the little “ruler” in the upper right corner). The vertical bar on the first column icon represents the total lookup time summed across each row.

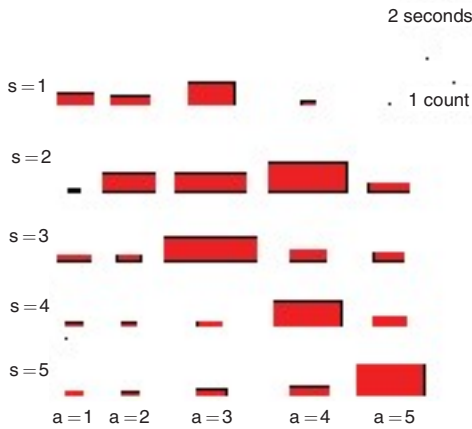
slightly higher (0.268 versus 0.259, less than 0.01) than the score corresponding to his classification based on choices. Moreover, this subject would be classified as *SOPH* under the logit specification. Regarding *SOPH* subjects, it is not clear theoretically what their lookup patterns would be. But, the low *LM* scores do show that they do not look like *L1*, *L2*, or *EQ*.

C. Pupil Dilation

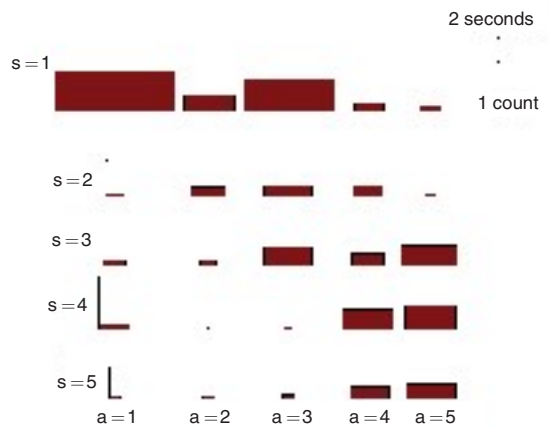
As noted in the introduction, pupils dilate when people are aroused or make cognitively difficult decisions. Our first exploratory step is to treat pupil dilation as a dependent variable and see whether the degree of behavioral deception by the sender is correlated with pupil dilation. It may be that pupil dilation is so poorly measured, or so weakly linked to deception, that there is no reliable correlation. However, we see that deception is reliably correlated with pupil dilation.

To correlate pupil dilation with senders' messages, average pupil sizes are calculated for various time periods before and after the sender's message decision. Then, we try to predict averaged

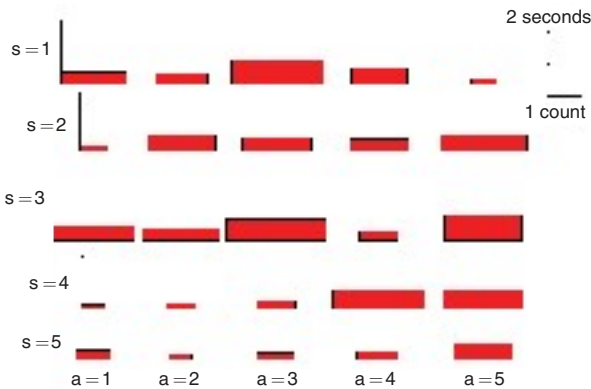
Panel A. Sender payoffs



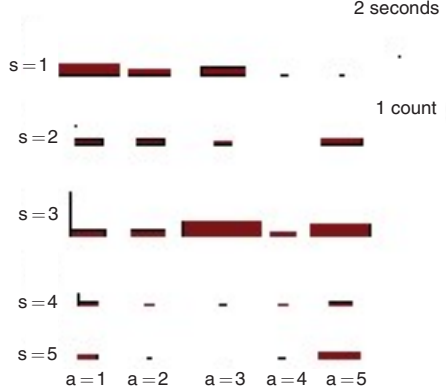
Panel B. Receiver payoffs

FIGURE 6: LOOKUP ICON GRAPH FOR $b = 2$, HIDDEN BIAS–STRANGER, TYPE = L1

Panel A. Sender payoffs



Panel B. Receiver payoffs

FIGURE 7: LOOKUP ICON GRAPH FOR $b = 2$, HIDDEN BIAS–STRANGER, TYPE = L2

Notes: Each row reports the lookup counts and time for the “true state row” corresponding to the given true state. The width of each box is scaled by the number of lookups and the height by the length of lookups (scaled by the little “ruler” in the upper right corner). The vertical bar on the first column icon represents the total lookup time summed across each row.

pupil dilation using the bias b and the amount of deception (measured by the absolute distance between states and messages, $|M - S|$).

To record their message M , senders are instructed to look at a series of decision boxes on the right side of the screen, which contain the numbers 1 to 5 (corresponding to the possible numerical messages). The software records a decision after the subject has fixated on a single decision box for 0.8 seconds—that is, the subjects choose by using their eyes, not their hands.²³

²³ Allowing eye fixations to determine actual choices is widely used in research with monkeys. For humans, making choice hands free is an advantage if psychophysiological measurements are being recorded simultaneously (e.g., gal-

TABLE 7—INDIVIDUAL LOOKUP LINEAR MEASURE SCORES FOR VARIOUS LEVEL- k TYPES

Type	Subject ID	$L1$	$L2$	$L3/EQ$
$L1$	#2 (1 – 2)	<u>0.24</u>	0.22	0.19
	#3 (2 – 1)	<u>0.16</u>	0.15	0.14
	#6 (4 – 1)	<u>0.26</u>	0.24	0.18
	#7 (5 – 1)	<u>0.41</u>	0.33	0.28
	Average	<u>0.27</u>	0.23***	0.19†
$L2$	#8 (5 – 2)	<u>0.27</u>	<u>0.26</u>	0.21
	#9 (6 – 1)	0.22	<u>0.24</u>	0.19
	Average	0.24	<u>0.25</u>	0.20**
$SOPH$	#1 (1 – 1)	<u>0.17</u>	0.16	0.13
	#5 (3 – 2)	<u>0.16</u>	0.15	0.11
	#10 (6 – 2)	<u>0.21</u>	0.13	0.07
	Average	<u>0.18</u>	0.15	0.10

Note: Highest lookup scores underlined. Lookup scores if choice classifications correspond to lookups **boldfaced**. Note that they almost always coincide for $L1$ and $L2$ types.

, * and † denotes $p < 0.05$, $p < 0.01$, $p < 0.0001$ for signed rank sum test using both own and other cells for each state, each bias, and each subject (of that type) with total lookup time > 1 sec.

Since there is a time lag of at least 0.8 seconds between the instant subjects “made up their minds” and the recording of this decision,²⁴ the *decision time* is defined as the first time subjects view any of the boxes in the decision boxes area, (provided they continue to look at the decision box area for more than 98 percent of the time until the software records a decision).

Average pupil sizes are regressed on the amount of deception for different biases, the absolute size of the deception ($LIE_SIZE = |M - S|$), and bias and state dummies, controlling for subject random effects and individual learning trends (picked up by round number and squared round number variables interacted with individual dummies). All standard errors are robust standard errors clustered at the individual level. The specification is:

$$(1) \quad PUPIL_i = \alpha + \sum_{b=0}^2 \beta_{1b} LIE_SIZE \times BIAS_b + \sum_{b \neq 2} \beta_{2b} \times BIAS_b + \sum_{s \neq 3} \beta_{3s} STATE_s + \sum_{k=1}^K (\gamma_{k,1} ROUND \times SUBJ_k + \gamma_{k,2} ROUND^2 \times SUBJ_k) + \varepsilon$$

where the error term ε has elements $\varepsilon_{kt} = u_k + \eta_{kt}$ (subject random effects), and

$PUPIL_i$ = Average pupil (area) size at time frame i : 1.2 to 0.8 seconds, 0.8 to 0.4 seconds, 0.4 to 0 seconds before, and 0 to 0.4 seconds, 0.4 to 0.8 seconds after the decision time.²⁵ Here, we

vanic skin conductance on the palms, heart rate) since even small hand movements add noise to those measurements.

²⁴ This time lag can be longer if the subject is not perfectly calibrated and, hence, needs extra time to perform the required fixation. Another possible situation is when the subject “changed her mind” and looked at different decision boxes.

²⁵ Hence, we are aggregating 100 observations into one data point when averaging for each 400-millisecond interval. Rounds with very short response time are discarded if $PUPIL_i$ cannot be calculated.

TABLE 8—PUPIL SIZE REGRESSIONS FOR 400 MSEC INTERVALS

Y	$PUPIL_i$	−1.2~ −0.8 sec	−0.8~ −0.4 sec	−0.4~ 0.0 sec	0.0~ 0.4 sec	0.4~ 0.8 sec
constant	α	107.27 (2.81)	108.03 (2.55)	106.19 (2.57)	109.56 (2.05)	108.67 (2.16)
$LIE_SIZE \times BIAS_b$ interactions	β_{10}	2.83 (1.85)	2.36 (2.23)	3.07 (2.46)	5.35*** (1.76)	5.57** (2.19)
	β_{11}	−1.02 (1.26)	−0.46 (1.31)	−0.36 (1.28)	2.16* (1.21)	2.64** (1.15)
	β_{12}	2.06** (0.86)	1.52* (0.79)	1.47** (0.75)	1.83** (0.75)	2.00*** (0.74)
	N	414	415	414	415	414
	χ^2	323.86	235.43	194.40	258.49	352.49
	R^2	0.291	0.299	0.263	0.365	0.438

Notes: Robust standard errors in parentheses. (Dummies for biases, states, individual subjects and individual learning trends are included in the regression, but results are omitted.)

[†] Significant at the 0.1 percent level.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

normalize each individual's average pupil size to 100.²⁶

LIE_SIZE = The “size” of the lie or the amount of deception, measured by the absolute distance between states and messages, $(|M - S|)$.

$BIAS_b$, $STATE_s$, $SUBJ_k$ = Dummy variables for the bias b , true state s , and subject k

$ROUND$ = Round number t

The parameter α is the average pupil size. The β_1 coefficients give us the effect of deviating from reporting the true state (deceiving more) under different bias levels. The coefficients β_{2b} and β_{3s} give us the pure effects of different biases b (relative to $b = 2$) and states (relative to $S = 3$) which might influence dilation, and $\gamma_{k,1}$, $\gamma_{k,2}$ capture (individual) linear and quadratic learning effects.

Look first at the coefficients on the amount of deception in Table 8, interacted with bias (denoted β_{1b} where b is the bias parameter). Immediately after the decision is made (0 seconds to 0.4 seconds and 0.4 to 0.8 seconds later), the coefficients are significantly higher at about 2 percent for all biases. Sending less accurate messages is therefore correlated with pupil dilation when $b = 1$ or $b = 2$. Before the decision is made, the pupil dilation difference is still at 1.5–2 percent (though less significant) when $b = 2$.

Note that the bias condition by itself does not generate pupil dilation (i.e., nearly all the coefficients β_{2s} are insignificant and are omitted from Table 8). This finding implies arousal or cognitive difficulty is created by sending deceptive messages in bias conditions, not by bias per

²⁶ Pupil sizes are measured by area, in relative terms. Absolute pixel counts have little meaning since they vary by camera positions, contrast cutoffs, etc., which depend on individual calibrations. Hence, the eyetracker scales it to a pupil size measurement between 800–2000. Here, we normalize all observations by the average pupil size of each subject throughout the entire experiment and present all results in percentage terms. (To avoid potential bias created by eyetracker adjustments, all between-round adjustment stages were excluded when doing this normalization.) Therefore, “100” means 100 percent of an individual subject's typical pupil size.

se. Furthermore, these basic patterns are reproduced when we divide the samples into two halves and compare them, which provides some assurance of statistical reliability.²⁷

D. Results of the Display Bias–Partner Design

The supplemental Appendix reports results analogous to those in Tables 2–6, Table 8 and Figures 2–4 for the display bias–partner condition (Tables S2–S6 and S8, Figures S2–S4). Compared to the hidden bias–stranger condition there is more overcommunication (correlations of M and S around 0.5 even when $b = 2$) and more low-type classification (one third $L0$ types). These differences are probably due to the repeated game effects created by the partner matching. Subjects do also look at the bias parameter when it is available, but they look less often at receiver payoffs (which they need not look at to figure out what the bias b is).

The pupil dilation results are much stronger than in the hidden bias–stranger design. The coefficients on pupil dilation predicting the amount of deception are 2.8–4.5 percent, and are significant in all 400 millisecond intervals from $-1,200$ milliseconds to $+800$ milliseconds (where zero is the decision time). It is likely that the display bias–partner design is less demanding cognitively and lowered baseline pupil dilation. In fact, the increase in predictive power here could be construed as consistent with the cognitive difficulty story because showing the bias parameter and eliminating noise from the payoffs make the display bias–partner design easier in general. This simplification could decrease the baseline pupil dilation of truth telling in all conditions, which makes any additional dilation from deception easier to detect. Running similar regressions shows that using criteria of 99, 95, or 90 percent all yield similar results, though slightly weaker.

IV. Lie-Detection and Prediction

As noted, one goal of measuring eyetracking is to see whether these behavioral measures enable us to improve upon predictions of theory. This section reports whether using eyetracking data helps predict deception and uncover the underlying true states. The ability to detect private information in this way could eventually have many practical applications. And since private information often undermines efficiency, the ability to detect private information could be Pareto improving in some settings.

Here, we ask how well receivers could predict the true state using *only* messages and lookup patterns (and how much they could earn by using those predictions). That is, we pretend we don't know the true state for predictive purposes, forecast it from observables, then use knowledge of the true state to evaluate predictive accuracy. We focus only on $b = 1$ and $b = 2$ since truth telling is so prominent when $b = 0$.

For the dependent variable $STATE_j$, from 1–5, we ran an ordered logit regression

$$\log[\Pr(STATE \geq j)] = \theta_j + \sum_{b=1,2} (\beta_{1b}MESSAGE + \beta_{2b}ROW_{self} + \beta_{3b}ROW_{other})BIAS_b + \varepsilon$$

where lookups are consolidated into two integer variables, ROW_{self} and ROW_{other} , which are the states corresponding to the own (or opponent) payoff row which has the longest total lookup time of all rows.

²⁷ Because we measured eyetracking and pupil dilation from ten senders, it is useful to check how reliable these results are in two subsamples of five subjects each. The 400-msec intervals from $+0.4$ to $+0.8$ secs after decision time gives the highest R^2 's so we compare those. The β_{1b} coefficients across bias levels ($b = 0, 1, 2$) are the most important. They are 6.35**, 2.40, 2.11 for the first five subjects and 6.11***, 4.14***, and 3.00[†] for the second five subjects. For other intervals, as predictive power (R^2) falls the reliability across the two subsamples falls too, but the coefficient signs are almost always the same in the two subsamples and magnitudes are typically reasonably close.

TABLE 9—PREDICTING TRUE STATES (*Resampling 100 times*) (*SE in parentheses*)

X	Hidden bias–stranger	
MESSAGE \times BIAS = 1	β_{11}	0.46 [†] (0.12)
MESSAGE \times BIAS = 2	β_{12}	0.42 [†] (0.09)
ROW _{self} \times BIAS = 1	β_{21}	1.07 [†] (0.24)
ROW _{self} \times BIAS = 2	β_{22}	1.72 [†] (0.20)
ROW _{other} \times BIAS = 1	β_{31}	1.27 [†] (0.22)
ROW _{other} \times BIAS = 2	β_{32}	0.44 [†] (0.15)
total observations N^a	357	
N used in estimation	238.3	
N used to predict	118.7	
	Actual data	Holdout sample
Percent of wrong prediction ($b = 1$)	58.5	28.9
Percent of errors of size (1, 2, 3+) ($b = 1$)	(61, 28, 11)	(79, 19, 2)
Average predicted payoff ($b = 1$) ^b	87.5 (28.8)	101.7 [†] (2.1)
Percent of wrong prediction ($b = 2$)	77.9	37.9
Percent of errors of size (1, 2, 3+) ($b = 2$)	(60, 30, 10)	(72, 24, 4)
Average predicted payoff ($b = 2$) ^b	80.9 (26.9)	98.0 [†] (2.2)

Notes:^a Observation with less than 0.5 seconds lookup time and without the needed pupil size measures are excluded.^b Two-sample *t*-test conducted against the actual payoffs of receivers in the experiment who are paired with eyetracked senders.[†] Significant at the 0.1 percent level.

**Significant at the 5 percent level.

The coefficients β_{1b} represents the information about the state contained in the message the coefficient, β_{2b} measures the effects of the “most viewed row” of one’s own payoffs (i.e., the state number corresponding to the row that is viewed for the longest time), and β_{3b} represents the effects of the “most viewed row” of the opponent’s payoffs. The θ_j are state-specific constants.

To evaluate how well these specifications could predict new data, out-of-sample validation is used. Each observation is used with probability 2/3 to estimate the model, then the model forecasts on a holdout sample of the remaining 1/3 of the data. For each holdout observation, the estimated logit probabilities are used to calculate the expected state, which is rounded to the nearest integer to make a precise single-state prediction. This partial estimation-prediction procedure is performed for 100 random samples of the data. Average β s and (bootstrap) standard errors across the 100 resamplings are reported in Table 9.

The significance of β_{1b} in Table 9 indicates that the messages are informative about the states (as analyses reported above established). A smaller message indicates a smaller true state, even though standard game theory predicts that little information should be transmitted by the message (β_{1b} should be zero when $b = 2$).

The lookup data are significantly correlated with states as well. The coefficients β_{2b} , on the most-viewed *own* row variables, and the coefficients β_{3b} , on the most-viewed *other* row variables, are all positive and significant. Thus, lookup data improve predictability *even when controlling for the message*. For example, if the message is 4, but the lookup data indicate the subject was looking most often at the payoffs in the row corresponding to state 2, then the model could predict that the true state is 2, not 4. This is to be expected, since Table 6 indicates subjects look at the payoff rows corresponding to the true state five times more than other rows. However, note that this sort of prediction can come only from a setting in which attention is measured. In addition, if senders knew their eye movements were being used to infer the state, they could of course change their lookups and undermine the predictions.

The error rates in predicting states in the holdout sample are never greater than 40 percent. (Keep in mind that the error rates in equilibrium would be 60 percent and 80 percent.) Most of the wrong predictions from the logit model (70 percent) miss the state by one. The model accuracy is also substantially better than the actual performance of the receiver subjects in our experiments: Subjects “missed” (chose $A \neq S$) 58.5 percent of the time when $b = 1$ and missed 77.9 percent for $b = 2$.

An interesting calculation is how much these predictions could potentially add to the receiver payoffs (cf. “economic value” in Camerer, Ho, and Chong 2004). For biases $b = 1$ and $b = 2$, the average actual payoffs earned by receivers who faced eyetracked senders in the random sample were 87.5 and 80.9. If receivers had based their predictions on the models estimated in Table 8 and chose an action equal to the model predicted state (for the holdout sample), their expected payoffs would be 101.7 for $b = 1$ and 98.0 for $b = 2$. Since the maximum payoff possible is 110, this is a large economic value of about 60 percent of the increment between actual and maximum payoffs.²⁸ In fact, these payoffs are already close to what subjects actually earn when $b = 0$ and there is no bias (100.85 in Table 3).²⁹ These economic value statistics suggest that it could be possible to almost erase the cost to receivers of not knowing the true state just by looking at attention along with messages.

An important caveat to these analyses is that we do not know what would happen if the senders knew that their pupil dilation and lookups were being used to predict the true state. Senders would try to signal-jam by looking at the payoffs corresponding to their message more often (a kind of faked sincerity), but it is possible that excessive pupil dilation or more detailed lookup patterns could distinguish such signal jamming. Putting senders under time pressure might also make it difficult for them to use such a deliberately misleading strategy. In any case, such experiments are natural follow-ups and could be easily done.

V. Conclusion

This paper reports experiments on sender-receiver games with an incentive gap between senders and receivers, such as managers or security analysts painting a rosy picture about a firm’s earnings prospects. Senders observe a state S , an integer 1–5, and choose a message M . Receivers observe M (but not S) and choose an action A . The sender prefers that the receiver choose an action $A = S + b$, which is b units higher than the true state, where $b = 0$ (truth telling is optimal), or $b = 1$ or $b = 2$. The bias number b is the size of the incentive gap. Receivers know the payoff structure, so they should be suspicious of inflated messages M .

Our experimental results show “overcommunication”—messages are more informative of the state than they should be, in equilibrium. This result is consistent with a level- k model of communication anchored at level-0 truth telling. To explore the cognitive foundations of overcommunication, eyetracking was used to record what payoffs the sender subjects are looking at, and how widely their pupils dilate (expand) when they send a message.³⁰

²⁸ For $b = 1$, economic value = $(101.7 - 87.5)/(110 - 87.5) = 63$ percent. For $b = 2$, economic value = $(98 - 80.9)/(110 - 80.9) = 59$ percent. Analogous out-of-sample prediction results for the display bias–partner design are reported in supplemental Appendix Table S9. Results are weaker than that of the hidden bias–stranger design, having a modest economic value of 44 and 24 percent.

²⁹ Such gains in the hidden bias–stranger design are not surprising since subjects are forced to look at the payoff table to discover the bias parameter, and they focus disproportionately on the “true state” row along the way.

³⁰ The sender-receiver paradigm also expands the quality of research on lie detection in general: Deception in these games is spontaneous and voluntary (most studies use instructed lying); and both players have a clear and measurable financial incentive to deceive and to detect deception (most studies lack one or both types of incentives).

The lookup data show that senders look disproportionately at the payoffs corresponding to the true state. They do not appear to be thinking strategically enough by putting themselves “in the other’s shoes”; looking and choice are roughly consistent with a cognitive hierarchy specified by the level- k model, starting from truth telling.

Senders’ pupils also dilate when they send deceptive messages ($M \neq S$) and dilate more when the deception $|M - S|$ is larger in magnitude. In a simpler pilot design that is prone to memory and repeated game effects (the display bias–partner design), these behavioral results are also present. Together, these data are consistent with the underlying assumptions of the level- k model, and that figuring out how much to deceive another player is cognitively difficult. Gneezy (2005) and Sjaak Hurkens and Kartik (2008) found that changing the known costs to others from deception lowers deception by subjects, suggesting that guilt plays a role in limiting deception. Complementing this finding, we find that guilt does not appear to be the sole driver of overcommunication, because senders who look at receiver payoffs more often are also more deceptive. In fact, Santiago Sánchez-Pagés and Marc Vorsatz (2007) show that overcommunication is caused by the tension between normative social behavior and incentives for lying.

Furthermore, combining sender messages and lookup patterns, one can predict the true state and lower the miss rate of subjects by one half. Those predictions increase receiver payoffs up to 16–21 percent, which is an economic value of more than half of the maximum increase above what subjects actually earn in the experiment.

There are many directions for future research.

Within this paradigm, eyetracking receivers would be useful for establishing their degree of strategic sophistication in making inferences from messages. More generally, economic theories often talk vaguely about the costs of decision making or difficulty of trade-offs. Pupil dilation gives us one way to start measuring these costs.

Many economic models also specify a cognitive algorithm that maps acquired information into choices (e.g., dynamic programming applications which require looking ahead). The idea of allocating attention has itself gotten attention in economics (Della Vigna 2009) and in macroeconomic studies of “rational inattention” (e.g., Christopher Sims 2006). In both cases, measuring attention directly through (now video-based) eyetracking could improve tests of theories which make predictions about both attention and choice, and how they interact, as in previous mouse-tracking studies, such as Costa-Gomes, Crawford, and Broseta (2001), Johnson et al. (2002), and Costa-Gomes and Crawford (2006). Given the novelty of using these two methods in studying games, the results should be considered exploratory and simply show that such studies can be done and can yield surprises (e.g., the predictive power of lookups and pupil dilation for inferring private state information).

In the realm of deception, two obvious questions for future research are: Are there substantial individual differences in the capacity or willingness to deceive others for a benefit? And, can experience teach people to be better at deception, and at detecting deception? Both questions are important for extrapolating these results to domains in which there is self-selection and possibly large effects of experience (e.g., used-car sales or politics). In other domains of economic interest, the combination of eyetracking and pupil dilation could be applied to study any situation in which the search for information and cognitive difficulty are both useful to measure, such as “directed cognition” (Gabaix et al. 2006), perceptions of advertising and resulting purchase, and attention to trading screens with multiple markets (e.g. with possible arbitrage relationships).

REFERENCES

- Aboyoun, Darren, and James M. Dabbs, Jr. 1998. “The Hess Pupil Dilation Findings: Sex or Novelty?” *Social Behavior and Personality*, 26(4): 415–20.

- Beatty, Jackson.** 1982. "Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources." *Psychological Bulletin*, 91(2): 276–92.
- Bernheim, B. Douglas.** 2008. "The Psychology and Neurobiology of Judgment and Decision Making: What's in it for Economists?" In *Neuroeconomics: Decision Making and the Brain*, ed. Paul W. Glimcher, Colin F. Camerer, Ernst Fehr, and Russell A. Poldrack, 115–26. Amsterdam: Elsevier.
- Berrien, F. K., and G. H. Huntington.** 1943. "An Exploratory Study of Pupillary Responses During Deception." *Journal of Experimental Psychology*, 32(5): 443–49.
- Bhatt, Meghana, and Colin F. Camerer.** 2005. "Self-Referential Thinking and Equilibrium as States of Mind in Games: MRI Evidence." *Games and Economic Behavior*, 52(2): 424–59.
- 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(5): 1323–40.
- 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(1): 79–120.
- Bradley, M. T., and Michel P. Janisse.** 1979. "Pupil Size and Lie Detection: The Effect of Certainty on Detection." *Psychology: A Journal of Human Behavior*, 16(4): 33–39.
- Bradley, M. T., and Michel P. Janisse.** 1981. "Accuracy Demonstrations, Threat, and the Detection of Deception: Cardiovascular, Electrodermal, and Pupillary Measures." *Psychophysiology*, 18(3): 307–15.
- Brennan, Michael J.** 2004. "How Did It Happen?" *Economic Notes*, 33(1): 3–22.
- Bull, Ray, and Gabriel Shead.** 1979. "Pupil-Dilation, Sex of Stimulus, and Age and Sex of Observer." *Perceptual and Motor Skills*, 49(1): 27–30.
- Cai, Hongbin, and Joseph Tao-Yi Wang.** 2006. "Overcommunication in Strategic Information Transmission Games." *Games and Economic Behavior*, 56(1): 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): 861–98.
- Camerer, Colin F., Eric J. Johnson, Talia Rymon, and Sankar Sen.** 1993. "Cognition and Framing in Sequential Bargaining for Gains and Losses." In *Frontiers of Game Theory*, ed. Ken Binmore, Alan Kirman, and Piero Tani, 27–47. Cambridge, MA: MIT Press.
- Chapman, C. Richard, Shunichi Oka, David H. Bradshaw, Robert C. Jacobson, and Gary W. Donaldson.** 1999. "Phasic Pupil Dilation Response to Noxious Stimulation in Normal Volunteers: Relationship to Brain Evoked Potentials and Pain Report." *Psychophysiology*, 36(1): 44–52.
- Chen, Ying.** 2007. "Perturbed Communication Games with Honest Senders and Naive Receivers." Unpublished.
- Costa-Gomes, Miguel A., and Vincent P. Crawford.** 2006. "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study." *American Economic Review*, 96(5): 1737–68.
- Costa-Gomes, Miguel A., Vincent P. Crawford, and Bruno Broseta.** 2001. "Cognition and Behavior in Normal-Form Games: An Experimental Study." *Econometrica*, 69(5): 1193–235.
- Crawford, Vincent P.** 2003. "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions." *American Economic Review*, 93(1): 133–49.
- Crawford, Vincent P.** 2008. "Look-ups as the Windows of the Strategic Soul." In *The Foundations of Positive and Normative Economics: A Handbook*, ed. Andrew Caplin and Andrew Schotter, 249–80. Oxford: Oxford University Press.
- Crawford, Vincent P., and Joel Sobel.** 1982. "Strategic Information Transmission." *Econometrica*, 50(6): 1431–51.
- Della Vigna, Stefano.** 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature*, 47(2): 315–72.
- Dickhaut, John W., Kevin A. McCabe, and Arijit Mukherji.** 1995. "An Experimental Study of Strategic Information Transmission." *Economic Theory*, 6(3): 389–403.
- Dionisio, Daphne P., Eric Granholm, William A. Hillix, and William F. Perrine.** 2001. "Differentiation of Deception Using Pupillary Responses as an Index of Cognitive Processing." *Psychophysiology*, 38(2): 205–11.
- Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg.** 2006. "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model." *American Economic Review*, 96(4): 1043–68.
- Gneezy, Uri.** 2005. "Deception: The Role of Consequences." *American Economic Review*, 95(1): 384–94.
- Goldwater, Bram C.** 1972. "Psychological Significance of Pupillary Movements." *Psychological Bulletin*, 77(5): 340–55.
- Hall, Brian J., and Kevin J. Murphy.** 2003. "The Trouble with Stock Options." *Journal of Economic Perspectives*, 17(3): 49–70.

- Heilveil, Ira.** 1976. "Deception and Pupil Size." *Journal of Clinical Psychology*, 32(3): 675–76.
- Hicks, Robert A., Tom Reaney, and Lynn Hill.** 1967. "Effects of Pupil Size and Facial Angle on Preference for Photographs of a Young Woman." *Perceptual and Motor Skills*, 24(2): 388–90.
- Hurkens, Sjaak, and Navin Kartik.** 2009. "Would I Lie to You? On Social Preferences and Lying Aversion." *Experimental Economics*, 12(2): 180–92.
- Jacob, Brian A., and Steven D. Levitt.** 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics*, 118(3): 843–77.
- Janisse, Michel P.** 1973. "Pupil Size and Affect: Critical Review of Literature since 1960." *Canadian Psychologist*, 14(4): 311–29.
- Janisse, Michel P., and M. T. Bradley.** 1980. "Deception, Information and the Pupillary Response." *Perceptual and Motor Skills*, 50(3): 748–50.
- Johnson, Eric J., Colin F. Camerer, Sankar Sen, and Talia Rymon.** 2002. "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining." *Journal of Economic Theory*, 104(1): 16–47.
- Kahneman, Daniel, and Jackson Beatty.** 1966. "Pupil Diameter and Load on Memory." *Science*, 154(3756): 1583–85.
- Kang, Min Jeong; Ming Hsu, Ian M. Krajbich, George Loewenstein, Samuel M. McClure, Joseph Tao-yi Wang, and Colin F. Camerer.** 2009. "The Wick in the Candle of Learning: Epistemic Curiosity Activates Reward Circuitry and Enhances Memory." *Psychological Science*, 20(8): 963–73.
- Kartik, Navin.** 2009. "Strategic Communication with Lying Costs." *Review of Economic Studies*, 76(4): 1359–95.
- Kartik, Navin, Marco Ottaviani, and Francesco Squintani.** 2007. "Credulity, Lies, and Costly Talk." *Journal of Economic Theory*, 134(1): 93–116.
- Lin, Hsiou-wei, and Maureen F. McNichols.** 1998. "Underwriting Relationships, Analysts' Earnings Forecasts and Investment Recommendations." *Journal of Accounting and Economics*, 25(1): 101–27.
- Lubow, R. E., and Ofer Fein.** 1996. "Pupillary Size in Response to a Visual Guilty Knowledge Test: New Technique for the Detection of Deception." *Journal of Experimental Psychology: Applied*, 2(2): 164–77.
- Michaely, Roni, and Kent L. Womack.** 1999. "Conflict of Interest and the Credibility of Underwriter Analyst Recommendations." *Review of Financial Studies*, 12(4): 653–86.
- Nietzsche, Friedrich.** 1878. *Human, All Too Human: A Book for Free Spirits*. 2nd ed., trans. R. J. Hollingdale. Cambridge, UK: Cambridge University Press, 1996.
- Núñez, Jennifer Maria, B.J. Casey, Tobias Egner, Todd Hare, and Joy Hirsch.** 2005. "Intentional False Responding Shares Neural Substrates with Response Conflict and Cognitive Control." *NeuroImage*, 25(1): 267–77.
- Oka, Shunichi, C. Richard Chapman, and Robert C. Jacobson.** 2000. "Phasic Pupil Dilation Response to Noxious Stimulation: Effects of Conduction Distance, Sex, and Age." *Journal of Psychophysiology*, 14(2): 97–105.
- Rosovsky, Henry, and Matthew Hartley.** 2002. "Evaluation and the Academy: Are We Doing the Right Thing? Grade Inflation and Letters of Recommendation." Cambridge, MA: American Academy of Arts and Sciences.
- Sanchez-Pages, Santiago, and Marc Vorsatz.** 2007. "An Experimental Study of Truth-Telling in a Sender-Receiver Game." *Games and Economic Behavior*, 61(1): 86–112.
- Selten, Reinhard.** 1991. "Properties of a Measure of Predictive Success." *Mathematical Social Sciences*, 21(2): 153–67.
- Sims, Christopher A.** 2006. "Rational Inattention: Beyond the Linear-Quadratic Case." *American Economic Review*, 96(2): 158–63.