Pinocchio's Pupil: Using Eyetracking and Pupil Dilation

To Understand Truth-telling and Deception in Games

Joseph Tao-yi Wang, Michael Spezio and Colin F. Camerer[*]

Abstract

We conduct laboratory experiments on sender-receiver games with an incentive for biased transmission (such as security analysts painting a rosy picture about earnings prospects). Our results confirm earlier experimental findings of "overcommunication"—messages are more informative of the true state than they should be, in equilibrium theory. Furthermore, we used eyetracking to show that senders look much less at receiver payoffs compared to their own payoffs. At the same time, the senders' pupils dilate when they send deceptive messages, and dilate more when the deception is larger in magnitude. Together, these data are consistent with the hypothesis that figuring out how much to deceive another player is cognitively difficult. Using a combination of sender messages, lookup patterns, and pupil dilation, we can predict the true state about twice as often as predicted by equilibrium. Using these measures would enable receiver subjects to earn 6-8 percent more than they actually do.

In the inferior and middling stations of life… The good old proverb, therefore, that honesty is the best policy, holds, in such situations, almost always perfectly true. …In the superior stations of life the case is unhappily not always the same. In the courts of princes... flattery and falsehood too often prevail over merit and abilities.

~ Adam Smith, Theory of Moral Sentiments, III.5

I. Introduction

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 in Merrill Lynch used a five-point rating system (1=Buy to 5=Sell) to predict how the stock would perform. They usually gave two separate 1-5 ratings for short run (0-12 months) and long run (more than 12 months) performance. 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).[2] In one case, the online direct marketing firm LifeMinders, Inc. (LFMN), Blodget first reported a rating of 2-1 (short run "accumulate"—long run "buy") just before Merrill Lynch got investment banking business from 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 emailed fellow analysts that "LFMN is at $4. I can't believe what a POS [piece of shit] that thing is."[3] He was later banned from the security industry for life[4] and fined $4 million.[5]

---

[1] For a detailed description of the tech-stock bubble and how it happened, see Michael J. Brennan (2004). For evidence regarding analyst recommendations affected by conflicts of interest, see Lin and McNichols (1998) and Roni Michaely and Kent L. Womack (1999). Also note experimental results reported in Hunton and McEwen (1997), which gave real firm data to financial analysts under hypothetical incentive schemes and show that both eyetracked cognitive information search strategies and the incentive structure correlated with forecasting accuracy.

[2] See Complaint in Securities and Exchange Commission v. Henry M. Blodget, 03 CV 2947 (WHP) (S.D.N.Y.) (2003), paragraph 11-12.

[3] See Complaint in Securities and Exchange Commission v. Henry M. Blodget, 03 CV 2947 (WHP) (S.D.N.Y.) (2003), paragraph 70-72.

[4] See Securities and Exchange Commission Order Against Henry M. Blodget (2003).

This case is an example of a biased transmission game. Biased transmission 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 inference of the uninformed players. Given these behavioral patterns, mechanisms can be designed to encourage telling the truth given likely behavior.[6]

Incentives for biased transmission 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.[7] 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.[8] In universities, grade inflation and well-polished recommendation letters help schools promote their graduates.[9] Other examples of incentives for biased transmission include government-expert relationships in policy making, doctor-patient relationships in health care choices, teacher cheating on student tests[10] and the floor-committee relationship in Congress.

This paper reports experiments on a biased transmission game (sometimes called a "cheap talk" or strategic information transmission game; see Vincent P. Crawford and Joel Sobel, 1982). In the game, a sender learns the true state (a number S) and sends a message M to a receiver who

---

[5] See United States District Court Final Judgement on Securities and Exchange Commission v. Henry M. Blodget 03 Civ. 2947 (WHP) (S.D.N.Y.) (2003).

[6] See Theodore Groves (1973), Jerry Green and Jean-Jacques Laffont (1977) and Roger B. Myerson (1979).

[7] See Brennan (2004), pp. 8-9, and Brian J. Hall and Kevin J. Murphy (2003), pp. 60-61.

[8] In fact, 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, while in the prior two months he 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, Civil Action No. H-04-0284 (Harmon) (S.D. Tx.) (2004), paragraph 81-82.

[9] See, for example, Henry Rosovsky and Matthew Hartley (2002).

then chooses an action A. The receiver prefers to choose an action which matches the state, but the sender wants the receiver to choose an action closer to S+b, where b is a bias parameter. The value of b is varied across rounds. When b=0 senders prefer to just announce S (i.e., M=S) and they almost always do. 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), just as Blodget never rated a stock 4 or 5 (the equivalent of 1-2 in our game).

An advantage of the biased transmission game for studying deception is game theory makes precise equilibrium predictions about how much informed agents will exaggerate what they know, when they know that other agents are fully-informed about the game's structure and the incentives to exaggerate. And while in most other deception studies[11], subjects are instructed to lie or give weak or poorly controlled incentives,[12] subjects in experiments like ours choose voluntarily whether to deceive others or not (see also John Dickhaut et al., 1995, Andreas Blume et al., 1998, 2001 and Hongbin Cai and Joseph T. Wang, 2005).[13] Senders and receivers also have clear measurable economic incentives to deceive and to detect deception.[14]

---

[10] For example, Jacob and Levitt (2003) show how public school teachers cheat on student standardized tests in response to high-power incentive systems based on these test scores.

[11] For a survey of studies on (skin-conductance) polygraph, see Theodore R. Bashore and Paul E. Rapp (1993). For lie-detection studies in psychology, see the reviews of Robert E. Kraut (1980) and Aldert Vrij (2000). For a comprehensive discussion of different cues used to detect lies, see Bella M. DePaulo et al. (2003). For individual differences in lie-detection (Secret Service, CIA and sheriffs do better), see Paul Ekman and Maureen O'Sullivan (1991) and Paul Ekman et al. (1999). More recently studies in neuroscience using functional magnetic resonance imaging (fMRI) include Sean A. Spence et al. (2001), D. D. Langleben et al. (2002) and F. Andrew Kozel et al. (2004).

[12] One exception is Samantha Mann et al. (2005) which used footage of real world suspect interrogation to test lie-detecting abilities of ordinary police. However, a lot of experimental control is lost in this setting. One interesting findings in this study is that counter to conventional wisdom, the more subjects relied on stereotypical cues such as gaze aversion to detect lies, the *less* accurate they were.

[13] Most lie-detection studies have three drawbacks: (1) They do not use naturally-occurring lies (because it is then difficult to know whether people are actually lying or not). Instead, most studies create artificial lies by giving subjects true and false statements (or creating a "crime scenario") and instructing them to either lie or tell the truth, sometimes to fool a lie-detecting algorithm or subject. However, instructed deception can be different than naturally-occurring

Besides measuring choices in these games, our experiment uses "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 4 milliseconds.[15] This is 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 the rules.[16]

The eyetracking apparatus also measures how much subjects' pupils "dilate" (expand in width and area due to arousal). Pupils dilate under stress,[17] cognitive difficulty,[18] arousal and pain.[19] Pupillary responses have also been measured in the lie-detection literature for several

---

voluntary deception, and the ability to detect instructed deception might be different than detecting voluntary deception. (2) The incentives to deceive in these studies are typically weak or poorly controlled (e.g., in Spence et al. (2001) all subjects were told that they successfully fooled the investigators who tried to detect them; in Mark G. Frank and Paul Ekman (1997), subjects were threatened a of "sitting on a cold, metal chair inside a cramped, darkened room labeled ominously XXX, where they would have to endure anywhere from 10 to 40 randomly sequenced, 110-decibel startling blasts of white noise over the course of 1 hr" but never actually enforcing it.). (3) Subjects are typically not economically motivated to detect deception. Experiments using the biased-transmission paradigm from game theory address all these drawbacks.

[14] Most lie-detection studies have three drawbacks: (1) They do not use naturally-occurring lies (because it is then difficult to know whether people are actually lying or not). Instead, most studies create artificial lies by giving subjects true and false statements (or creating a "crime scenario") and instructing them to either lie or tell the truth, sometimes to fool a lie-detecting algorithm or subject. However, instructed deception can be different than naturally-occurring voluntary deception, and the ability to detect instructed deception might be different than detecting voluntary deception. (2) The incentives to deceive in these studies are typically weak or poorly controlled (e.g., in Spence et al. (2001) all subjects were told that they successfully fooled the investigators who tried to detect them; in Mark G. Frank and Paul Ekman (1997), subjects were threatened a of "sitting on a cold, metal chair inside a cramped, darkened room labeled ominously XXX, where they would have to endure anywhere from 10 to 40 randomly sequenced, 110-decibel startling blasts of white noise over the course of 1 hr" but never actually enforcing it.). (3) Subjects are typically not economically motivated to detect deception. Experiments using the biased-transmission paradigm from game theory address all these drawbacks.

[15] Previous studies (see footnote 13) used a "Mouselab" system in which moving a cursor into a box opens the box's contents. One small handicap of this 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.

[16] See Camerer et al. (1993); Costa-Gomes et al. (2001); Johnson et al. (2002); Costa-Gomes and Crawford (2005); and the recent Gabaix et al. (2006).

[17] See R. A. Hicks et al. (1967), R. Bull and G. Shead (1979), and Darren C. Aboyoun and James N. Dabbs (1998).

[18] See Jackson Beatty (1982) and B. C. Goldwater (1972).

[19] See C. Richard Chapman et al. (1999) and Shunichi Oka et al. (2000).

decades (that's why poker players often wear sunglasses if they are allowed to).[20] These studies suggest that pupil dilation might be used to infer deceptive behavior because senders find deception stressful or cognitively difficult.

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

1. We replicate the results of Cai and Wang's (2005) original experimental study using a different matching protocol (partner rather than stranger) and a more obtrusive measurement technique. That is, the correlation of messages M and actions A with the actual state S, and players' payoffs, decline with the bias b. At the same time, when b>0 the correlations between states and messages and actions are much larger than predicted by theory, which means senders are being more truthful than they should be, a phenomenon we call "overcommunication".

2. The lookup data suggests informed players are not very strategically sophisticated: They mostly look at their own payoffs, and at payoffs for the true state S. Which payoffs they look at most often is a modest predictor of their later choices, and a better predictor than equilibrium predictions.

3. Senders' pupils dilate more widely when they are misrepresenting by a larger amount. Using pupil dilation as indicators, we can predict the propensity to misrepresent the state, and the degree of misrepresentation, with some accuracy. The true state can also be predicted with some accuracy from a combination of messages, lookup patterns, and pupil dilation.

---

[20] See for example, F. K. Berrien and G. H. Huntington (1942), I. Heilveil (1976), Michel P. Janisse (1973), M. T. Bradley and Michel P. Janisse (1979, 1981), Michel P. Janisse and M. T. Bradley (1980), R. E. Lubow and Ofer Fein (1996), and Daphne P. Dionisio et al. (2001).

Since economists are used to judging theories only by whether they predict choices accurately, it is useful to ask what direct measurement of eye fixations and pupil dilation can add. The inferential strategy from eyetracking in our study is similar to previous studies of offers and lookups in three-period alternating-offer bargaining (Colin F. Camerer et al., 1993; Eric J. Johnson et al., 2002) which adopt a mouse-tracking technology (Mouselab). In those experiments, opening offers typically fell between an equal split of the first-period surplus and the subgame perfect equilibrium prediction (assuming self-interest). These offers could be caused by limited strategic thinking (i.e., players do not always look ahead to the second and third round payoffs of the game), or by computing an equilibrium by looking ahead, adjusting for fairness concerns of other players. Offers alone cannot distinguish between these two theories, but lookups can. The failure to look at payoffs in future periods showed that the deviation of offers from equilibrium was (at least partly) due to limited strategic thinking, rather than entirely due to equilibrium adjustment for fairness.[21]

Miguel Costa-Gomes et al. (2001) use the same mouse-tracking technology in a different way that is also powerful. In the dominance-solvable games they study, two natural decision rules are those in which players optimize against perceived random play (L1) or optimize against perceived random play excluding dominated strategies (D1). But L1 and D1 choices are the same in most games. However, D1 players have to look at the payoffs of other players (and detect dominance relations in others' payoffs) and L1 players don't. When they classify players into L1 or D1 using choices alone, they find a roughly equal mixture of the two rules. But when they use

---

[21] Furthermore, comparing across rounds, when players do look ahead at future round payoffs their resulting offer are closer to the self-interested equilibrium prediction (see Eric J. Johnson and Colin F. Camerer, 2004). Thus, the lookup data can actually be used to predict choices, to some degree.

lookup patterns, they find mostly L1's and few D1's. Thus, if their study had used only choices, and not lookups, they would have reached the wrong conclusion.[22]

In the accounting literature, James E. Hunton and Ruth A. McEwen (1997) asked analysts under hypothetical incentive schemes to make earnings forecast based on real firm data, and investigated factors that affect the accuracy of these forecasts. Using an eye-movement computer technology (Integrated Retinal Imaging System, IRIS), they find that analysts who employ a "directive information search strategy" make more accurate forecasts, both in the lab and in the field, even after controlling for years of experience. This indicates that eyetracking may provide an alternative measure of experience or expertise that is not simply captured by seniority. Had they not observed the eye movements, they could not have measured the difference in information search which are linked to accuracy.

These three studies just illustrate the potential for using cognitive data, besides choices, for distinguishing between competing theories or inspiring new theory. In the biased-transmission games, the overcommunication of the true state that we observe is consistent with two rough accounts, strategizing and guilt, or cognitive difficulty. Senders may feel guilty about deceiving the receivers and potentially costing the receivers money. According to this theory, senders will look at the receiver payoffs (since seeing those payoffs is the basis of guilt) and their pupils will dilate when they misrepresent the state (i.e., choose M different from S) due to emotional arousal from guilt. In this story, the guilt springs from the senders' realization that their actions are costing the receivers money, which requires them to look at the receiver payoffs.[23]

---

[22] One might wonder why we should care whether subjects were following L1 or D1 rules if they gave the same predictions in choices. However, there are other games that the two rules predict very differently, and the point is to see if the same identified rule can explain behavior in more games.

[23] There is little doubt that guilt sometimes exists and affects strategic behavior. For example, Uri Gneezy (2005) find that changing the costs to others affects deception by subjects. Eyetracking helps us explore this insight further using data on whether potential deceivers actually know those costs.

A different story is that senders do not feel guilty, but find it cognitively difficult to figure out how much to misrepresent the state.   For example, senders might believe that some other senders always tell the truth, and receivers might therefore believe messages are truthful. Then strategic senders have to think hard about how much to misrepresent the state to take advantage of the receivers' naïveté (as in Vincent P. Crawford, 2003 and Macro Ottaviani and Francesco Squintani, 2004). In this story, senders do not have to pay much attention to receiver payoffs but their pupils will dilate because of the cognitive difficulty of misrepresentation.

Taken together, the choices, eye fixations, and pupil dilation can roughly adjudicate between these two stories.   Senders spend little time looking at the receiver payoffs but their pupils are dilated when they misrepresent the state (when b>0).   Moreover, the time spent looking at the receiver payoffs increases as bias increases, but remains around half the lookup time of the sender payoffs.   These facts point in the direction of the cognitive difficulty story rather than strategizing plus guilt.[24]

This is the first study in experimental economics to use a combination of eyetracking and pupil dilation, and is, of course, hardly conclusive.   But the pupil dilation results by themselves suggest that the implicit assumption in theories of "cheap talk" in games with communication— namely, that deception has no cost— is not completely right.   Mark Twain famously quipped, "If you tell the truth, you don't have to remember anything."[25]   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 the form of looking patterns and pupil dilation.   Future theories could build in an

---

[24] In fact, when the senders were asked after the experiment whether they considered sending a number different from the true state deception, 9 of the subjects said yes, while another 3 said no, but gave excuses such as "it's part of the game" or "the other player knows my preference difference."   Only 1 subject said no without any explanation.   These debriefing results also suggest that guilt has played little role in the experiment.

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.

II. The Biased-Transmission Game

In each round of the experiments, subjects play a game of strategic information transmission, involving "cheap talk." 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, as described in more detail in 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 equal probability. The receiver knows the bias b, but not the realization of the state S. Both players commonly know the basic structure of the game.

The sender then sends a message to the receiver, from the set of messages $\mathbf{M} = \{1, 2, 3, 4, 5\}$.[26] 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 $u_R = 110 - 10 \cdot |S - A|^{1.4}$, and $u_S = 110 - 10 \cdot |S + b - A|^{1.4}$, 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 the true state plus the bias b. Figure 1 shows the screen display for b=1 and S=4.

---

[25] Quotation taken from Mark Twain's Notebook, 1894.
[26] Following Cai and Wang (2005), we use the specific 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).

As in Cai and Wang (2005), the most informative equilibrium for b=2 is "babbling", in which the sender sends an uninformative message, while the receiver ignores the message and chooses A=3 based on her prior beliefs. When b=1, the most informative equilibrium requires the senders to send messages {1,2} (i.e., randomize between saying 1 or 2) when the S is 1 or 2, and send {3,4,5} when S is 3-5. When b=1 the receivers should choose action A=1 or 2 when seeing M={1,2}, and A=4 when seeing M={3,4,5}. When b=0, truth-telling by choosing M=S (and receivers choosing A=M) is the most informative equilibrium

To be sure subjects learn, and to collect a lot of trials to pool across, the same game is played 45 times among the two paired players with random choices of bias b (and random states) in each round.[27] Because we could only eyetrack one subject at a time, we used a partner protocol in which a pair of subjects played repeatedly in a fixed-role protocol. Only the senders were hooked up to the mobile Eyelink eyetracker. The results reported below focus entirely on the eye fixations and pupil dilation of senders, and the message choices of senders and action choices of receivers.

We record subject choices and focus on the most informative equilibrium in the one-shot game.[28] Informativeness is measured by the correlation between actions and the true states, and by sender and receiver payoffs (more informative equilibria have higher expected payoffs). In addition, if we assume a natural language interpretation of the message, we can also measure the "informativeness" of senders' messages by the correlation between the true states of the world and

---

[27] Cai and Wang (2005) used a random matching scheme that guarantee two subjects never see each other again. However, due to the eyetracking nature of this experiment, we have subjects paired against the same opponent throughout the 45 rounds they play.

[28] We do not consider possible dynamic equilibrium that might sustain higher information transmission level. Nonetheless, this is not a problem for bias = 0 or 2. When bias = 2, babbling is the only equilibrium in the one shot game and backward induction yields the babbling equilibrium for all finitely repeated games; when bias = 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 bias = 2.

the messages they sends. How "trusting" the receivers are can be measured by the correlation between the messages they receive and the actions they take.[29]

Subjects were 24 Caltech undergraduates (12 pairs) recruited from the Social Science Experimental Laboratory subject pool. They earned between $12 and $24 in addition to a $5 show-up fee. To compare across pairs, we use the same set of randomly drawn biases and states for 9 of the 12 pairs, and use two other sets of parameters for the remaining 3 pairs to see if there were any effects for using the same parameters.[30]

Note that 24 subjects might appear to be a small sample size.[31] But most experimental studies with larger samples have fewer choices per subject. Our eyetracked subjects play 45 games, and make a very large number of eye fixations; so we have a lot of data for each subject and can often draw confident statistical conclusions from these sample sizes.[32]

III. Results

III.A Comparative Statics and Behavior

What are the comparative static results? Looking at subjects' choices (M and A), we find that the key comparative static prediction of Crawford and Sobel (1982) holds in the data. In other words, as the bias b increases, the information transmitted decreases, measured either by the

---

[29] 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 (2005). Moreover, many behavioral theories of lying, such as Crawford (2003) and Ottaviani and Squintani (2004), also lead to this sort of natural language interpretation since naïve receivers would take the message at face value.

[30] We did not see any effect in both the main results and the split-sample test.

[31] Ironically, a subject size of 24 is perceived as a *large* sample size for psychophysical studies.

[32] As we note below, for a primary analysis predicting pupil dilation from observables, a split-sample test comparing two groups of six subjects yields comparable results in the two sub-samples.

correlation between State S and Action A, or by receiver payoffs, confirming the findings of Cai and Wang (2005).

Table 1 shows that the actual information transmitted, measured by the correlation between states S, actions A, and messages M, decreases as b increases from 0 to 1 to 2. 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 0.5.

Since poor information transmission harms both players' payoffs, the decline in their payoffs as the bias increases is an economic measure of how deception affects payoffs. Across the values of b, the average receiver payoffs decreased from a near perfect 109.14 to 94.01 to 85.52; the sender payoffs decreased from 109.14 to 93.35, and then to 41.52.

Importantly, when the bias is positive, information transmission is higher (measured by correlations among S, M and A) and payoffs are higher than predicted by standard economic theory. These data replicate the "overcommunication" (too much truth-telling) reported in Cai and Wang (2005).[33]

What does the raw data look like? Figure 2-4 display the three dimensions of the raw choice data-- states, messages and actions--for b = 0, 1 and 2, respectively. Each Figure is a 5-by-5 display where the true states 1-5 correspond to the five rows and the sender messages 1-5 correspond to the five columns. Within each cell, the average receiver action is given in numbers, with a pie chart that gives the breakdown of actions chosen by the receiver. Actions in the pie-chart are represented by a gray-scale, ranging from white (action 1) to black (action 5), indicating receivers' response to the senders' messages. The area of the pie-chart in each cell is scaled by the number of occurrences for the corresponding state and message. 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.

For example, when b=0, and there is no conflict of interest, large pie-charts are concentrated on the diagonal, 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 true state increases, meaning that the receivers follow senders' recommendation when choosing their actions. Thirdly, the distribution of state frequencies conditional on each message (i.e., 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 standard theory.

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 large pie charts off the diagonal. Consistent with the findings in Cai and Wang (2005), this departure is lopsided—only the upper diagonal of Figure 3 is populated with large pie charts.[34] That is, for a given state, the most common messages are the state itself or higher messages. Furthermore, the largest pie charts of each row are mainly on the line to the right of the diagonal (i.e., the state S+1), consistent with the L1 sender behavior discussed in Cai and Wang (2005). Within the upper diagonal, the pie-chart gets darker and darker going down and right, showing how the receivers "correctly" respond to the messages and increase their actions as the state and message increase. Since the conditional distribution of states (columns in Figure 3) shift from a mass point on the true state (as in Figure 2) to a distribution skewed toward state 3 to 5, some information is transmitted. However, this

distribution is not consistent with the {1, 2}-{3, 4, 5} partition equilibrium predicted by standard theory which requires states within each partition to have the same conditional distribution of states.[35]

Finally, when b=2, standard theory predicts a babbling equilibrium. If they were playing this equilibrium, the pie-charts in each cell would be roughly the same size (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 4 do not all show the same uniform distribution of state frequencies.[36] However, many senders still sent message 5, especially for states 2 to 5. And a substantial amount of receivers did chose action 3, as predicted in the babbling equilibrium. Therefore, Figure 4 seems to be a mixture of truth-telling and babbling.

III.B Lookup Patterns

What numbers do senders look at? Table 3 shows the number of separate fixations (the fixation threshold is 50msec) and average lookup time for various parameters of the game. Senders clearly are thinking carefully about the game because they look up the state and the bias parameter 2-6 times per round, for about 1 second. (The low time per lookup is a reminder that the eye glances around very rapidly, making frequent quick fixations, as is typical of other tasks including reading.) Senders also look at their own payoffs about twice as often, and for twice as long, compared to receiver payoffs. Interestingly enough, the ratio of lookup time for sender and receiver

---

[34] 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.

[35] If subjects were playing according to the partition equilibrium, column 1 and 2 should both have equal (1/2) probability on state 1 and 2, and zero probability elsewhere, indicating the state being in partition {1,2}, while column 3 to 5 should all have equal probability (1/3) on states 3 through 5, and zero elsewhere (indicating the state being in partition {3,4,5}).

[36] For instance, if the message is 1, the true state will never be above 3.

payoffs are always close to 2 as the bias increases (more guilt involved), and is not consistent with the strategizing and guilt story of deception. Moreover, for b=2, which has the most guilt involved, if we use the median of the lookup time for receiver payoffs to split the sample into two groups, the high (receiver payoffs) lookup time group actually has more deception than the low group, also inconsistent with a guilt story predicting that the more one cares about other's payoffs, the less one should "cheat" them.[37]

Do lookup times shrink over time? A natural concern with lookup data is whether subjects are simply memorizing the payoff tables, which would undermine the value of eyetracking. While memorization is unlikely because the states S and bias b vary randomly across the 45 rounds, and looking at the screen is so effortless (memorizing is harder than looking), it is useful to see whether response times drop sharply across rounds. Considering the average response times across three blocks of 15 rounds, response times do drop substantially when the bias is b=0 (from around 5 seconds to 2 seconds) but drop much less in the b=1 and b=2 rounds (only about 20 percent from the first 15 rounds to the last 15 rounds). This indicates substantial learning for b=0, but not much for b=1 and 2, and is also consistent with a cognitive difficult story of deception.

What state payoffs do senders look at? Table 4 shows that subjects have about five times more fixations (5.9 lookups per round) and lookup time (1.71 seconds per round) on the payoff rows corresponding to the true state than on 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 1 or 2.[38]

---

[37] For the high group, the correlation between states and messages is 0.545, and the average LIE_SIZE is 0.875; for the low group, the correlation is 0.688, and the average LIE_SIZE is 0.705.

[38] Note that the Table 5 indicates significant 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 could predict

An efficient way to convey information about fixations and lookup time visually is with an icon graph (developed by Eric Johnson, cf. Johnson et al., 2002), as in Figure 5. Each box in Figure 5 represents the attention paid to the payoff corresponding to each state-action combination. Figure 5a represents attention to the sender payoff boxes and Figure 5b represents attention to the receiver payoff boxes. The width of the box is a linear function of the average number of fixations on that box, and the height is a linear function of the average total looking time in that box. Boxes which are wide and long were looked at repeatedly and for a longer time. The bars in the first columns represent the sum of looking time across each row. Longer bars represent longer time (for that state).

Figure 5 shows the icon graph for the rows corresponding to the true states when the bias is 1.[39] The first thing to notice is that subjects spend much more time looking at their own payoffs (Figure 5a) than the payoffs of receivers (Figure 5b), as the Table 4 statistics show. Subjects' lookups are also more frequent and longer for actions that are equal to the actual state or the state plus one. The looking patterns can be supportive of a quantal response equilibrium,[40] which predicts actions concentrated around A=4 (when actual states are 3, 4, or 5) and mixing A=1 and 2 (when states are 1 or 2). However, state 2 has the longest lookups on A=3, which is at odds with the QRE prediction of choosing mainly A=1 and 2. Furthermore, state 5 also does not fit the pattern: Senders look frequently at payoffs corresponding to actions 1, 3 and 4, and also look relatively often at the receiver payoff boxes in this state. (Both informative equilibrium and QRE predict that

---

what the actual state was rather reliably. Of course, it is not clear how the senders would behave had they known that their lookup patterns were monitored by the receivers.

[39] 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 about a quarter as often as they look at their own payoffs.

[40] Quantal response equilibrium in extensive form games was introduced by Richard D. McKelvey and Thomas R. Palfrey (1998), and Cai and Wang (2005) applied it to biased transmission games to explain their experimental findings.

state 5 should be treated similarly to states 3-4.) The looking patterns may reflect the fact that this state 5 is the only state in which both subjects prefer the same action choice (action 5).[41]

Figure 6 shows the lookup icon graphs for bias b=2. 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. Since these are states they typically choose, it appears that they are using a one-step rule in which they anticipate that receivers will naively respond to their own message choices M by choosing action A=M (see Crawford, 2003). However, when the state is 4 or 5 this pattern crumbles and they spread attention across more actions. When state = 5 and nothing is better than telling the truth, there is generally less lookup activity.

After examining subjects' lookup patterns, we turn to their pupil dilation responses and see whether we can improve upon prediction of subject behavior.

III.C Pupil Dilation

To correlate pupil dilation with senders' messages, we calculate average pupil sizes for various time periods before and after the sender's message decision, and see if we can predict 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 is calibrated to record 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.[42]

---

[41] Interestingly enough, due to discreteness, this is the only state that has perfectly aligned preferences.
[42] Allowing eye fixations to determine 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., galvanic skin conductance on the palms, heart rate) since even small hand movements add noise to those measurements.

Since there is a time lag of at least 0.8 second between the instant subjects "made up their minds" and the recording of this decision,[43] we define the *decision time* 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 95 percent of the time until the software records a decision.[44]

Average pupil sizes are regressed on the amount of deception for different biases, or the sizes of the lie (LIE_SIZE = |M-S|), as well as bias and state dummies, controlling for individual fixed effects and individual learning trends (picked up by round number and squared round number variables interacted with individual fixed effects. The specification is:

(1) $$\text{PUPIL}_i = \alpha + \sum_{b=0}^{2} \beta_{1b} \cdot LIE\_SIZE \cdot BIAS_b$$

$$+ \sum_{b \neq 2} \beta_{2b} \cdot BIAS_b + \sum_{s \neq 3} \beta_{3s} \cdot STATE_s + \sum_{k \neq 6} \alpha_k \cdot SUBJ_k$$
$$+ \sum_{k=1}^{J} \left(\gamma_{k,1} ROUND \cdot SUBJ_k + \gamma_{k,2} ROUND^2 \cdot SUBJ_k \right) + \varepsilon$$

where

PUPIL_i = Average pupil (area) size[45] 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.[46] Here,

we normalize each individual's average pupil size to 100.[47]

---

[43]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.

[44] Running similar regressions (with raw pupil size) shows that using a criterion of 98 percent or 99 percent would yield stronger results than that of 95 percent. Moreover, even a noisy 90 percent would still produce the same qualitative results, though some results are less significant. Last but not the least, simply using the exact time the software records the decision (after the 0.8 second time lag) would also give us stronger results.

[45] Note that we are aggregating 100 observations into 1 data point when averaging for each 400 milliseconds interval.

[46] Rounds with very short response time are discarded if the corresponding PUPIL_i cannot be calculated.

[47] Pupil sizes are measured by area, in relative terms. Absolute pixel counts have little meaning since it varies 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 performing this normalization.) Therefore, 100 means 100 percent of an individual subject's typical pupil size.

LIE_SIZE = The "size" of the lie or the amount of deception, measured by the absolute distance

between states and messages, (|M-S|).

$BIAS_b$ = Dummy variable for the bias between the sender and the receiver.

$STATE_s$ = Dummy variable for the true state.

$SUBJ_k$ = Dummy variable for subject k.

ROUND = Round number

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 other effects of different biases b (relative to b=2) and states (relative to S=3), while coefficients α_k capture individual differences (relative to subject 6), 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 5, interacted with bias (denoted β_{1b} where b is the bias parameter). Right before the decision is made (-0.4 seconds to 0 seconds, where 0 seconds is the decision time), the coefficient on the amount of deception is 2.69 percent higher when b=1 and 3.13 percent when b=2. These effects are significant in all 400 millisecond intervals from 800 milliseconds before the decision, to 800 milliseconds after the decision. Sending more deceptive messages is therefore correlated with pupil dilation when b=1 or b=2.

Note that the bias condition by itself does not generate pupil dilation (the coefficients β_{2s} are insignificant). That is, it is not bias, per se, which creates arousal or cognitive difficulty; it is sending more deceptive messages in the bias conditions. Furthermore, the β_{3s} coefficients show that state S=1 tends to undilate pupils (relative to the benchmark state S=3), but the states 4 and 5 dilate pupils. These coefficients show that some states create more cognitive difficulty, but the important

19

interaction between the amount of deception and bias (coefficients β_{1b}) occurs even after controlling for the bias b and the state S. Furthermore, these basic patterns are reproduced when the sample is divided in half, which provides some assurance of statistical reliability even though the absolute sample size is modest.[48]

As noted, the goal of measuring eyetracking and pupil dilation is to see whether these behavioral measures enable us to improve upon predictions of theory. Below we use lookup data and pupil dilation to try to predict states.

IV. Lie-detection and Prediction

IV.A Predicting Deception

We have shown in Table 5 the result of predicting pupil dilation from states, bias, and the interaction between bias and deception. From a practical point of view, it is useful to know what happens when we run the regression in reverse—can the amount of deception be predicted using pupil dilation?

Table 6 summarizes some results using logit and ordered logit estimation to predict whether subjects deceive or not, and the amount of deception, using bias and state dummies, and the difference in pupil size between the beginning and the end of the decision process (also controlling for individual fixed effects and individual learning effects). Since b=0 leads to truth-telling almost

---

[48] Because we measured eyetracking and pupil dilation from 12 senders (a small sample compared to other studies in which measurements per subject are less frequent or much easier), it is useful to check how reliable these results are in two subsamples of six subjects each. The 400-msec interval from +0.4 to +0.8 secs after decision time gives the highest R^2's so we compare those. The β_{3s} coefficients on states give similar patterns—states S=1-2 are negative and S=4-5 are positive (-9.11**, -1.90, 3.92*, 5.18* for first six; -8.73***, -4.25*, 4.10*, 6.84*** for second six, where asterisk (*) notation follows Table 6.) The β_{1b} coefficients across bias levels (b=0, 1, 2) are the most important. They are 4.79, 3.44**, 2.42* for the first six subjects and 12.94*, 4.63***, and 4.47*** for the second six subjects. For other intervals,

all the time, there is little variation in the dependent variables so we exclude b=0 and focus solely on the periods in which b=1 or b=2. Specifically, for the binary dependent variable LIE (=0 if M=S, =1 otherwise), we estimate the logit regression:

$$(2) \quad \log[\Pr(Y=1)] = \theta + \beta_1 \cdot BIAS_1 + \sum_{s \neq 3} \beta_{2s} \cdot STATE_s + \sum_{d=1,2} \beta_{3d} \cdot \left(\frac{PUPIL_{end} - PUPIL_{ini}}{10} \right) \cdot BIAS_d + Controls + \varepsilon$$

With the discrete dependent variable LIE_SIZE= |M–S|, ranging from 0 through $J = 4$, we run the ordered logit regression, for $j = 0, 1, 2, 3, 4 (= J)$,

$$(3) \quad \log[\Pr(Y \geq j)] = \theta_j + \beta_1 \cdot BIAS_1 + \sum_{s \neq 3} \beta_{2s} \cdot STATE_s + \sum_{d=1,2} \beta_{3d} \cdot \left(\frac{PUPIL_{end} - PUPIL_{ini}}{10} \right) \cdot BIAS_d + Controls + \varepsilon$$

where the *Controls* include individual fixed effects and individual learning effects as in the pupil dilation regressions of the previous section, and

$PUPIL_{ini}$ = Average pupil size (area) size in the *first* 0.4 seconds of the decision process starting as soon as the decision-making screen is displayed.

$PUPIL_{end}$ = Average pupil size (area) size in the *last* 0.4 seconds before the decision time

The parameter β_1 represents the effect of having bias 1 instead of 2, β_{2s} represents the effects of the true state (other than 3), and β_{3d} represents the effect of the difference between the final and initial pupil size, and θ_j (or θ) are the constants. The "all" regressions use all observations. The "part" regressions perform a (2/3, 1/3) coin toss for each observation to determine whether to use it in estimation or not. Hence, the "part" regression typically uses two-thirds of the data to estimate the regression, and then use those estimated coefficients to predict whether there is a LIE or not and the LIE_SIZE of the lie (both predictions are rounded to the nearest integer) for the holdout data

---

as predictive power (R^2) falls the reliability across the two subsamples falls but the coefficient signs are almost always the same in the two subsamples and magnitudes are typically reasonably close.

(those not used in estimation).[49] This partial estimation-prediction procedure is repeated 100 times. The coefficients and standard errors reported below are the mean and standard deviation of estimates across these 100 repetitions (a bootstrap procedure).

As shown in Table 6, β_1 is negative and significant for the size of the lie (LIE_SIZE), but not for whether one lies or not (LIE), indicating that different bias conditions influence the amount of deception (LIE_SIZE) but not whether deception occurs at all (LIE). The effect of the true state is only robustly significant for state S=5, when people deceive less often and to a smaller extent (indicating a "ceiling" effect at the top of the state space). The change in pupil dilation times b=2 interaction (β_{32}) is positive and significant (though the β_{31} coefficients are not). Thus, pupil dilation can be used to predict whether a person is being deceptive, and how deceptive they are, when the bias condition encourages deception most (b=2). The bottom rows of the "(part)" columns of the table show that when part of the sample is used to forecast the rest of the data, the forecasts are wrong about 20 percent of the time for lying or not (LIE) and 34 percent of the time for the size of the deception (LIE_SIZE), but the errors of deception size tend to be small (around 80 percent are only off by one). Keep in mind that if senders were playing the most informative equilibria, deceptions could only be accurately predicted 40 percent and 20 percent of the time for b = 1 and 2, respectively. So the hit rates of 80 percent and 66 percent (100 percent minus the error rates reported in the table) are a substantial improvement over the simplest prediction of equilibrium game theory.

---

[49] Here we use the estimated probabilities to calculate the expected outcome (lie/not lie or size of the lie) and round to the nearest integer. Using the most likely outcome (choose the highest probability) yield almost identical results.

IV.B Predicting the True State from Lookups and Pupil Dilation

Although the previous section shows how we may predict deception to a modest degree, one might still wonder how much the receivers can gain by seeing senders' pupil dilation or even the lookup patterns. Hence, now we ask how well a combination of lookup patterns and pupil dilation can predict the true state. For the dependent variable STATE j, ranging from 1 to 5, model 1 is an ordered logit regression

$$\log[\Pr(Y \geq j)] = \theta_j + \beta_1 \cdot BIAS_1 + \sum_{b=1,2}(\beta_{2b} \cdot MESSAGE + \beta_{3b} \cdot ROW_{self} + \beta_{4b} \cdot ROW_{other}) \cdot BIAS_b + \varepsilon$$

where lookups are consolidated into two integer variables:

ROW_{self} = The state of the own-payoff row which has the longest total lookup time of all own-payoff state rows

ROW_{other} = The state of the opponent-payoff row which has the longest total lookup time of all opponent-payoff state rows

The coefficients β_1 represent the effect of having bias b=1 instead of b=2, β_{2b} represents the information about the state contained in the message, β_{3b} represents the effects of the "most viewed row" of one's own payoffs, and β_{4b} represents the effects of the "most viewed row" of the opponent's payoffs. The θ_j are state-specific constants.

In alternative specifications (models 2), we include pupil size effects (β_5 and β_6), as in the previous section. To do this we estimate the ordered logit regression

$$(4) \quad \log[\Pr(Y \geq j)] = \theta_j + \beta_1 \cdot BIAS_1 + \sum_{b=1,2}(\beta_{2b} \cdot MESSAGE + \beta_{3b} \cdot ROW_{self} + \beta_{4b} \cdot ROW_{other}) \cdot BIAS_b$$

$$+ \beta_5 \cdot \left(\frac{PUPIL_{end} - PUPIL_{ini}}{10}\right) + \beta_6 \cdot \left(\frac{PUPIL_{end} - PUPIL_{ini}}{10}\right) \cdot MESSAGE + \varepsilon$$

As in the partial-sample results in Table 6, we estimate these three models using roughly 2/3 of the data, then forecast the actual state using the estimated coefficients for the remaining 1/3 of the data. This procedure is performed 100 times; average βs and (bootstrap) standard errors across the 100 resamplings are reported in Table 7.

Table 7 shows that β_1 is negative (and almost significant) for both specifications, indicating possible differences between small and large biases. The significance of β_{2b} indicates that the messages are informative about the states.[50] The smaller the message, the smaller is the true state, even though standard game theory predicts that little information should be transmitted in the message (none should be transmitted, when b=2).

The lookup data are significantly correlated with states as well. The coefficients β_{3b}, on the most-viewed *own* row variables, are positive and significant in both models. The coefficients β_{4b}, on the most-viewed *other* row variables, are positive but are smaller than own-row coefficients, and are only significant in model 1 when b=2. The important point here is that lookup data improve predictability *even when controlling for the message.* In fact, if the message is 4, but the lookup data indicate the subject was looking most often at the payoffs in row corresponding to state 2, then the model might predict that the true state is 2, not 4.

Pupil size effects are evident too. The coefficient β_5 on the change in pupil size (i.e., pupil dilation) is significantly positive, which means higher states generally produce more dilation (as shown in Table 5). Interacting pupil dilation with the message sent has a negative estimated β_6. That means that when the pupil is more dilated, the weight placed on the message falls—i.e., messages are less informative about the true state because a deception is more likely.

---

[50] We also tried yet another model, which included message dummies instead of MESSAGE, but the results are almost the same.

To predict the true state, we again use the estimated logit probabilities to calculate the expected state, and round it to the nearest integer. The error rates in predicting states in the holdout sample are still substantial. Table 7 shows that the state is predicted incorrectly about 30 percent and 60 percent of the time when b=1 and b=2, respectively. This is better than the actual performance of the receiver subjects, however: They "missed" (A≠S) 56.2 percent of the time for b=1, and 70.9 percent for b=2. (Nevertheless, keep in mind that the error rates in equilibrium would be 60 percent and 80 percent which are even worse).

Around 75 percent of these erroneous predictions from the logit model only miss the state by one, for both specifications and both bias levels b. This is comparable with the actual performance of the receiver subjects, whose "misses" were only off by one unit 80 percent and 67 percent of the time, when b=1 and b=2. This means that including lookup data and pupil dilation *can* improve accuracy even for incorrect prediction, especially for b=2. Also note that the "misses" of the subjects and the logit model in Table 7 are both inconsistent with equilibrium, which predicts that 100 percent of the "misses" will be off by exactly one unit when b=1, and 50 percent of the misses will be one unit off when b=2.

An interesting calculation is how much these predictions might add to the receiver payoffs (cf. "economic value" in Colin F. Camerer et al., 2004). For biases b=1 and b=2, the average actual payoffs for receivers were 93.4 and 86.2 (which are higher than predicted by theory, 87.4 and 71.6, because of the empirical bias toward truth-telling which helps receivers). If receivers had based their predictions on the models estimated in Table 7, and chose the same action as the predicted state (for the holdout sample), their expected payoffs would be around 100 for b=1 and 92

for b=2, which is a modest economic value of 6-8 percent.[51]   These average payoffs are all significantly higher than both what subjects actually earn in the experiments and that predicted by equilibrium theory.

V. Conclusion

This paper reports experiments on sender-receiver games with an incentive for biased transmission (such as managers or security analysts painting a rosy picture about a corporation'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. But receivers know the payoff structure, so they should be suspicious of inflated messages M.

Equilibrium analysis predicts that when b=1 there will be some partial pooling of the truth; when b=2 the only equilibrium is "babbling" in which messages reveal nothing about the true states. Our experimental results confirm earlier experimental findings of "overcommunication"—messages are more informative of the state than they should be, in equilibrium theory. To explore the cognitive foundations of overcommunication, we used eyetracking to record what payoffs the sender subjects are looking at, and how widely their pupils dilate (expand) when they send messages. The biased transmission 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).

---

[51] Of course, this calculation assumes the receivers could measures lookups and pupil dilation without senders altering their lookup patterns because they knew they were being watched and studied. Whether such techniques actually add value is beyond the scope of this paper.

The lookup data show that senders do not look at receiver payoffs as frequently or as long as that of their own payoffs, so they do not appear to be thinking very strategically. Nor does it seem that guilt plays an important role. At the same time, the senders' pupils dilate when they send deceptive messages ($M \neq S$), and dilate more when the deception $|M-S|$ is larger in magnitude. Together, these data are consistent with the hypothesis that figuring out how much to deceive another player is cognitively difficult. The cognitive measures are reliable enough that deception is correlated with pupil dilation, and reversing the regression enables us to predict whether a subject is deceiving mildly well from the dilation response (when the bias parameter $b=2$ is largest). Furthermore, using a combination of sender messages, lookup patterns, and pupil dilation, one can predict the true state about twice as often as predicted by equilibrium, and increase receiver payoffs by 6-8 percent compared to what subjects actually earn in the experiment.

There are many directions for future research. We see our unique contribution as bringing a combination of eyetracking (used in just three types of games so far) and pupil dilation, a measure of cognitive difficulty, to bear on the kind of simple game that lies in the heart of many economic and social questions. Economists often talk loosely about the costs of decision making or difficulty of tradeoffs; pupil dilation gives us one way to start measuring these costs. 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 pupil dilation).

In the realm of deception, two obvious questions for future research are whether there are substantial individual differences in the capacity or willingness to deceive others for a benefit, and whether experience can teach people to be better at deception, and at detecting deception. Both are important for extrapolating these results to domains in which there is self-selection and possibly

large effects of experience (e.g., politics). In other domains of economic interest, the combination of eyetracking and pupil dilation could be used to study any situation in which the search for information and cognitive difficulty are both useful to measure, such as "directed cognition" (Xavier Gabaix et al., 2006), perceptions of advertising and resulting choices, and attention to trading screens with multiple markets (e.g., with possible arbitrage relationships).

References

Complaint in Securities and Exchange Commission V. Henry M. Blodget, 03 Cv 2947 (Whp) (S.D.N.Y.), Securities and Exchange Commission Litigation Release No. 18115, April 23, 2003. Washington, DC: Securities and Exchange Commission, 2003.

Order against Henry M. Blodget, Securities and Exchange Commission Administrative Proceedings, File No.3-11322, October 31, 2003. Washington, DC: Securities and Exchange Commission, 2003.

Second Amended Complaint in Securities and Exchange Commission V. Richard A. Causey, Jeffrey K. Skilling and Kenneth L. Lay, Civil Action No. H-04-0284 (Harmon) (S.D. Tx.), Securities and Exchange Commission Litigation Release No. 18776, July 8, 2004. Washington, DC: Securities and Exchange Commission, 2004.

United States District Court Final Judgement on Securities and Exchange Commission V. Henry M. Blodget 03 Civ. 2947 (Whp) (S.D.N.Y.), Securities and Exchange Commission Litigation Release No. 18115, Washington, DC: Securities and Exchange Commission, 2003.

Aboyoun, Darren C. and Dabbs, James N. "The Hess Pupil Dilation Findings: Sex or Novelty?" *Social Behavior and Personality*, 1998, *26*(4), pp. 415-19.

Bashore, Theodore R. and Rapp, Paul E. "Are There Alternatives to Traditional Polygraph Procedures." *Psychological Bulletin*, 1993, *113*(1), pp. 3-22.

Beatty, Jackson. "Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources." *Psychological Bulletin*, 1982, *91*(2), pp. 276-92.

Berrien, F. K. and Huntington, G. H. "An Exploratory Study of Pupillary Responses During Deception." *Journal of Experimental Psychology*, 1943, *32*(5), pp. 443-49.

Blume, Andreas; DeJong, Douglas V.; Kim, Yong-Gwan and Sprinkle, Geoffrey B. "Evolution

of Communication with Partial Common Interest." *Games and Economic Behavior*, 2001, *37*(1), pp. 79-120.

____. "Experimental Evidence on the Evolution of Meaning of Messages in Sender-Receiver Games." *American Economic Review*, 1998, *88*(5), pp. 1323-40.

Bradley, M. T. and Janisse, Michel P. "Accuracy Demonstrations, Threat, and the Detection of Deception - Cardiovascular, Electrodermal, and Pupillary Measures." *Psychophysiology*, 1981, *18*(3), pp. 307-15.

____. "Pupil Size and Lie Detection - the Effect of Certainty on Detection." *Psychology*, 1979, *16*(4), pp. 33-39.

Brainard, David H. "The Psychophysics Toolbox." *Spatial Vision*, 1997, *10*, pp. 433-36.

Brennan, Michael J. How Did It Happen?, Unpublished paper, 2004.

Bull, R. and Shead, G. "Pupil-Dilation, Sex of Stimulus, and Age and Sex of Observer." *Perceptual and Motor Skills*, 1979, *49*(1), pp. 27-30.

Cai, Hongbin and Wang, Joseph T. "Overcommunication in Strategic Information Transmission Games." *Games and Economic Behavior*, 2005, *forthcoming*.

Camerer, Colin F.; Ho, Teck-Hua and Chong, Juin-Kuan. "A Cognitive Hierarchy Model of Games." *Quarterly Journal of Economics*, 2004, *119*(3), pp. 861-98.

Camerer, Colin F.; Johnson, Eric J.; Rymon, Talia and Sen, Sankar. "Cognition and Framing in Sequential Bargaining for Gains and Losses," K. G. Binmore, A. P. Kirman and P. Tani, *Frontiers of Game Theory*. Cambridge: MIT Press, 1993, 27-47.

Chapman, C. Richard; Oka, Shunichi; Bradshaw, David H.; Jacobson, Robert C. and Donaldson, Gary W. "Phasic Pupil Dilation Response to Noxious Stimulation in Normal Volunteers: Relationship to Brain Evoked Potentials and Pain Report." *Psychophysiology*,

1999, *36*(1), pp. 44-52.

Cornelissen, Frans W.; Peters, Enno M. and Palmer, John. " The Eyelink Toolbox: Eye Tracking with Matlab and the Psychophysics Toolbox." *Behavior Research Methods, Instruments & Computers*, 2002, *34*, pp. 613-17.

Costa-Gomes, Miguel; Crawford, Vincent P. and Broseta, Bruno. "Cognition and Behavior in Normal-Form Games: An Experimental Study." *Econometrica*, 2001, *69*(5), pp. 1193-235.

Crawford, Vincent P. "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions." *American Economic Review*, 2003, *93*(1), pp. 133-49.

Crawford, Vincent P. and Sobel, Joel. "Strategic Information Transmission." *Econometrica*, 1982, *50*(6), pp. 1431-51.

DePaulo, Bella M.; Lindsay, James J.; Malone, Brian E.; Muhlenbruck, Laura; Charlton, Kelly and Cooper, Harris. "Cues to Deception." *Psychological Bulletin*, 2003, *129*(1), pp. 74-118.

Dickhaut, John; McCabe, Kevin and Mukherji, Arijit. "An Experimental Study of Strategic Information Transmission." *Economic Theory*, 1995, *6*, pp. 389-403.

Dionisio, Daphne P.; Granholm, Eric; Hillix, William A. and Perrine, William F. "Differentiation of Deception Using Pupillary Responses as an Index of Cognitive Processing." *Psychophysiology*, 2001, *38*(2), pp. 205-11.

Ekman, Paul and O'Sullivan, Maureen. "Who Can Catch a Liar?" *American Psychologist*, 1991, *46*, pp. 913-20.

Ekman, Paul; O'Sullivan, Maureen and Frank, Mark G. "A Few Can Catch a Liar." *Psychological Science*, 1999, *10*, pp. 263-66.

Frank, Mark G. and Ekman, Paul. "The Ability to Detect Deceit Generalizes Acrosss Different

Types of High-Stake Lies." *Journal of Personality and Social Psychology*, 1997, *72*(6), pp. 1429-39.

Gabaix, Xavier; Laibson, David; Moloche, Guillermo and Weinberg, Stephen. " Information Acquisition: Experimental Analysis of a Boundedly Rational Model." *American Economic Review*, 2006, *forthcoming*.

Gneezy, Uri. "Deception: The Role of Consequences." *American Economic Review*, 2005, *95*(1), pp. 384-94.

Goldwater, B. C. "Psychological Significance of Pupillary Movements." *Psychological Bulletin*, 1972, *77*(5), pp. 340-55.

Green, Jerry and Laffont, Jean-Jacques. "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods." *Econometrica*, 1977, *45*(2), pp. 427-38.

Groves, Theodore. "Incentives in Teams." *Econometrica*, 1973, *41*(4), pp. 617-31.

Hall, Brian J. and Murphy, Kevin J. "The Trouble with Stock Options." *Journal of Economic Perspectives*, 2003, *17*(3), pp. 49-70.

Heilveil, I. "Deception and Pupil Size." *Journal of Clinical Psychology*, 1976, *32*(3), pp. 675-76.

Hicks, R. A.; Reaney, T. and Hill, L. "Effects of Pupil Size and Facial Angle on Preference for Photographs of a Young Woman." *Perceptual and Motor Skills*, 1967, *24*(2), pp. 388-&.

Hunton, James E. and McEwen, Ruth A. "An Assessment of the Relation between Analysts' Earnings Forecast Accuracy, Motivational Incentives and Cognitive Information Search Strategy." *Accounting Review*, 1997, *72*(4), pp. 497-515.

Jacob, Brian A. and Levitt, Steven D. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics*, 2003, *118*(3), pp. 843-77.

Janisse, Michel P. "Pupil Size and Affect - Critical Review of Literature since 1960." *Canadian*

Psychologist, 1973, *14*(4), pp. 311-29.

Janisse, Michel P. and Bradley, M. T. "Deception, Information and the Pupillary Response." *Perceptual and Motor Skills*, 1980, *50*(3), pp. 748-50.

Johnson, Eric J.; Camerer, Colin; Sen, Sankar and Rymon, Talia. "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining." *Journal of Economic Theory*, 2002, *104*(1), pp. 16-47.

Johnson, Eric J. and Camerer, Colin F. "Thinking Backward and Forward in Games," I. Brocas and J. Castillo, *The Psychology of Economic Decisions, Vol.2: Reasons and Choices.* Oxford University Press, 2004,

Kozel, F. Andrew; Revell, Letty J.; Lorberbaum, Jeffrey P.; Shastri, Ananda; Elhai, Jon D.; Horner, Michael David; Smith, Adam; Nahas, Ziad; Bohning, Daryl E. and George, Mark S. "A Pilot Study of Functional Magnetic Resonance Imaging Brain Correlates of Deception in Healthy Young Men." *Journal of Neuropsychiatry and Clinical Neurosciences*, 2004, *16*, pp. 295-305.

Kraut, Robert E. "Humans as Lie Detectors: Some Second Thoughts." *Journal of Communication*, 1980, *30*, pp. 209-16.

Langleben, D. D.; Schoroeder, L.; Maldjian, J. A.; Gur, R. C.; McDonald, S.; Ragland, J. D.; O'Brien, C. P. and Childress, A. R. "Brain Activity During Simulated Deception: An Event-Related Functional Magnetic Resonance Study." *NeuroImage*, 2002, *15*(3), pp. 727-32.

Lin, Hsiou-wei and McNichols, Maureen F. "Underwriting Relationships, Analysts' Earnings Forecasts and Investment Recommendations." *Journal of Accounting and Economics*, 1998, *25*(1), pp. 101-27.

Lubow, R. E. and Fein, Ofer. "Pupillary Size in Response to a Visual Guilty Knowledge Test:

New Technique for the Detection of Deception." *Journal of Experimental Psychology-Applied*, 1996, *2*(2), pp. 164-77.

Mann, Samantha; Vrij, Aldert and Bull, Ray. Detecting True Lies: Police Officers' Ability to Detect Suspects' Lies, Unpublished paper, 2005.

McKelvey, Richard D. and Palfrey, Thomas R. "Quantal Response Equilibria for Extensive Form Games." *Experimental Economics*, 1998, *1*(1), pp. 9-41.

Michaely, Roni and Womack, Kent L. "Conflict of Interest and the Credibility of Underwriter Analyst Recommendations." *Review of Financial Studies*, 1999, *12*(4), pp. 653-86.

Myerson, Roger B. "Incentive Compatibility and the Bargaining Problem." *Econometrica: Journal of the Econometric Society*, 1979, *47*(1), pp. 61-74.

Oka, Shunichi; Chapman, C. Richard and Jacobson, Robert C. "Phasic Pupil Dilation Response to Noxious Stimulation: Effects of Conduction Distance. Sex, and Age." *Journal of Psychophysiology*, 2000, *14*(2), pp. 97-105.

Ottaviani, Macro and Squintani, Francesco. Non-Fully Strategic Information Transmission, Unpublished paper, 2004.

Pelli, Denis G. "The Videotoolbox Software for Visual Psychophysics: Transforming Numbers into Movies." *Spatial Vision*, 1997, *10*, pp. 437-42.

Rosovsky, Henry and Hartley, Matthew. "Evaluation and the Academy: Are We Doing the Right Thing? Grade Inflation and Letters of Recommendation," Cambridge, MA: American Academy of Arts and Sciences, 2002.

Spence, Sean A.; Farrow, Tom F. D.; Herford, Amy E.; Wilkinson, Iain D.; Zheng, Ying and Woodruff, Peter W. R. "Behavioural and Functional Anatomical Correlates of Deception in Humans." *NeuroReport*, 2001, *12*(13), pp. 2849-53.

Vrij, Aldert. *Detecting Lies and Deceit: The Psychology of Lying and the Implications for Professional Practice*. Chichester: Wiley and Sons, 2000.

Appendix: Methods

Eyetracking data and button responses were recorded using the mobile Eyelink II head-mounted eyetracking system (SR Research, Osgoode, Ontario). Eyetracking data were recorded at 250 Hz. The mobile Eyelink II is a pair of tiny cameras mounted on a lightweight rack facing toward the subjects' eyes, and supported by comfortable head straps. Subjects can move their heads and a period of calibration adjusts for head movement to infer accurately where the subject is looking. New nine-point calibrations and validations were performed prior to the start of each experiment in a participant's session. Accuracy in the validations typically was better than 0.5° of visual angle. Experiments were run under Windows XP (Microsoft, Inc.) in Matlab (Mathworks, Inc., Natick, MA) using the Psychophysics Toolbox (David H. Brainard, 1997; Denis G. Pelli, 1997) and the Eyelink Toolbox (Frans W. Cornelissen et al., 2002).

Eyetracking data were analyzed for fixations using the Eyelink Data Viewer (SR Research, Hamilton, Ontario). In discriminating fixations, we set saccade velocity, acceleration, and motion thresholds to 30°/sec, 9500°/sec^2, and 0.15°, respectively. Regions of interest (ROIs), or the boxes subject look up, were drawn on each task image using the drawing functions within the Data Viewer. Measures of gaze included Fixation Number (i.e., the total number of fixations within an ROI) and Fractional Dwell Time (i.e., the time during a given round spent fixating a given ROI divided by the total time between image onset and response). Only those fixations beginning between 50ms following the onset of a task image and offset of the task image were considered for analysis.

All task images were presented on a CRT monitor (15.9 in x 11.9 in) operating at 85 or 100 Hz vertical refresh rate with a resolution of 1600 pixels x 1200 pixels, and at an eye-to-screen distance of approximately 24 inches, thus subtending ~36 degrees of visual angle.

Table 1: Actual Information Transmission

BIAS	Corr(S, M)	Corr(M, A)	Corr(S, A)	Predicted Corr(S, A)
0	0.990	0.998	0.988	1.000
1	0.725	0.738	0.721	0.707
2	0.630	0.571	0.497	0.000

Table 2: Sender and Receiver's Payoffs

BIAS	u_S (std)	u_R (std)	Predicted u_R (std)
0	109.14 (4.07)[a]	109.14 (4.07)[a]	110.00 (0.00)
1	93. 35 (20.75)	94.01 (19.86)	87.38 (18.88)
2	41.52 (49.98)	85.52 (25.60)	71.59 (27.26)

Note: [a] Payoffs are exactly the same for senders and receivers due to the symmetry of the payoffs when b=0.

Table 3: Average Sender Fixation Counts and Lookup Time across Game Parameters

BIAS	Response time (sec.)	State		Bias		Sender Payoffs		Receiver Payoffs	
		Fixation (count)	Lookup (sec.)	Fixation (count)	Lookup (sec.)	Fixation (count)	Lookup (sec.)	Fixation (count)	Lookup (sec.)
0	2.99	2.6	0.65	2.1	0.41	3.0	0.73	1.4	0.27
1	6.10	5.0	1.47	3.9	0.99	8.1	2.29	3.9	1.05
2	8.62	6.2	1.72	5.5	1.52	10.6	3.03	5.4	1.50
overall	6.16	4.7	1.34	4.0	1.02	7.6	2.14	3.7	1.00

Table 4: Average Fixation Counts and Lookup Time per Row

BIAS	True State Rows		Other Rows	
	Fixation Counts (counts per row)	Lookup Time (sec. per row)	Fixation Counts (counts per row)	Lookup Time (sec. per row)
0	2.2	0.54	0.5	0.11
1	6.8	2.06	1.3	0.32
2	7.8	2.24	2.0	0.57
overall	5.9	1.71	1.3	0.36

Table 5: Pupil Size Regressions for 400 milliseconds Intervals (standard errors in parentheses)

Y	$PUPIL_i$	-1.2~ -0.8sec	-0.8~ -0.4sec	-0.4~ 0.0sec	0.0~ 0.4sec	0.4~ 0.8sec
constant	α	99.56***	99.53**	103.78***	103.12***	102.90***
		(4.90)	(4.54)	(4.15)	(4.00)	(4.04)
LIE_SIZE * BIAS interactions	β_{10}	-1.51	-9.84	-1.08	-2.18	8.16
		(5.52)	(5.09)	(5.34)	(4.05)	(4.64)
	β_{11}	0.94	2.19*	2.69***	4.30***	3.89***
		(0.94)	(0.88)	(0.80)	(0.76)	(0.77)
	β_{12}	3.08***	3.15***	3.13***	2.77***	3.18***
		(0.91)	(0.85)	(0.78)	(0.74)	(0.76)
BIAS effects (b=2 benchmark)	β_{20}	1.96	1.01	-1.60	-0.38	1.94
		(1.57)	(1.46)	(1.34)	(1.28)	(1.30)
	β_{21}	2.99	1.86	1.23	-0.42	1.46
		(1.71)	(1.59)	(1.45)	(1.39)	(1.41)
STATE effects (S=3 benchmark)	β_{31}	-1.58	-0.69	-1.34	-7.52***	-8.88***
		(1.39)	(1.27)	(1.16)	(1.11)	(1.12)
	β_{32}	-1.37	-1.08	-1.40	-2.17	-2.87*
		(1.51)	(1.41)	(1.29)	(1.21)	(1.23)
	β_{34}	1.26	3.53*	1.89	4.94***	3.97**
		(1.50)	(1.40)	(1.27)	(1.21)	(1.23)
	β_{35}	1.49	1.89	2.09	6.20***	5.69***
		(1.67)	(1.55)	(1.42)	(1.35)	(1.38)
	N	497	495	497	510	505
	F	5.63	8.72	13.71	14.33	16.43
	R^2	0.291	0.407	0.530	0.535	0.574

Note: t-Test p-values lower than *5 percent, ** 1 percent, and *** 0.1 percent.

Note: Regarding individual differences (using subject 6 as benchmark), before the decision time, only two or three of the α_k are significant, mainly for subject 2 and 7. After the decision is made, individual differences increase slightly and have 3-5 significant α_k.

Table 6: Order Logit Results of Deception and Size of Deception (standard errors in parentheses)

X	Y	LIE (all)	LIE (part)	LIE_SIZE (all)	LIE_SIZE (part)
BIAS=1	β_1	-0.57	-0.64	-0.65*	-0.69*
		(0.44)	(0.55)	(0.27)	(0.22)
STATE=1	β_{21}	-0.68	-0.75	-0.03	-0.04
		(0.53)	(0.46)	(0.35)	(0.28)
STATE=2	β_{22}	0.09	0.20	0.19	0.18
		(0.57)	(0.52)	(0.34)	(0.27)
STATE=4	β_{24}	-0.17	-0.04	-1.05*	-1.11*
		(0.62)	(0.51)	(0.40)	(0.34)
STATE=5	β_{25}	-5.72*	-6.20*	-5.10*	-5.42*
		(0.76)	(1.25)	(0.55)	(0.64)
$(PUPIL_{end}-PUPIL_{ini})$ * BIAS=1	β_{31}	-0.12	-0.10	0.12	0.14
		(0.23)	(0.21)	(0.15)	(0.13)
$(PUPIL_{end}-PUPIL_{ini})$ * BIAS=2	β_{32}	1.22*	1.37*	0.97*	0.96*
		(0.33)	(0.54)	(0.22)	(0.25)
total observations N^{a}		366	366	366	366
N used in estimation		366	242.0	366	243.8
N used to predict		366	124.0	366	122.2
Percent of wrong prediction (b=1)		7.0	16.6	26.5	34.3
Percent of wrong prediction (b=2)		11.1	21.3	26.0	34.2
Percent of errors of size (1,2,3+) (b=1)		N/A	N/A	(94,4,2)	(87,11,3)
Percent of errors of size (1,2,3+) (b=2)		N/A	N/A	(79,15,6)	(78,15,7)

Note: * Denotes $p<0.05$ (t-test).
[a] 20 observations without the needed pupil size measures are excluded.

Table 7: Predicting True States (Resampling 100 times) (standard errors in parentheses)

X	Y	Actual	STATE (model 1)	STATE (model 2)
BIAS=1	β_1	-	-1.50	-1.72
		-	(0.83)	(0.88)
MESSAGE * BIAS = 1	β_{21}	-	0.81**	1.15**
		-	(0.24)	(0.28)
MESSAGE * BIAS = 2	β_{22}	-	0.77**	0.96**
		-	(0.23)	(0.27)
ROW $_{self}$ * BIAS=1	β_{31}	-	0.97**	0.94*
		-	(0.23)	(0.29)
ROW $_{self}$ * BIAS=2	β_{32}	-	0.96*	1.02*
		-	(0.29)	(0.33)
ROW $_{other}$ * BIAS=1	β_{41}	-	0.23	0.19
		-	(0.17)	(0.19)
ROW $_{other}$ * BIAS=2	β_{42}	-	0.34*	0.28
		-	(0.16)	(0.17)
$(PUPIL_{end}-PUPIL_{ini})$	β_5	-	-	1.17**
		-	-	(0.27)
$(PUPIL_{end}-PUPIL_{ini})$	β_6	-	-	-0.25**
* MESSAGE		-	-	(0.07)
total observations N [a]		208	208	201
N used in estimation		-	138.4	134.3
N used to predict		-	69.6	66.7
Percent of wrong prediction (b=1)		56.2	29.4	33.5
Percent of errors of size (1,2,3+) (b=1)		(80, 15, 5)	(73, 21, 6)	(74, 20, 6)
Average predicted payoff (b=1) [b]		93.4 (22.3)	100.7* (2.8)	99.3* (3.3)
Percent of wrong prediction (b=2)		70.9	60.2	58.7
Percent of errors of size (1,2,3+) (b=2)		(67, 26, 7)	(74, 21, 5)	(79, 16, 5)
Average predicted payoff (b=2) [b]		86.2 (23.8)	91.8* (3.3)	93.1* (3.4)

Note: * and ** Denotes p<0.05 and p<0.001 (t-test)
  [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 receiver subjects in the experiment.
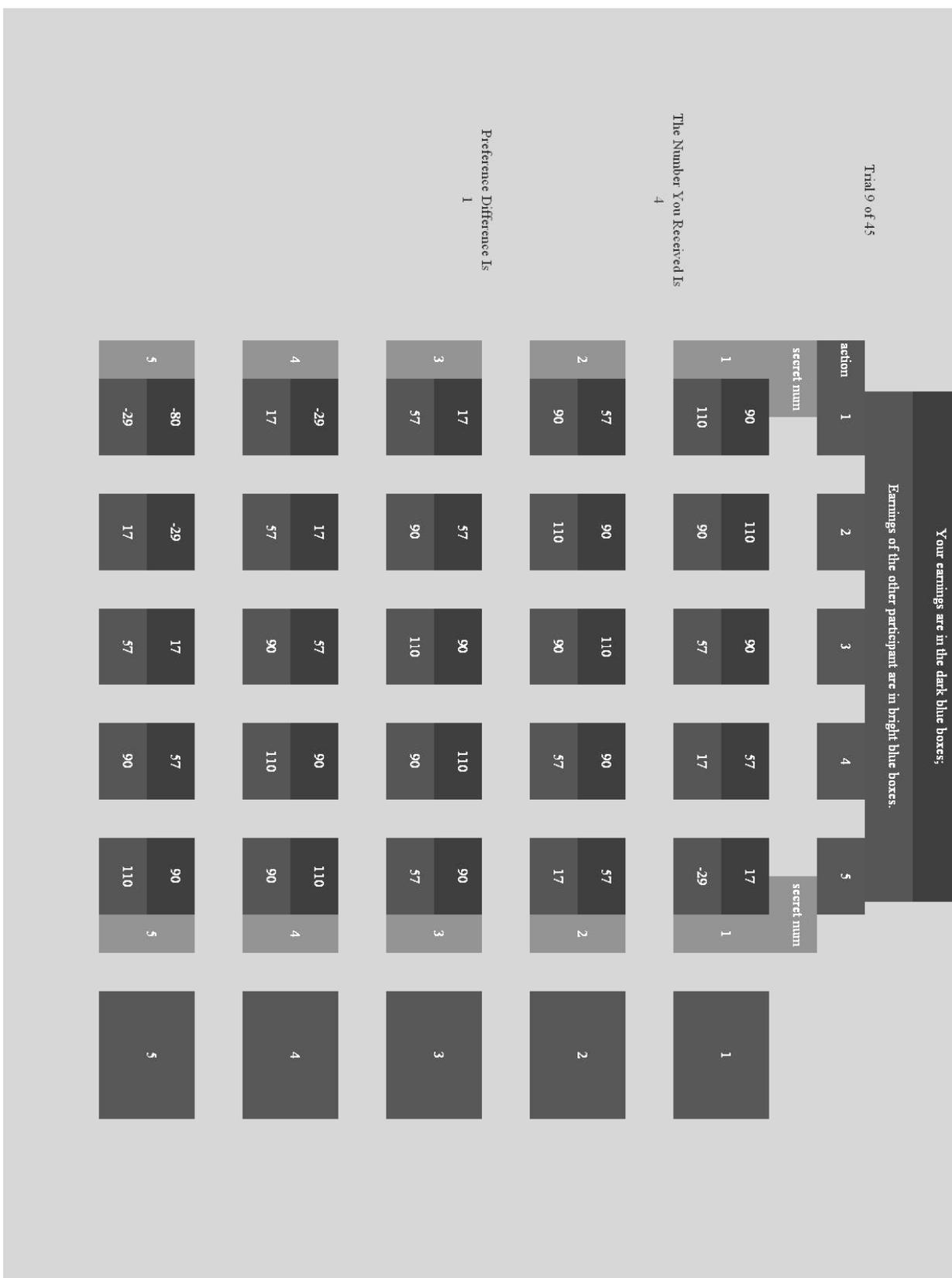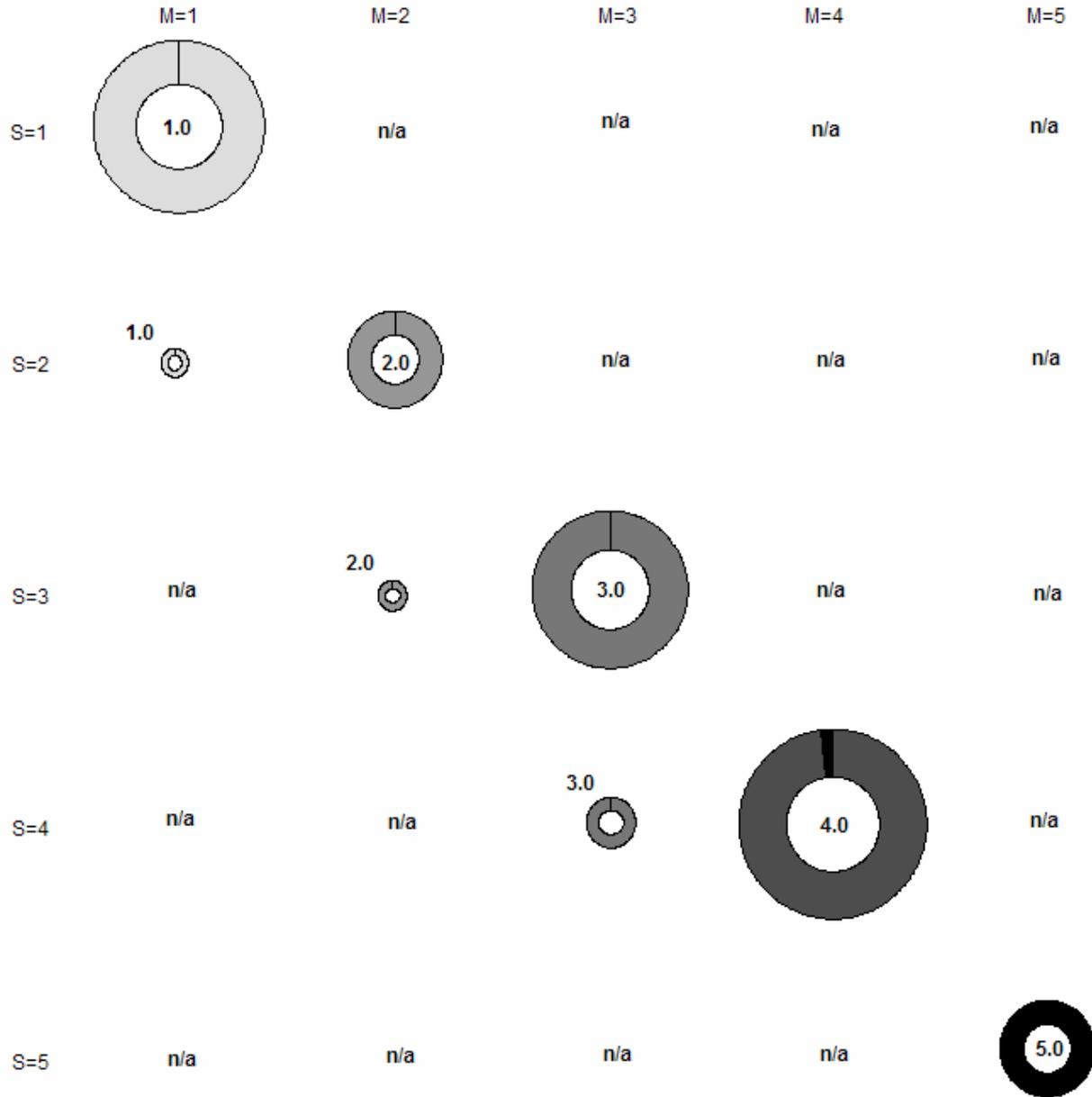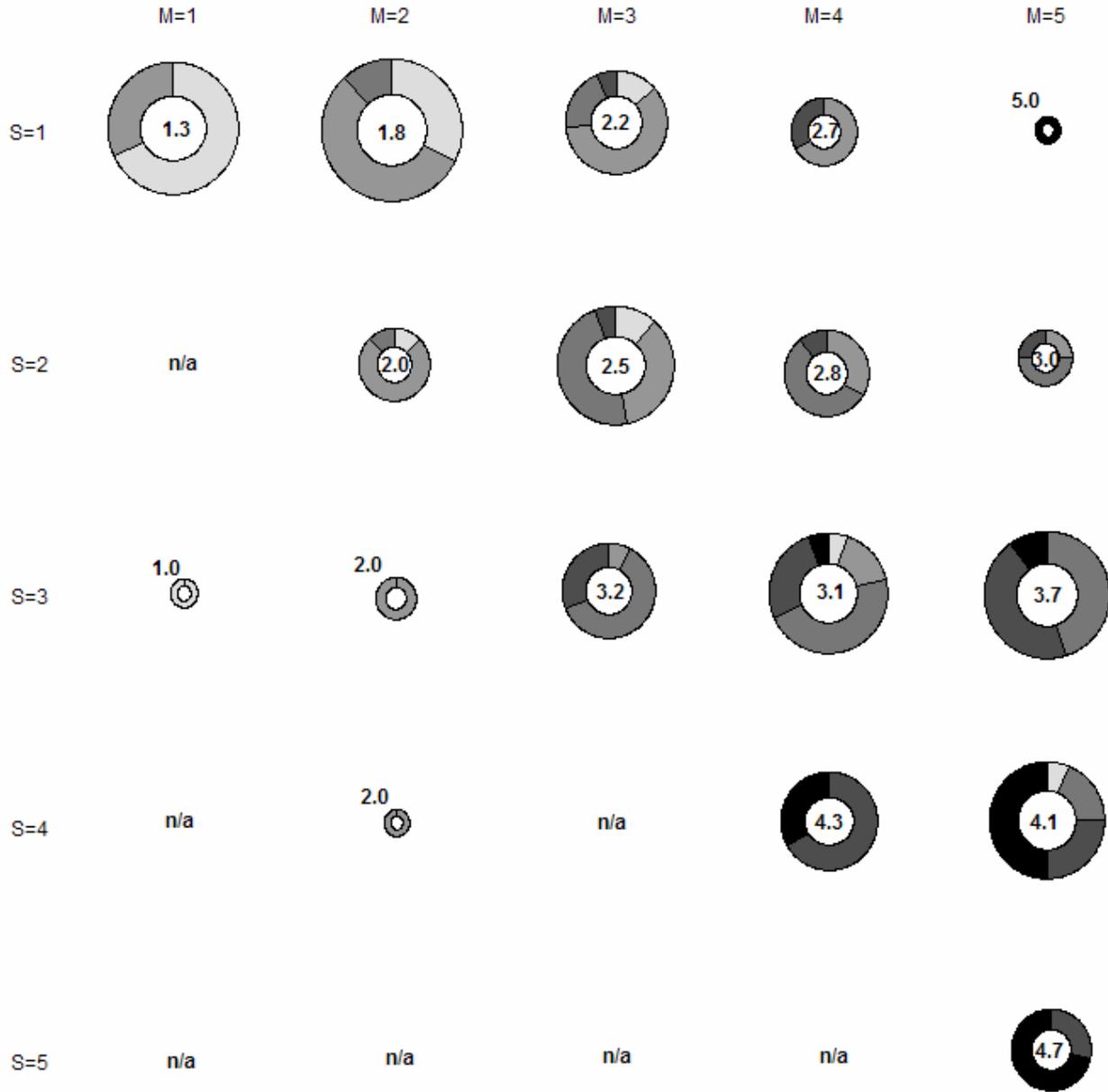
Figure 1: Sender Screen for b=1 and S=4

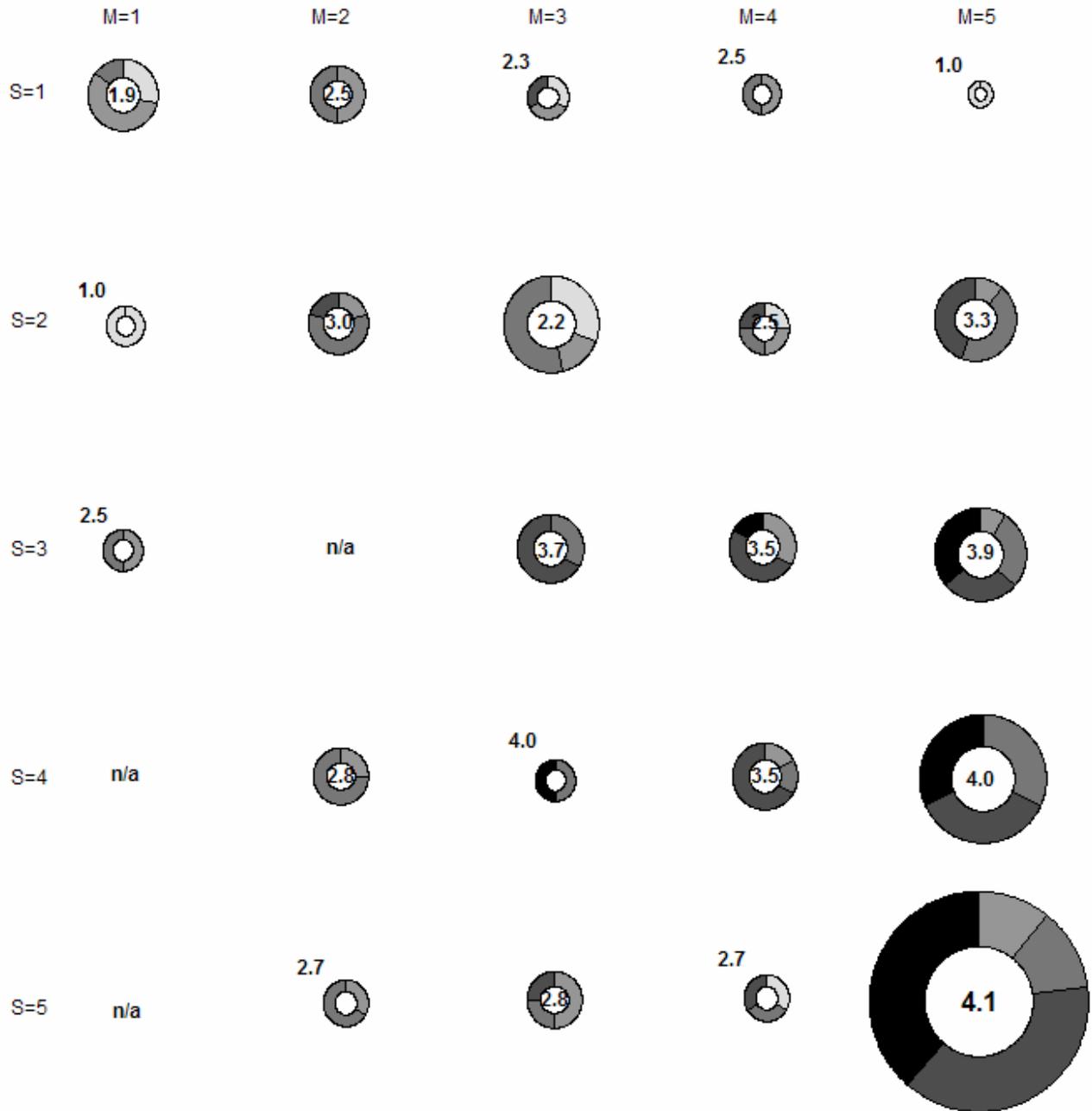**Figure 2: Raw Data Pie Charts (b=0)**

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 break down 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.

Figure 3: Raw Data Pie Chart (b=1)

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 break down 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.

Figure 4: Raw Data Pie Chart (b=2)

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 break down 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.

Figure 5: Lookup Icon Graph for b=1
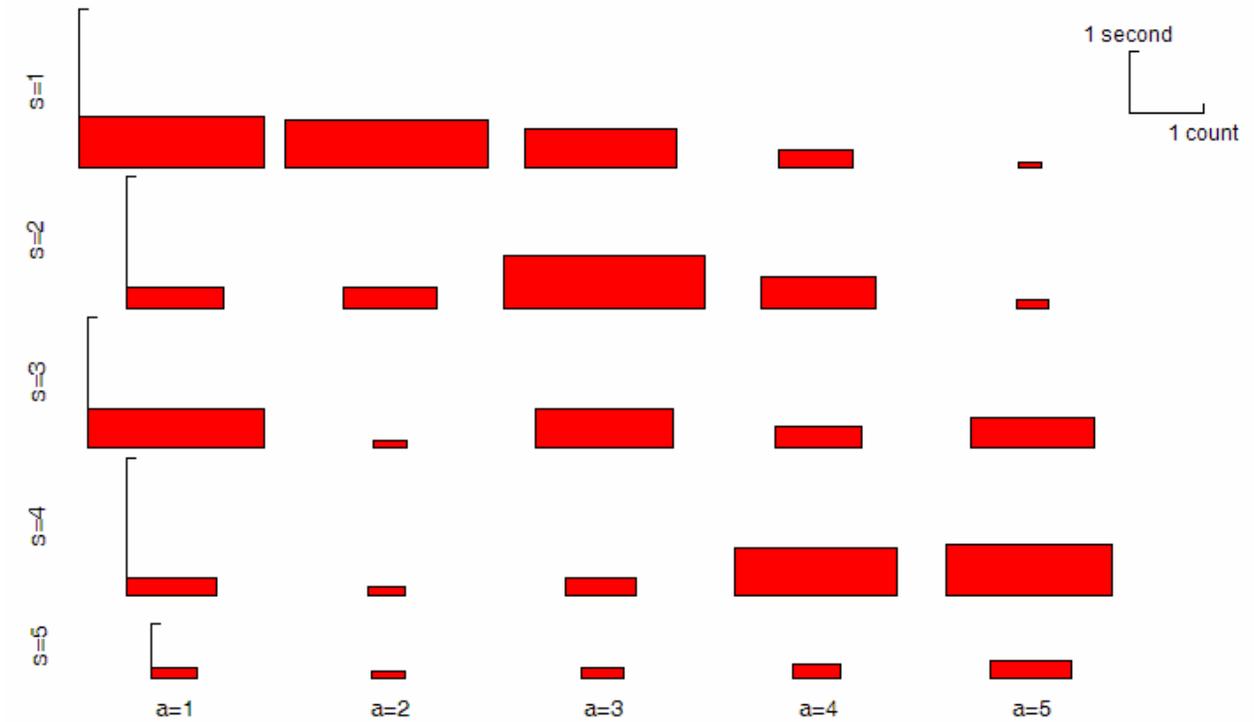
Part (a): Sender Payoffs
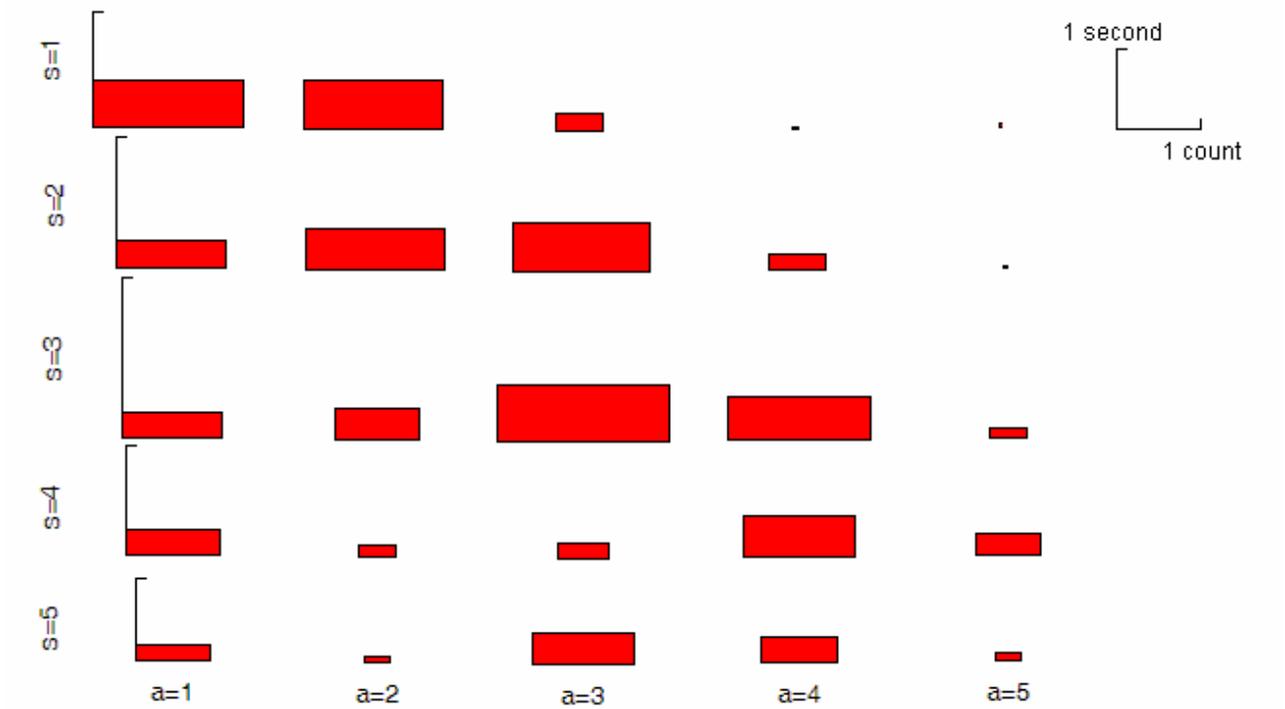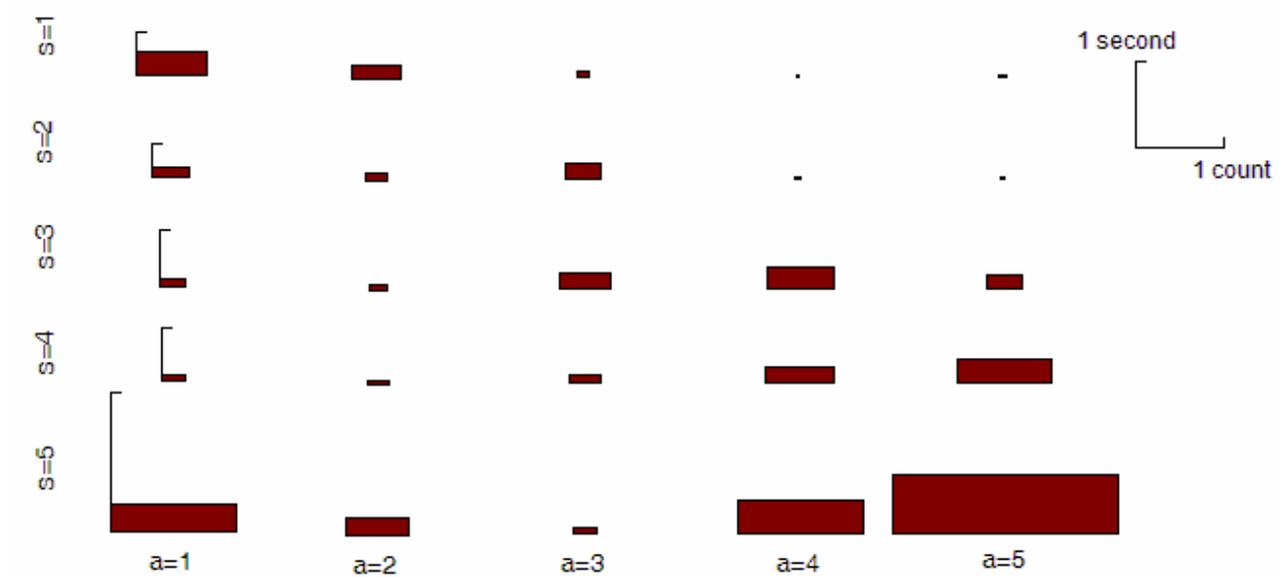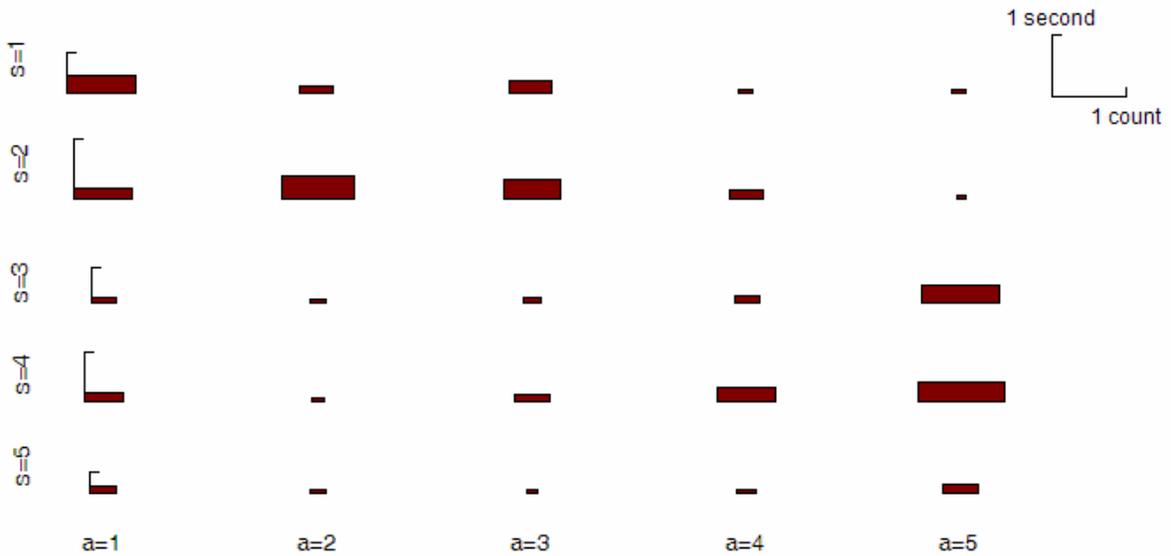

Part (b): Receiver Payoffs

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.

Figure 6: Lookup Icon Graph for b=2

Part (a): Sender Payoffs

Part (b): Receiver Payoffs

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.

Appendix: Experiment Instructions

The experiment you are participating in consists of 1 session, having 45 rounds. At the end of the last session, you will be asked to fill out a questionnaire and paid the total amount you have accumulated during the course of the sessions in addition to a $5 show-up fee. Everybody will be paid in private after showing the record sheet. You are under no obligation to tell others how much you earned.

During the experiment all the earnings are denominated in FRANCS. Your dollar earnings are determined by the FRANC/$ exchange rate: 200 FRANCS = $1.

In each round, the computer program generates a secret number that is randomly drawn from the set {1,2,3,4,5}. The computer will display this secret number on member A's screen. After receiving the number, member A will send the message "The number I received is XX," to member B by staring at box XX. Hearing the message from member A, member B will then choose an action. In particular, member B can choose action 1, 2, 3, 4, or 5, using the game pad. Earnings of both members depend on the secret number and member B's action.

Member B's earnings is higher when member B's action is closer to the secret number, while member A's earnings is higher when member B's action is closer to the secret number **plus the preference difference**. The preference difference is either 0, 1 or 2, with equal chance, and will also be displayed and announced at the beginning of each round.

For example, if the preference difference is 2 and the secret number is 3, member B's earnings are higher if his or her action is closer to 3. However, member A's earnings is higher when member B's action is closer to 3 + 2 = 5. The earning tables are provided to you for convenience.

To summarize, in each round, the computer will display the preference difference and the secret number on member A's screen. Then, member A stares at a box (on the right) containing the desired message. Member B will hear the preference difference and the message "The number I received is XX," and then choose an action. The secret number is revealed after this choice, and earnings are determined accordingly.

Practice Session: 3 Rounds

Session 1: 45 Rounds

Member B: Please make sure you record the earnings in your record sheet. **Your payments will be rounded up.** Thank you for your participation.

Appendix Figure 1: Icon graph of lookups (rectangle width) and looking time (shaded area) for b=0
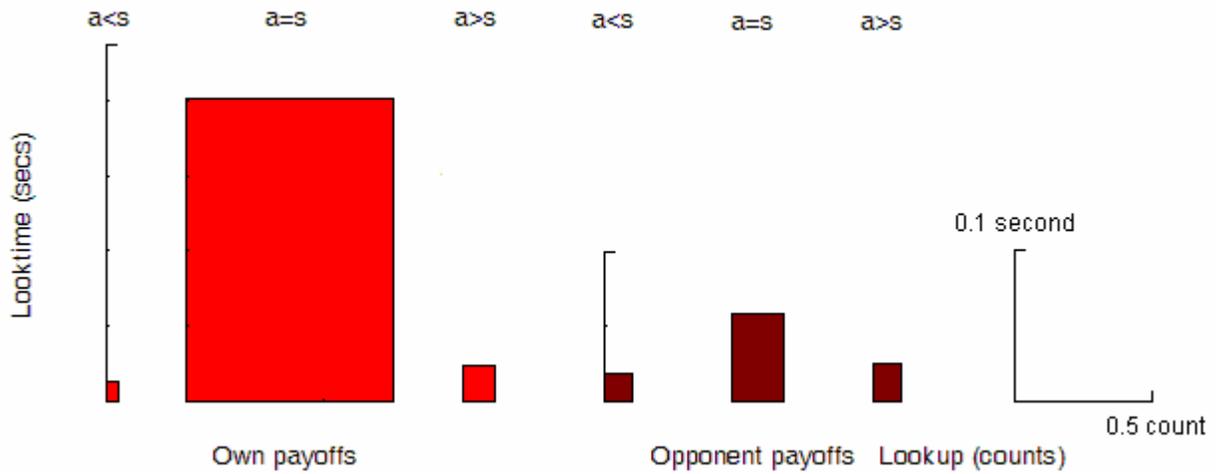
Appendix Table 1: Average response time change for different biases

Bias	N	Average for first 15 rounds	N	Average for middle 15 rounds	N	Average for last 15 rounds
0	38	4.69	47	2.79	55	1.99
1	73	7.74	60	4.83	59	5.36
2	67	9.02	68	8.78	51	7.90
overall	178	7.57	175	5.81	165	5.02

* The numbers of observations are slightly different because we exclude 10 rounds where subjects had to used the keyboard to make their decision. Also, subject #4 had severe pain and the experimenter was forced to stop the experiment at the end of round 33.

Note: Since the bias was randomly determined each round, and subject #4 stopped at round 33 (due to excess pain wearing the eyetracker), numbers of observations are not equal. Dropping subject #4 does not change the results.