Biased Recommendations from Biased and Unbiased Experts

Wonsuk Chung and Rick Harbaugh*

Forthcoming *Journal of Economics and Management Strategy*

Abstract

When can you trust an expert to provide honest advice? We develop and experimentally test a recommendation game where an expert helps a decision maker choose among two actions that each benefit the expert and an outside option that does not. For instance, a salesperson recommends one of two products to a customer who may purchase nothing. Behavior is largely consistent with predictions from the cheap talk literature. For sufficient symmetry, recommendations are *persuasive* in that they benefit the expert by lowering the chance that the decision maker takes the outside option. If the expert is known to be biased toward either action, such as a salesperson receiving a higher commission on one product, the decision maker partially *discounts* a recommendation for it and is more likely to take the outside option. If the bias is uncertain, then biased experts lie even more, while unbiased experts follow a *political correctness* strategy of pushing the opposite action so as to be more persuasive. Even if the expert is known to be unbiased, when the decision maker already favors an action the expert *panders* toward it, and the decision maker partially discount the recommendation. The comparative static predictions hold with any degree of lying aversion up to pure cheap talk, and most subjects exhibit some limited lying aversion. The results highlight that transparency of expert incentives can improve communication, but need not ensure unbiased advice.

*Chung: Korea Insurance Research Institute, woschung@indiana.edu; Harbaugh: Department of Business Economics and Public Policy, Kelley School of Business, Indiana University, riharbau@indiana.edu. For helpful comments we thank the editors and referees, and Alexei Alexandrov, Tim Cason, Archishman Chakraborty, Bill Harbaugh, Tarun Jain, Wooyoung Lim, Dmitry Lubensky, Stephen Morris, Michael Raith, Eric Rasmusen, Eric Schmidbauer, Irene Skricki, Joel Sobel, Yossi Spiegel, Lise Vesterlund, Jimmy Walker, Stephanie Wang, Alistair Wilson, and seminar and conference participants at the University of Pittsburgh, Michigan State University, Booth School of Business, Simon School of Business, Indian School of Business, the Western Economics Association meetings, the Fifth Conference on the Economics of Advertising and Marketing at Qinghua University's School of Management and Economics, and the 2013 Boulder Summer Conference on Consumer Financial Decision Making.

1 Introduction

Experts provide information about different choices to decision makers. But experts often benefit more from some choices than from others, such as a salesperson who earns a higher commission on a more expensive product and receives nothing if the customer walks away. Can an expert's recommendation still be persuasive with such conflicts of interest, or will it be completely discounted? What if the decision maker suspects that the expert benefits more from one choice but is not sure? And how is communication affected if the expert knows that the decision maker is already leaning toward a particular choice?

These issues are important to the design of incentive and information environments in which experts provide advice. In recent years, the incentives of mortgage brokers to recommend high-cost loans (Agarwal et al., 2014), of financial advisors to provide self-serving advice (Egan et al., 2016), of doctors to prescribe expensive drugs (Iizuka, 2012), and of media firms to push favored agendas (DellaVigna and Kaplan, 2007) have all come under scrutiny. Can such problems be resolved by requiring disclosure of any conflicts of interest, or is it necessary to eliminate biased incentives?[1] And are unbiased incentives always sufficient to ensure unbiased advice?

To gain insight into such questions, several papers have applied the cheap talk approach of Crawford and Sobel (1982) to discrete choice environments where an expert has private information about different actions and the decision maker has an outside option that is the expert's least favored choice (e.g., De Jaegher and Jegers, 2001; Chakraborty and Harbaugh, 2007, 2010; Che, Dessein, and Kartik, 2013).[2] Based on this literature, we develop and test a simplified recommendation game in which an expert knows which of two actions is better for a decision maker, and may have an incentive to push one of the actions more than the other. The decision maker can take either action or pursue an outside option. Despite its simplicity, the model captures several key phenomena from the literature.

First, for sufficient payoff symmetry, recommendations are "persuasive" in that they benefit the expert by reducing the chance that the decision maker walks away without taking either action. Even though a recommendation is only cheap talk, it is still credible since it raises the expected value of one action at the opportunity cost of lowering the expected value of the other action. Such "comparative cheap talk" is persuasive since the higher expected value of the recommended action

[1] Conflict of interest disclosure may be imposed by regulators (see for instance, the SEC's regulation of investment advisors), or voluntarily adopted, as many medical journals have done for authors. Regulations can impose more equal incentives, e.g., requirements for "firewalls" that limit the gains to stock analysts from pushing their firm's clients, and incentives may also be adjusted voluntarily to increase credibility, e.g., Best Buy promotes its "non-commissioned sales professionals" whose "first priority is to help you make the right purchasing decision".

[2] See also Bolton, Freixas, and Shapiro (2007), Inderst and Ottaviani (2012), Chung (2012), and Chakraborty and Harbaugh (2014).

is now more likely to exceed the decision maker's outside option. For instance, a customer is more likely to make a purchase if a recommendation persuades him that at least one of two comparably priced products is of high quality. In our experimental results, we find that recommendations are usually accepted, and are almost always accepted when the decision maker's outside option is unattractive.

Second, when the expert is known to be biased in the sense of having a stronger incentive to push one action, a recommendation for that action is "discounted" in that the decision maker is more likely to ignore the recommendation and stick with the outside option. Therefore, in equilibrium, the expert faces a tradeoff where one recommendation generates a higher payoff if it is accepted while the other recommendation is more likely to be accepted. In our experiment, we find that experts are significantly more likely to recommend the more incentivized action than the other action, while decision makers are significantly less likely to accept a recommendation for the more incentivized action than the other action.

Third, when the decision maker is known to already favor one action before listening to the expert, the expert benefits by "pandering" to the decision maker and recommending that action even when the other action is actually better. Hence, biased recommendations can arise even when the expert's incentives for each action are the same (Che, Dessein, and Kartik, 2013). The decision maker anticipates such pandering and, just as in the asymmetric incentives case, discounts a recommendation for that action so that the incentive to lie is mitigated. In our experiment, we find that experts are significantly more likely to recommend the favored action than the other action, and decision makers are significant more likely to discount such a recommendation than when the prior distribution of decision maker values is symmetric.

Finally, when the decision maker is unsure of whether the expert is biased toward an action, a recommendation for that action is suspicious and hence discounted by the decision maker, so an unbiased expert has a "political correctness" incentive to recommend the opposite action even if it is not the better choice (c.f., Morris, 2001). For instance, if a salesperson is suspected to benefit more from selling one product than another product but in fact has equal incentives, then pushing the other product is more attractive since it is more likely to generate a sale. In our experiment we find that, as predicted, lack of transparency induces unbiased experts to make the opposite recommendation from that made by biased experts. Decision makers do not appear to sufficiently discount such recommendations, suggesting that they may not always anticipate how lack of transparency warps the incentives of even unbiased experts.[3]

Most of the experimental literature on cheap talk has focused on testing different implications of the original Crawford and Sobel (1982) model, and typically finds that experts are somewhat

[3] Our "unbiased" experts are still biased against the outside option, rather than sharing the decision maker's preferences (e.g., Morgan and Stocken, 2003).

lying averse (e.g., Dickhaut, McCabe, and Mukherji, 1995).[4] Based on this literature we expect subjects to be reluctant to lie, and in particular we expect the strength of this aversion to vary across subjects (Gibson et al., 2013).[5] Therefore we depart from a "pure" cheap talk approach and assume that the expert has a lying cost that is drawn from a continuous distribution with support that could be arbitrarily concentrated near zero.[6] Lying aversion is not necessary for communication in our game but its inclusion makes the model more testable by eliminating extra equilibria that involve babbling/pooling, that have messages implying the opposite of their literal meanings, and that have strategic mixing between messages.[7]

In our experimental tests, we cannot control for varying subject preferences against lying, so the exact lying rates and acceptance rates cannot be predicted beforehand. However, we find that the comparative static predictions of the model are the same for any distribution of lying costs, so we can test these predictions even without knowing the exact distribution of subject preferences against lying. Moreover, since communication incentives in the model with lying costs are still driven primarily by the endogenous opportunity cost of messages, these comparative static predictions are the same as in the most intuitive equilibrium of the limiting case of pure cheap talk. Hence the model's predictions are consistent with the insights generated by the theoretical cheap talk literature.

Our experimental results on the benefits of transparency contrast with results in the literature based on different models. Cain, Loewenstein, and Moore (2005) find experimentally that subjects do not fully discount recommendations by experts with known biases. We find that discounting rates by decision makers are instead slightly higher than the best response rate of discounting.

[4]See also Cai and Wang (2006), Sánchez-Pagés and Vorsatz (2007), and Wang, Spezio and Camerer (2010). This literature has been extended to test Battaglini's (2002) results on cheap talk by competing experts in a multidimensional version of the Crawford-Sobel model (Lai, Lim, and Wang, 2015; Vespa and Wilson, 2016). In a costly communication model, Lafky and Wilson (2015) investigate how incentives for information provision can both encourage communication and make it less credible. Blume, DeJong, Kim, and Sprinkle (1998, 2001) test players' ability to successfully coordinate on the equilibrium meaning of messages when preferences are aligned or partly aligned. There is also a large literature on pre-play communication about strategic intentions in games with complete information (Crawford, 2003).

[5]We do not consider the range of factors that can affect lying aversion (Gneezy, 2005; Sutter, 2009; Sobel, 2016; Gneezy et al., 2018). To reduce the reputational costs of lying (Abeler, Nosenzo, and Raymond, 2016), we use random matching and emphasize the anonymity of the procedure to subjects.

[6]Since messages have a direct effect on payoffs as in a (costly) signaling game, the model is a form of "costly talk" (Kartik, Ottaviani, and Squintani, 2007) or, as lying costs become arbitrarily small, "almost cheap talk" (Kartik, 2009).

[7]With no lying aversion or homogeneous lying aversion, the expert is indifferent in equilibrium and uses a mixed strategy of lying or not. Heterogeneous lying aversion purifies these strategies since experts with lower lying costs strictly prefer to lie while those with higher lying costs strictly prefer to tell the truth. As shown in Section 4.3, we find evidence for heterogeneous lying aversion in our experiment.

They also find that lying is less common when biases are undisclosed because disclosure can allow experts to feel more license to exaggerate. We find that, as predicted, lying rates are higher when decision makers are unsure if the expert is biased. In the theory literature, Li and Madarász (2008) find that disclosure of the expert's bias in the Crawford-Sobel model can harm communication. In a model closer to ours, Chakraborty and Harbaugh (2010) find that disclosure of biases ensures the existence of informative equilibria, while lack of transparency can undermine communication but need not preclude it.

The model's predictions are related to the literature on credence goods, which examines recommendations to buy a cheap or an expensive version of a product (Darby and Karni, 1973).[8] Building on the Pitchik and Schotter (1987) model, De Jaegher and Jegers (2001) consider a doctor who recommends either a cheap or an expensive treatment to a patient whose condition is either severe or not; the expensive treatment works for both conditions, but the cheap treatment works only if the condition is not severe. For some parameter ranges, their model has a mixed strategy equilibrium with aspects of pandering and discounting since a cheap treatment is more attractive for the patient and an expensive treatment is more lucrative for the doctor.

The predictions are also related to the literature on cheap talk in repeated games with a binary decision and no outside option. Sobel (1985) shows that, in such games, recommendations are discounted based on the strength of the expert's incentive to push their favored action relative to reputational costs. Morris (2001) shows how an unbiased expert avoids making the same recommendation as a biased expert so as to maintain a reputation for not being biased. Our one-period model captures similar tradeoffs in an easily testable framework. Rather than leading to a future reputational loss, being perceived as biased leads to the immediate loss that the decision maker is more likely to take the outside option.

In the following section, we present the simplified recommendation game and provide testable hypotheses based on its equilibrium properties. We then outline how we test the hypotheses experimentally and report on the experiment results.

2 Recommendation Game

A decision maker chooses one of two actions, A or B, or an outside option C. These options have respective values to the decision maker of v_A, v_B, and v_C. An expert knows v_A and v_B but the decision maker only knows their distribution. The decision maker knows v_C but the expert only knows its distribution. The expert earns $\pi_A > 0$ or $\pi_B > 0$ if the decision maker chooses action A or B respectively, but earns nothing ($\pi_C = 0$) if the outside option C is chosen. The expert first sends a message $m \in \{m_A, m_B\}$ to the decision maker. The decision maker observes the message

[8] Dulleck, Kerschbamer, and Sutter (2011) test this problem in a pricing rather than cheap talk context.

m, learns the value v_C, and then chooses the action A or B or the outside option C to maximize her expected payoffs given her beliefs. The equilibrium concept is Perfect Bayesian Equilibrium so, along the equilibrium path, the decision maker's beliefs follow Bayes rule based on the expert's communication strategy, which maps the expert's information to the message space.

We tailor this general model to facilitate testing in our laboratory experiment. First, we assume that the action values v_A and v_B have a two-point distribution where one action is good and one is bad with equal probability, $\Pr[(v_A, v_B) = (a, 0)] = \Pr[(v_A, v_B) = (0, b)] = 1/2$ for given $0 < a, b \leq 1$. Second, we assume that the outside option value v_C is independently and uniformly distributed over the interval $[0, 1]$. Finally, to reflect previous experimental evidence, we assume that sending m_A when $v_B > v_A$ or sending m_B when $v_A > v_B$ may incur a lying cost d. We start the analysis with the "pure cheap talk" case of $d = 0$.

Let P_A and P_B be the respective probabilities that A and B recommendations are accepted, and conjecture an equilibrium where the decision maker either follows the expert's recommendation or takes the outside option, so the expert's expected payoff from recommending A or B is $\pi_A P_A$ or $\pi_B P_B$, respectively. For a pure cheap talk equilibrium, the expert cannot strictly benefit from always sending one message or the other, so in equilibrium the payoffs must be equal,

$$\pi_A P_A = \pi_B P_B. \tag{1}$$

Hence a higher commission on an action must be offset by a lower probability that a recommendation for the action is accepted. Given the uniform distribution of v_C, these probabilities are $P_A = \Pr[E[v_A|m_A] > v_C] = E[v_A|m_A]$ and $P_B = \Pr[E[v_B|m_B] > v_C] = E[v_A|m_A]$, so the condition is

$$\pi_A E[v_A|m_A] = \pi_B E[v_B|m_B]. \tag{2}$$

If the expert never lies, then $E[v_A|m_A] = a$ and $E[v_B|m_B] = b$, so complete honesty is only possible if incentives and values are symmetric or if any asymmetries exactly offset each other. But even without such a coincidence, some communication is still possible if the decision maker's estimates are updated based on the equilibrium probability that the expert is lying.

To see this, going forward we assume that $\pi_A a \geq \pi_B b$ so the overall incentive and value asymmetry favors A. Let the expert's communication strategy be represented by the probabilities of a "lie" for each state, $\alpha = \Pr[m_A|v_B > v_A]$ and $\beta = \Pr[m_B|v_A > v_B]$. Conjecture that the expert might falsely claim that A is better, $\alpha \geq 0$, but never falsely claims B is better, $\beta = 0$, so

$$\begin{aligned} P_A &= E[v_A|m_A] = \Pr[v_A > v_B|m_A]a = \frac{1-\beta}{1-\beta+\alpha}a = \frac{a}{1+\alpha}, \\ P_B &= E[v_B|m_B] = \Pr[v_B > v_A|m_B]b = \frac{1-\alpha}{1-\alpha+\beta}b = b, \end{aligned} \tag{3}$$

and also $E[v_B|m_A] = b\alpha/(1+\alpha)$ and $E[v_A|m_B] = 0$. As α rises, $E[v_A|m_A]$ falls, so an equilibrium is possible where the decision maker suspiciously discounts a recommendation for the favored action enough that (2) holds. Substituting from (3), in this equilibrium the expert lies in favor of A and B with respective probabilities

$$\alpha = \frac{\pi_A a}{\pi_B b} - 1,\ \beta = 0. \tag{4}$$

Existence of the equilibrium requires $\alpha \in [0,1]$, which from (4) holds if and only if $\pi_A a/\pi_B b \in [1,2]$. It also requires that, as assumed above, the decision maker prefers the recommended action to the other action, $E[v_A|m_A] \geq E[v_B|m_A]$ and $E[v_B|m_B] \geq E[v_A|m_B]$. Substituting α and β from (4), these reduce to $\pi_A a \leq \pi_B(a+b)$ and to $b \geq 0$, which holds by assumption. Combining the conditions, the overall incentive and value asymmetry cannot be too large:

$$1 \leq \frac{\pi_A a}{\pi_B b} \leq \min\{2, \frac{a+b}{b}\}. \tag{5}$$

We use this symmetry restriction, which is necessitated by our simplifying assumption that there are only two states,[9] in our experimental parameterizations.

Proposition 1 (Pure Cheap Talk) *For incentives and values satisfying (5), there exists a pure cheap talk equilibrium with acceptance and lying rates given by (3) and (4).*

Turning now to lying aversion, we assume that the expert knows his own lying cost d, which is the same for lying in either direction, but the decision maker only knows that d has a common-knowledge distribution G with no mass points and support on $[0, \overline{d}]$ for some $\overline{d} > 0$. Since experts have heterogeneous lying costs, any lying must come from experts with lower lying costs, so we are now interested in the pure strategies of expert types with different lying costs.

If each message offers the same monetary payoff, no lying occurs, and if one message offers a higher monetary payoff, then lying must be in that direction only, so $\alpha = 0$ or $\beta = 0$. Continuing to focus on the case where (5) holds, conjecture an equilibrium where $\alpha \geq 0$ and $\beta = 0$. Letting d_A be the lying cost of the marginal expert type who is just indifferent between lying in favor of A or not, the equilibrium condition is

$$\pi_A P_A - d_A = \pi_B P_B, \tag{6}$$

where P_A and P_B are still given by the expected values of each recommended action. Substituting P_A and P_B into (6) and noting that $d_A(\alpha)$ is the quantile function $G^{-1}(\alpha)$, in equilibrium

$$\alpha = \frac{\pi_A a}{\pi_B b + d_A(\alpha)} - 1,\ \beta = 0, \tag{7}$$

[9] If the symmetry conditions are violated then communication can become more complicated or break down. In a richer state space, comparative cheap talk of the form analyzed in this paper remains credible and influential for arbitrary asymmetries (Chakraborty and Harbaugh, 2010; Chung, 2012).

where α and β now represent the fraction of expert types with a pure strategy of lying toward A and B respectively.

Since G is continuous and strictly increasing in α so is d_A, implying that there exists a unique α that solves (7). The only alternative equilibrium possibility has $\alpha = 0$ and $\beta > 0$, but this requires $\pi_A a < \pi_B b$. Hence the equilibrium is unique.

Proposition 2 (Cheap Talk with Lying Aversion) *For incentives and values satisfying (5), the unique equilibrium with lying aversion has acceptance and lying rates given by (3) and (7).*

Lying aversion has the qualitative effect of selecting the simple type of cheap talk equilibrium analyzed in Proposition 1, and the quantitative effect of reducing the amount of lying. The strength of lying aversion can be represented by stochastic dominance of different G distributions. As lying aversion increases, the marginal type $d_A(\alpha)$ that satisfies the equilibrium condition (6) becomes larger, lying decreases, and P_A increases accordingly. For instance, if G is uniform on $[0, \overline{d}]$, then $d_A(\alpha) = \alpha\overline{d}$, so the equilibrium lying rate α is decreasing in $\overline{d}$ from (7). Conversely, if lying aversion is very weak, e.g., $\overline{d}$ is sufficiently small, then lying and acceptance rates are arbitrarily close to the pure cheap talk case. We will take G as given and test comparative static predictions for changes in the expert's incentives and the decision maker's values.

3 Testable Implications

We now derive testable comparative static predictions for how expert incentives and decision maker values affect behavior. We use Proposition 1 to provide intuitive proofs for the special case of pure cheap talk, and in the Appendix we use Proposition 2 to show that the same testable implications extend to the general case that allows for lying aversion. Because individuals vary in the strength of lying aversion, the model does not make point predictions but only comparative static predictions. In this section, we also extend the model to allow for uncertainty over sender preferences, and state the results in Proposition 3.

3.1 Symmetric Baseline

To provide a baseline for testing the effects of asymmetries, suppose $\pi_A = \pi_B = \pi$ and $a = b = v$ for some $\pi > 0$ and $v \in (0, 1]$. From (4) there is a cheap talk equilibrium without lying, $\alpha = \beta = 0$, so the recommended action is definitely better. In this equilibrium, the expected value of either recommended action is v, so the decision maker follows the recommendation if the value of the outside option v_C is lower than v. Therefore, the decision maker takes the recommended action with probability $\Pr[v_C \leq v] = v$. Without a recommendation, the expected value of either action is $v/2$, so the decision maker will take one of these actions rather than the outside option with probability $\Pr[v_C \leq v/2] = v/2$. Thus, a cheap talk recommendation doubles the probability that one of the two actions is accepted and thereby doubles the expert's expected payoff. As shown in the Appendix, if payoffs and/or values are higher for one action than the other, communication is still persuasive as long as (5) is satisfied.[10]

In the Symmetric Baseline (SB) treatment of the experiment, we set $\pi_A = \pi_B = .8$ and $a = b = 1$, so the expert receives .8 if either A or B is chosen regardless of which is actually better, while the decision maker receives 1 from choosing the better action and 0 from choosing the worse action. In this symmetric case, there is a one-to-one relation between the acceptance probability and the expert's payoffs, so we test whether the probability that the expert's recommendation for A or B is accepted after communication, P^{SB}, is higher than the theoretical probability that A or B is chosen without communication.[11]

Persuasiveness Hypothesis: *Communication increases the probability that the decision maker chooses A or B rather than the outside option, $P^{SB} > 1/2$, thereby benefiting the expert.*

To see how communication is persuasive in this case and more generally, Figure 1(a) shows the expert's preferences over decision maker estimates of v_A and v_B in $[0,1]^2$. The figure can capture any prior distribution of (v_A, v_B), and shows the case where the prior distribution is evenly distributed on the points $(1,0)$ and $(0,1)$ as in the SB treatment.[12] Given any such points, the decision maker's estimates must be a convex combination of them, and hence on a line between them. Above the dashed line, $E[v_B] > E[v_A]$, so the decision maker chooses B with probability $P_B = E[v_B]$ and the expert earns an expected payoff of $\pi_B E[v_B]$, as represented by the expert's isopayoff curves. Similarly, below the dashed line, the expert earns $\pi_A E[v_A]$.

If the decision maker believes that a recommendation for A or B is always honest, the posterior estimates are $E[(v_A, v_B)\,|m_A] = (1,0)$ and $E[(v_A, v_B)\,|m_B] = (0,1)$. Since $\pi_A = \pi_B$ in this treatment, these estimates for each message are on the same isopayoff curve so there is no incentive to lie and hence such beliefs are part of an equilibrium, as confirmed by equation (4) where $\alpha = \beta = 0$. Moreover this isopayoff curve is higher than that through the prior $E[(v_A, v_B)] = (.5, .5)$, so the expert benefits from either recommendation relative to no communication.

[10] In a logit discrete choice model, Chung (2012) finds that expert preferences are always quasiconvex or quasiconcave, and switch to the latter for sufficient asymmetry, implying that any communication makes the expert worse off (Chakraborty and Harbaugh, 2010).

[11] If decision makers are risk-averse, or are competitive and want to hurt the expert, the theoretical acceptance probability will be lower than 1/2, which biases this test against supporting our hypothesis. However, the opposite holds if they are risk-seeking or altruistic.

[12] See Figure 2(a) in Chakraborty and Harbaugh (2010) for the corresponding case where the distribution of v has full support on $[0,1]^2$.

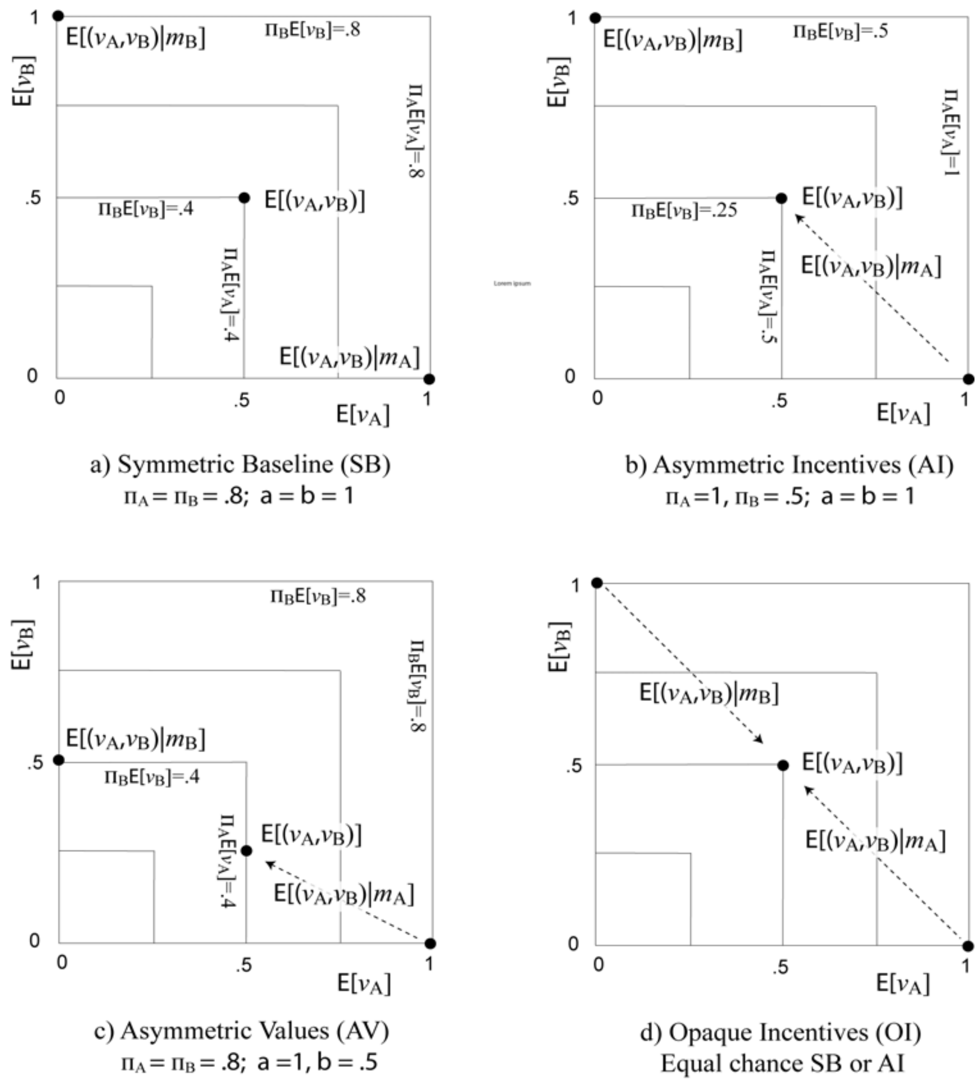


Figure 1: Incentives, Values, and Equilibrium Estimates

3.2 Asymmetric Incentives

When the expert has a financial incentive to recommend one choice, the decision maker should be suspicious of and hence less likely to follow a recommendation for that choice. Such discounting then reduces the gains from lying, so in equilibrium only a fraction of experts lie. We are interested in how this equilibrium interaction between discounting by the decision maker and the probability of lying by the expert changes as incentives change.

Since A is favored overall, $\pi_A a \geq \pi_B b$, an increase in π_A raises the asymmetry in favor of A and hence raises the probability α of lying in favor of A from (4), which implies a lower probability $P_A = a/(1+\alpha)$ that a recommendation for A is accepted. Similarly, a decrease in π_B leads to an increase in α and a fall in P_A. In each case the probability of a lie in favor of B remains at $\beta = 0$, and the probability of accepting a B recommendation remains at $P_B = b$.

In our experiment, we test this discounting hypothesis in the Asymmetric Incentives (AI) treatment by setting $\pi_A = 1$ and $\pi_B = .5$ while keeping $a = b = 1$ as in the SB treatment. With this increase in π_A and decrease in π_B relative to the SB treatment, we predict more lying toward A than B in the AI treatment itself, $\alpha^{AI} > \beta^{AI}$, and more lying toward A than in the SB treatment, $\alpha^{AI} > \alpha^{SB}$. The chance that the expert follows the expert's A recommendation, P_A^{AI}, is predicted to fall accordingly.

Discounting Hypothesis: *A higher expert incentive for A or lower expert incentive for B increases the probability of a lie that A is better,* $\alpha^{AI} > \beta^{AI}$ *and* $\alpha^{AI} > \alpha^{SB}$*, resulting in a lower probability that a recommendation for A is accepted,* $P_A^{AI} < P_B^{AI}$ *and* $P_A^{AI} < P_A^{SB}$.

Figure 1(b) shows this AI treatment, in which the greater incentive to recommend A implies that the isopayoff curves are discontinuous at the dashed line. The expert has an incentive to lie in favor of A if $E[v_A|m_A] = E[v_B|m_B]$, but as the lying rate α increases $E[v_A|m_A] = 1/(1+\alpha)$ falls, and $E[(v_A, v_B)\,|m_A]$ moves toward the prior $(.5, .5)$. If $\pi_A \leq 2\pi_B$, then for some $\alpha \in [0,1]$ the estimates reach the same isopayoff curve as that of $E[(v_A, v_B)\,|m_B]$, so the incentive to lie is balanced out by the skepticism of the decision maker. Since $\pi_A = 2\pi_B$ in the treatment, $E[v_A|m_A]$ would fall all the way to the prior without lying aversion, but with lying aversion the payoff for falsely claiming A is better is reduced by d_A, so in equilibrium $E[v_A|m_A]$ decreases less and the posterior $E[(v_A, v_B)\,|m_A]$ does not reach the prior $E[(v_A, v_B)]$. The comparative static predictions are unaffected by the strength of lying aversion.

3.3 Asymmetric Values

Che, Dessein and Kartik (2013) show how an expert often has an incentive to "pander" by recommending an action that the decision maker is already leaning toward, with a resulting loss in credibility that hurts both sides. To capture a simplified version of pandering, we consider the case where the incentives are symmetric but the distribution is asymmetric with $a > b$, e.g., a consumer is known to favor the design of product A.[13]

Pushing A can benefit the expert more, but is more suspicious to the decision maker, which reduces the expert's incentive to push A. As seen from (4), an increase in a or a decrease in b strictly increases α and has no effect on $\beta = 0$. From (3), an increase in a has a positive direct effect on P_A that is partially counteracted by the lower credibility from the rise in α,[14] but no

[13] Che, Dessein and Kartik (2013) analyze a strong version of pandering where the expert strictly rather than weakly prefers that the decision maker choose the better of the two actions.

[14] For the limiting pure cheap talk case in the sufficiently symmetric parameter range, substituting in (4) yields $P_A = b\frac{\pi_B}{\pi_A}$ and the effects cancel out.

effect on P_B. For a decrease in b, the effect on P_A via α is negative and the direct effect on P_B is also negative, so both acceptance rates fall.

In the Asymmetric Values (AV) treatment, we set $a = 1$ and $b = .5$ and keep $\pi_A = \pi_B = .8$ as in the SB treatment. The decrease in b relative to the SB treatment should induce more lying toward A and a fall in acceptance rates.

Pandering Hypothesis: *A higher relative decision-maker value for A increases the probability of a lie that A is better,* $\alpha^{AV} > \beta^{AV}$ *and* $\alpha^{AV} > \alpha^{SB}$, *and a lower-decision maker value for B decreases the probability that a recommendation for either action is accepted,* $P_A^{AV} < P_A^{SB}$ *and* $P_B^{AV} < P_B^{SB}$.

Figure 1(c) shows this AV treatment, in which the prior is evenly distributed on $(1, 0)$ and $(0, .5)$. If the decision maker took recommendations literally so $E[(v_A, v_B)\,|m_A] = (1, 0)$ and $E[(v_A, v_B)\,|m_B] = (0, .5)$, the expert would always lie toward A as seen from the isopayoff curves. As the lying rate α increases, the decision maker discounts the recommendation and $E[v_A|m_A] = 1/(1+\alpha)$ falls as $E[(v_A, v_B)\,|m_A]$ approaches the prior. If the asymmetry in values is not too great, $a \leq 2b$, this discounting reduces the incentive to lie sufficiently for $E[(v_A, v_B)\,|m_A]$ to rest on the same isopayoff curve as $E[(v_A, v_B)\,|m_B] = b$, at which point the incentive to lie is eliminated. Since $a = 2b$ in the treatment, in a pure cheap talk equilibrium an A recommendation is fully discounted, $E[(v_A, v_B)\,|m_A] = E[(v_A, v_B)]$. Any lying aversion induces an equilibrium with less discounting, but with the same comparative static predictions.[15]

3.4 Extension – Opaque Incentives

Often a decision maker is uncertain whether the expert is biased or not, such as a customer who does not know whether a salesperson benefits more from selling one brand or another. To allows for this we now extend the model to assume the expert is biased toward A with probability $\gamma \in (0, 1)$ so $\pi_A > \pi_B$, and is otherwise unbiased so $\pi_A = \pi_B = \pi_u$ for some $\pi_u > 0$.[16] The decision maker knows the probability γ that the expert might be biased toward A. To focus on the incentive issue and simplify the analysis, suppose $a = b = v$ for some $v \in (0, 1]$.

[15] We do not test the interesting case of offsetting incentive and value asymmetries. From (4) and as seen in Figures 1(b) and 1(c), the incentive to lie is eliminated if $a = 2b$ and $2\pi_A = \pi_B$. Such countervailing incentives with the strong version of pandering are analyzed by Che, Dessein, and Kartik (2013, Online Appendix) and Chiba and Leong (2015). We also do not consider the case where the expert is uncertain over the decision maker's values, which directly mitigates the loss from pandering in this environment, and can allow for communication in other environments even with extreme biases (Chakraborty and Harbaugh, 2014).

[16] The closest analysis of incentive uncertainty is in Chakraborty and Harbaugh (2010, Online Appendix). We focus on the case where one expert type is unbiased, but the same analysis applies if both types are biased toward one of the actions with different strengths or they are each biased toward different actions. There may also be a range of possible expert biases, which can further undermine communication (Diehl and Kuzmics, 2018).

For the expert type that is biased toward A, recommending A is less suspicious than it would be if the bias were known, so the incentive to lie is strengthened. For the unbiased expert who has equal incentives to recommend A or B, the decision maker's suspicion of an A recommendation creates an incentive to recommend B even when A is actually better. When uncertainty is at the maximum, $\gamma = 1/2$, such lying by both types leaves the decision maker with no information unless there is some lying aversion. Letting subscripts b and u denote lying rates by biased and unbiased types, respectively, we have the following result. The proof is in the Appendix.

Proposition 3 (Opaque Incentives) *For symmetric values, (i) if $\gamma < 1/2$, there exists an informative pure cheap talk equilibrium with $\alpha_b = 1$, $\beta_b = 0$, $\alpha_u = 0$, $\beta_u \in (0,1)$; (ii) if $\gamma = 1/2$, there exists an uninformative pure cheap talk equilibrium with $\alpha_b = 1$, $\beta_b = 0$, $\alpha_u = 0$, $\beta_u = 1$, and an informative pure cheap talk equilibrium does not exist; (iii) if $\gamma > 1/2$ and π_A/π_B is sufficiently close to 1, there exists an informative pure cheap talk equilibrium with $\alpha_b \in (0,1)$, $\beta_b = 0$, $\alpha_u = 0$, $\beta_u = 1$; and (iv) in any equilibrium with lying costs, $\alpha_b > 0$, $\beta_b = 0$, $\alpha_u = 0$, $\beta_u > 0$.*

The Opaque Incentives (OI) treatment tests the maximum uncertainty case of $\gamma = 1/2$ where each expert is equally likely to be assigned the unbiased incentives in the SB treatment or the biased incentives in the AI treatment. We predict more lying by biased types toward A than in the AI treatment, and expect lying by unbiased types toward B where none was predicted in the AI and SB treatments. Regarding acceptance rates, since we predict lying in favor of both A and B in the OI treatment, and no lying in the SB treatment, lack of transparency is predicted to reduce acceptance rates for A and B recommendations relative to the SB treatment. Since there is lying in favor of A but not B in the AI treatment, the acceptance rate for B is expected to be lower than in the AI treatment. The effect on the acceptance rate for A relative to the AI treatment is ambiguous since the credibility of an A recommendation is enhanced by the presence of unbiased experts who do not lie in favor of A, but undermined by biased experts' stronger incentive to lie in favor of A.

Noting that the main insight of this model is similar to that of the political correctness theory of Morris (2001), we have the following hypothesis, which is proven in the Appendix.

Political Correctness Hypothesis: *For symmetric values, if $\gamma = 1/2$ so the expert is equally likely to be biased toward A or unbiased, then (i) biased and unbiased experts lie in opposite directions, $\alpha_b^{OI} > \beta_b^{OI}$ and $\alpha_u^{OI} < \beta_u^{OI}$, and are more likely to lie than if the expert is known to be biased or known to be unbiased, $\alpha_b^{OI} > \alpha^{AI}$,$\alpha^{SB}$ and $\beta_u^{OI} > \beta^{AI}$,$\beta^{SB}$, and (ii) the probability that a recommendation for either A or B is accepted is lower than if the expert is known to be unbiased, $P_A^{OI} < P_A^{SB}$ and $P_B^{OI} < P_B^{SB}$, and the probability that a recommendation for B is accepted is lower than if the expert is known to be biased, $P_B^{OI} < P_B^{AI}$.*

Figure 1(d) shows the OI treatment for this most unfavorable case where $\gamma = 1/2$. To see how communication with pure cheap talk unravels as lying encourages more lying, suppose a biased expert type always lies toward A, and an unbiased type never lies. An A recommendation is then a weighted average of $(v_A, v_B) = (1, 0)$ from the unbiased types who recommend A half the time, and $(v_A, v_B) = (1/2, 1/2)$ from the biased types who always recommend A, so $E[(v_A, v_B) \,|m_A] = (2/3, 1/3)$. Since a B recommendation is always from an unbiased type, $E[(v_A, v_B) \,|m_B] = (0, 1)$. Hence $P_B = E[v_B|m_B] = 1 > 2/3 = E[v_A|m_A] = P_A$, so unbiased types have an incentive to lie toward B. If they lie half the time toward B then the recommendation is discounted to $E[(v_A, v_B) \,|m_B] = (1/3, 2/3)$, which would set $P_A = P_B$. However $P_A = E[v_A|m_A]$ now falls further since only a quarter of A recommendations come from unbiased rather than biased types, which gives even more incentive for unbiased types to lie toward B, etc.[17] Only when both types always lie and $E[(v_A, v_B) \,|m_i]$ reaches the prior is the unbiased type made indifferent; at that point, the decision maker learns nothing. Allowing for some lying aversion reduces equilibrium lying in both directions, but the comparative static predictions for lying rates and acceptance probabilities are the same.

4 Experiment

4.1 Experimental Design

Following the exact model introduced in Section 2, we conduct an experiment using the parameter values for the different treatments in Section 3. In each treatment, a "consultant" (expert) has two projects, A and B, one of which is better for the "client" (decision maker). The probability that either project is the better one is 1/2 independently in each round. The client has an alternative project C with a value that is drawn from a uniform distribution independently in each round. The consultant knows the realized values of projects A and B and the client knows the realized value of the alternative project C. After learning the realized values of projects A and B, the consultant sends one of two messages to the client: either "I recommend Project A" or "I recommend Project B". The client learns the realized value of the alternative project C and sees the client's message. The client then chooses project A, B, or C and the decision and the outcome are displayed to both players.

We conduct four sessions with 20 subjects each, all drawn from undergraduate classes at Indiana University's Kelley School of Business in 2011. We follow a within-subject design in

[17] With lying costs, more lying by unbiased types also encourages more lying by biased types since P_B falls faster than P_A. As shown in the proof of the political correctness hypothesis, lying by each type is a strategic complement to lying by the other type.

which the same subjects are exposed to all four treatments in the sequence SB-AI-OI-AV.[18] This sequence minimizes the potential for confounding treatment effects with learning effects or experimenter-demand effects, since from each treatment to the next subjects do not have an incentive to either just follow their previous behavior or just switch to the opposite behavior. In the first treatment (SB), experts have no incentive to lie; in the second treatment (AI), experts have an incentive to lie if B is better; in the third treatment (OI), an expert is equally likely to be biased, with an incentive to lie if B is better, or unbiased, with an incentive to lie if A is better; in the final treatment (AV), all experts have an incentive to lie if A is better. However, learning effects may still confound some predictions as discussed below.

The experiments are conducted on computer terminals at the university's Interdisciplinary Experimental Lab using z-Tree software (Fischbacher, 2007). In each session, the subjects are randomly assigned to play the role of one of 10 consultants or 10 clients. Each of the four treatments (or "series" as they are described to subjects) in a session lasts 10 rounds; over these rounds, every client is matched anonymously with every consultant once and only once. At the end of the experiment, one round from each treatment is randomly chosen as the round that the subjects are paid for to reduce earnings effects. Subjects privately learn which rounds were chosen and privately receive cash payments. No record of their actions mapped to their identity is maintained. The monetary payoffs are 10 times larger in US dollar terms than indicated in Figure 1, but there is a 1 in 10 chance of each round being chosen, so the expected values for each round are the same. The average earnings in the 90-minute experiment are, excluding a $5 show-up payment, $22.02 for consultants and $27.78 for clients. The detailed procedures and verbal instructions are in Appendix B. Sample screenshots are in Appendix C.

4.2 Results and Analysis

In each of the four identical sessions there are four treatments, each with ten rounds. Since there are 20 subjects in a session, each round has 10 pairs of recommendations and decisions. To allow for some learning, we base the statistical analysis on behavior in the last 5 rounds of each treatment.[19] In these rounds there are 50 recommendations and 50 decisions in each session for each treatment, but these data points are not independent since players make multiple decisions

[18] Given the complexity of the OI treatment, an alternative between-subject design would require either a large number of rounds, in which case reputation effects would be hard to eliminate, or some practice rounds with and without expert bias for the subjects to better understand the game. The within-subject design keeps track of behavior from such "practice rounds" in the simpler SB and AI treatments and allows it to be compared with behavior in the more complicated OI and AV treatments.

[19] The significance results are qualitatively the same if we base the tests on behavior frequencies over all 10 rounds, with one exception, which is discussed in Footnote 23. Due to a coding error, data for the first round of the Asymmetric Values Treatment was not recorded, so only 9 rounds of data are available for this treatment.

and learn from interactions with other players. To allow for such interdependence, our statistical analysis follows the most conservative approach of analyzing behavior at the session level rather than the subject or subject-round level. In particular, we treat the frequency of a behavior in the last five rounds of a treatment for each session as the unit of analysis, and use paired tests to reflect the within-subject design. Since there are only four independent data points for a treatment, one for each session, statistical significance requires consistent differences in behavior frequencies across treatments within each of the four sessions. These frequencies are reported in Table 1. We do not formally model learning; however, for comparison, Figures 2 and 3 indicate frequencies for the first five rounds using dashed gray lines and frequencies for the last five rounds using solid gray lines. We limit our discussion of learning to cases involving large or unexpected differences between the first five and last five rounds.

For the SB treatment, the incentives and payoffs are symmetric, so the Persuasiveness Hypothesis implies that decision makers are less likely to choose C than if there were no communication. Since we treat the data generating process as being at the session level for each treatment, we are interested in whether the overall rate of acceptance for A and B is significantly above 1/2. Consistent with the prediction, from Table 1 we see that this frequency P^{SB}, which is a weighted average of P_A^{SB} and P_B^{SB}, is above 1/2 in each session.[20] For any possible distribution, the probability that this would occur if the hypothesis were false is lower than $(1/2)^4 = 1/16$, which is the p-value for the one-sided Wilcoxon signed rank test, as seen in Table 2. The one-sided t-test (which, in contrast to the subsequent tests, is a one sample test and thus not paired) indicates the difference is significant at the 5% confidence level. We report t-test results for completeness, though clearly the normality assumption is not valid in our context.

Figure 2(a) shows the overall acceptance rates, and Figure 3(a) shows that a decision maker is less likely to follow an expert's recommendation as the outside option becomes more attractive. For comparison with the other treatments, the "Expert Lies" bar in Figure 2(a) shows the frequency at which experts recommend A when B is the better option (α^{SB}) and the frequency at which experts recommends B when A is the better option (β^{SB}).

[20] Lower acceptance rates in the first two sessions may reflect the effects of one expert in each session who repeatedly lies for no material benefit, as discussed in Section 4.3. Consistent with such unexpected lying affecting subsequent decision-maker behavior, acceptance rates in the first five rounds of Sessions 1 and 2 are close to those of Sessions 3 and 4, but diverge in the last five rounds. Our use of session-level data prevents correlated error terms for subjects within a session from biasing inference.

	Var.	Ses. 1	Ses. 2	Ses. 3	Ses. 4	Overall
Symmetric Baseline Treatment						
Recommend A When B Better	α^{SB}	.14	.13	.11	.07	.11
Recommend B When A Better	β^{SB}	.14	.15	.04	.04	.10
A Recommendation Accepted	P_A^{SB}	.63	.77	.92	.96	.81
B Recommendation Accepted	P_B^{SB}	.39	.71	.88	1.00	.76
A or B Recommendation Accepted	P^{SB}	.52	.74	.90	.98	.79
Asymmetric Incentives Treatment						
Recommend A When B Better	α^{AI}	.35	.31	.52	.53	.41
Recommend B When A Better	β^{AI}	.21	.24	.04	.06	.13
A Recommendation Accepted	P_A^{AI}	.36	.32	.53	.53	.45
B Recommendation Accepted	P_B^{AI}	.68	.80	.71	1.00	.77
Opaque Incentives Treatment						
Biased: Recommend A When B Better	α_b^{OI}	.45	.45	.58	.78	.56
Biased: Recommend B When A Better	β_b^{OI}	.36	.07	.00	.31	.19
Unbiased: Recommend A When B Better	α_u^{OI}	.33	.15	.00	.00	.13
Unbiased: Recommend B When A Better	β_u^{OI}	.38	.58	.58	.50	.51
A Recommendation Accepted	P_A^{OI}	.58	.28	.68	.54	.52
B Recommendation Accepted	P_B^{OI}	.54	.64	.80	.92	.72
Asymmetric Values Treatment						
Recommend A When B Better	α^{AV}	.46	.68	.63	.70	.62
Recommend B When A Better	β^{AV}	.00	.05	.00	.09	.03
A Recommendation Accepted	P_A^{AV}	.46	.40	.68	.75	.58
B Recommendation Accepted	P_B^{AV}	.62	.30	.44	.40	.45

Table 1: Lying and Acceptance Frequencies in Rounds 6-10 of Each Treatment

In the AI treatment where action A is more incentivized, the Discounting Hypothesis states that decision makers are less likely to accept a recommendation for A since experts often lie in favor of A. Consistent with the predictions for false claim rates in this hypothesis, we see in Table 1 and Figure 2(b) that experts falsely claim A to be better 41% of the time and falsely claim B to be better 13% of the time. From Table 1, we see that false claims are higher for A than B ($\alpha^{AI} > \beta^{AI}$) in each session of the treatment, which would occur with probability $(1/2)^4 = 1/16$ in any distribution if the hypothesis were false, as seen in Table 2. We also see that false claims in favor of A are always higher than in the SB treatment ($\alpha^{AI} > \alpha^{SB}$), which again implies a p-value of 1/16. Using a paired one-sided t-test, both of these false claim differences are statistically significant at the 5% level or better. Going forward, we will refer to a difference as significant if

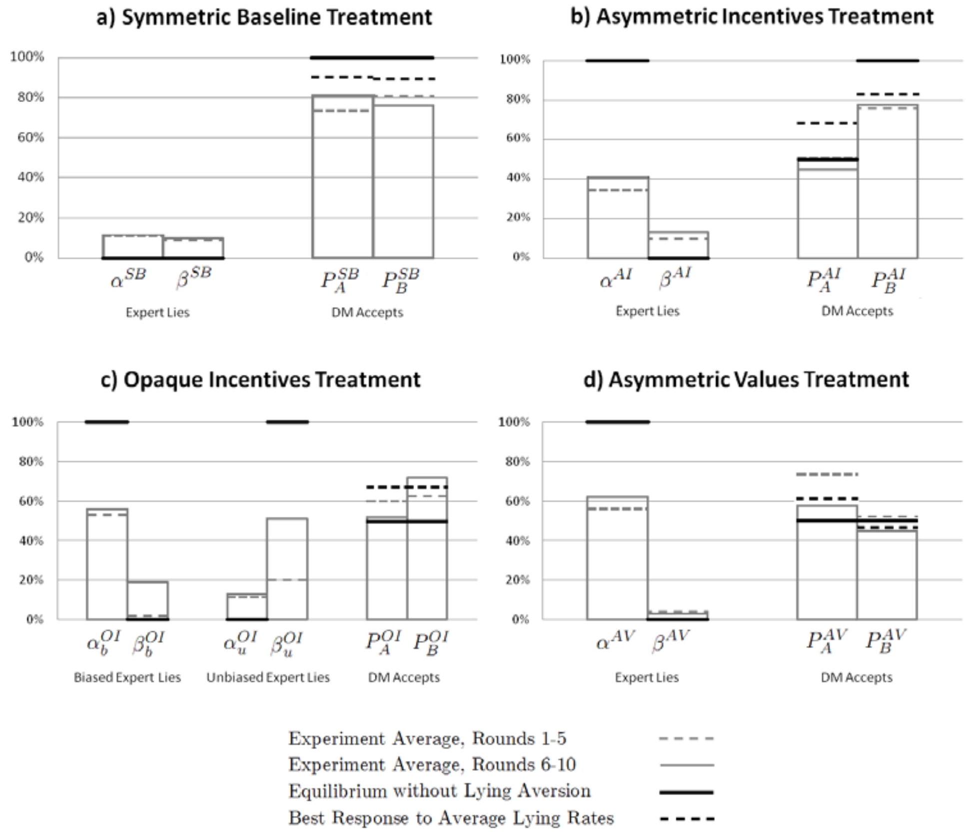


Figure 2: Lying and Acceptance Frequencies

the p-value for the Wilcoxon signed rank test reaches this lowest attainable level of $p = 1/16$ and the paired one-sided t-test also indicates significance at the 5% level or better.

Regarding acceptance rates in the AI treatment, as seen in Table 1 and Figure 2(b), decision makers accept the recommendation 45% of the time when A is recommended and 77% of the time when B is recommended. From Table 1, we see that acceptance rates in the treatment are lower for A than for B in each session ($P_A^{AI} < P_B^{AI}$). Also, acceptance rates for A are lower in the AI treatment than in the SB treatment ($P_A^{AI} < P_A^{SB}$). As seen in Table 2, both of these differences are significant. Based on Figure 3(b), decision makers are unlikely to accept A recommendations when the outside option v_C is favorable.

In the OI treatment, from the Political Correctness Hypothesis we expect that a biased expert type is more likely to falsely claim A is better and, of particular interest, an unbiased expert type is more likely to falsely claim B is better. Consistent with the prediction, the summary data in Table 1 and Figure 2(c) show that biased experts falsely claim A is better 56% of the time and falsely claim B is better 19% of the time, while unbiased experts falsely claim A is better 13% of the time and falsely claim B is better 51% of the time. Notice from the dashed gray line in Figure

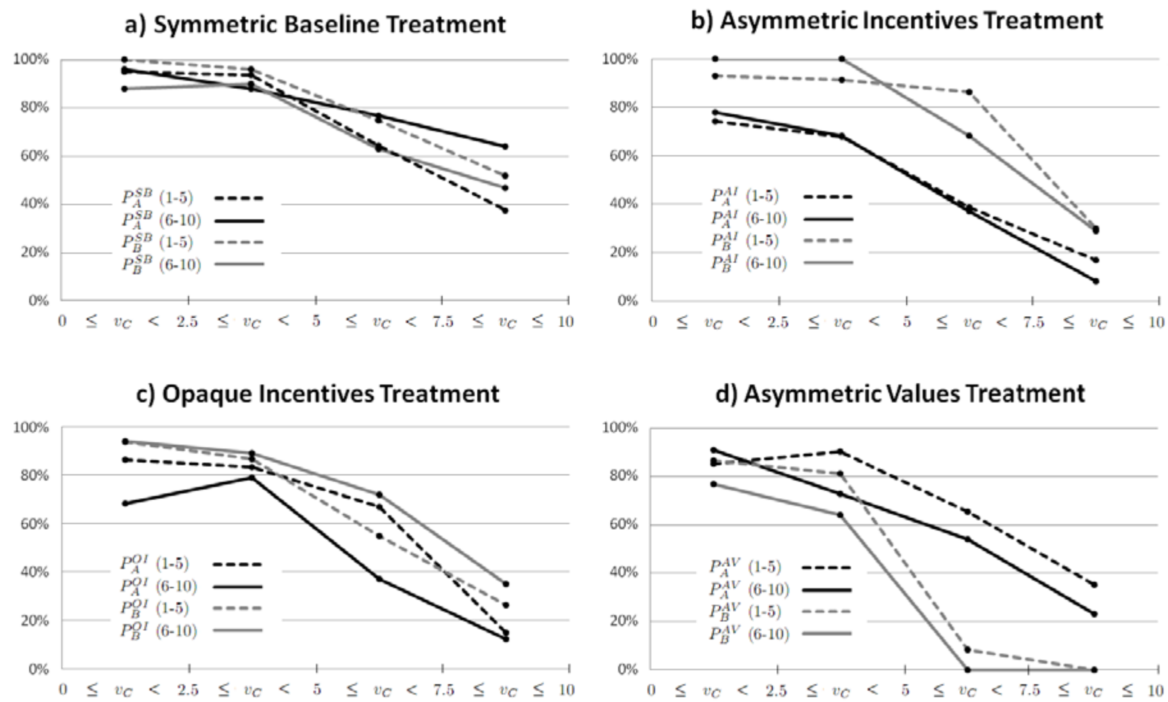


Figure 3: Acceptance Frequencies by Value of Outside Option

2(c) that unbiased experts do not immediately appear to recognize the benefits of lying toward B, but learn strongly over the course of the experiment. From Table 2, both of the false claim differences for the last five rounds are significant ($\alpha_b^{OI} > \beta_b^{OI}$ and $\alpha_u^{OI} < \beta_u^{OI}$). If we compare this behavior with that in the SB treatment, we find that in each session in the OI treatment biased experts are more likely to falsely claim A is better and unbiased experts are more likely to falsely claim B is better ($\alpha_b^{OI} > \alpha^{SB}$ and $\beta_u^{OI} > \beta^{SB}$), and both differences are significant. Comparing this behavior with that in the AI treatment, we find that biased experts are more likely to falsely claim A is better and unbiased types are more likely to falsely claim B is better ($\alpha_b^{OI} > \alpha^{AI}$ and $\beta_u^{OI} > \beta^{AI}$), and that the differences are significant. However, note that biased experts in the OI treatment may be continuing behavior that they learned in the AI treatment, so this higher rate of lying might reflect learning instead.[21]

For acceptance rates in the OI treatment, the results are more mixed. Theory predicts that decision makers should be more suspicious of an A recommendation than in the SB treatment ($P_A^{OI} < P_A^{SB}$) and the differences are significant. Theory also predicts that decision makers should

[21] Arguing for a learning interpretation, the extra incentive to lie in the Opaque Incentives treatment is weak for biased experts (compared to unbiased experts, whose political correctness behavior is the focus of this treatment). Arguing against such an interpretation, Figure 1(b-c) shows a bigger jump in lying between the treatments than between the first and second half of each treatment.

be more suspicious of a B recommendation than in either the SB or AI treatments ($P_B^{OI} < P_B^{SB}$ and $P_B^{OI} < P_B^{AI}$), though neither difference is significant. Hence, it seems that decision makers are correctly suspicious of an A recommendation, and that unbiased experts correctly anticipate this and falsely claim that B is better, but it does not appear that decision makers fully anticipate such lying by unbiased experts.[22] However, the power of our test is limited by the small sample size, so even if decision makers did fully anticipate lying toward B, it might not be significant in our test. From the gray dashed line in Figure 2(c), decision makers do not appear to become more suspicious of B recommendations over the course of the experiment, but rather are more likely to accept B recommendations in later rounds.[23] This pattern also appears in Figure 3(c), where decision makers with a favorable outside option become less rather than more suspicious of B recommendations over time. We did not use a role-reversal mechanism to speed up learning, so ten rounds might have provided insufficient opportunity for decision makers to successfully learn their optimal responses.

Finally, in the AV treatment, the Pandering Hypothesis states that experts are more likely to falsely claim that A is better than to falsely claim that B is better ($\alpha^{AV} > \beta^{AV}$). Consistent with this prediction, Table 1 and Figure 2(d) show false claim rates for A and B of 62% and 3%, respectively. The difference between these rates is statistically significant (see Table 2). Comparing the frequency of false claims for A in this treatment with that in the SB treatment (11%), the difference is also significant ($\alpha^{AV} > \alpha^{SB}$). Note that subjects who were assigned to be unbiased in the OI treatment were more likely to lie (toward A) in the subsequent AV treatment. After eventually recognizing that the indirect benefit from lying toward B in the OI treatment, these subjects may have been more likely to recognize the benefit from lying toward A in the AV treatment.[24] Regarding acceptance rates, as predicted they fall for both A and B in the AV treatment relative to the SB treatment ($P_A^{AV} < P_A^{SB}$ and $P_B^{AV} < P_B^{SB}$), but only the former decline is significant. Moreover, as seen from Figure 2(d) and Figure 3(d), decision makers take some time to become suspicious of A recommendations.[25]

[22] Such behavior may reflect level-k thinking in which subjects do not fully consider the entire chain of strategic interactions (e.g., Stahl and Wilson, 1995; Nagel, 1995; Crawford, 2003).

[23] Since behavior is initially closer to the equilibrium prediction, the difference in session means for P_B^{OI} and P_B^{SB} is significant if we include the first 5 rounds rather than restricting the analysis to the final 5 rounds.

[24] However the results have the same significance if we exclude this group. This pattern may also reflect an experimenter demand effect, whereby subjects anticipate that their behavior is expected to change between sessions and alter their behavior accordingly. However, we did not see such a reversal between the AI and OI treatments, and the OI randomization ensures that there is no predictable pattern across treatments in the direction of lying incentives.

[25] Acceptance probabilities are higher for A than B. While B is never a good choice when $v_C > 5$, A can still be a good choice for lower values of v_C if there is some chance that the expert is not lying.

	Hypothesis	Signed rank test	Paired t-test
Persuasiveness Hypothesis			
A and B Acceptance Rate: SB Treatment	$P^{SB} > 1/2$	.063	.032
Discounting Hypothesis			
A vs. B Lying Rate: AI Treatment	$\alpha^{AI} > \beta^{AI}$	.063	.037
A Lying Rate: AI vs. SB Treatment	$\alpha^{AI} > \alpha^{SB}$	.063	.010
A vs. B Acceptance Rate: AI Treatment	$P_A^{AI} < P_B^{AI}$	.063	.007
A Acceptance Rate: AI vs. SB Treatment	$P_A^{AI} < P_A^{SB}$	.063	.001
Political Correctness Hypothesis			
A vs. B Lying Rate: Biased OI Treatment	$\alpha_b^{OI} > \beta_b^{OI}$	.063	.017
A vs. B Lying Rate: Unbiased OI Treatment	$\alpha_u^{OI} < \beta_u^{OI}$	.063	.022
A Lying Rate: Biased OI vs. SB Treatment	$\alpha_b^{OI} > \alpha^{SB}$	.063	.008
B Lying Rate: Unbiased OI vs. SB Treatment	$\beta_u^{OI} > \beta^{SB}$	.063	.004
A Lying Rate: Biased OI vs. AI Treatments	$\alpha_b^{OI} > \alpha^{AI}$	.063	.018
B Lying Rate: Unbiased OI vs. AI Treatment	$\beta_u^{OI} > \beta^{AI}$	.063	.008
A Acceptance Rate: OI vs. SB Treatment	$P_A^{OI} < P_A^{SB}$	.063	.027
B Acceptance Rate: OI vs. SB Treatment	$P_B^{OI} < P_B^{SB}$	.438	.372
B Acceptance Rate: OI vs. AI Treatment	$P_B^{OI} < P_B^{AI}$	.188	.137
Pandering Hypothesis			
A vs. B Lying Rate: AV Treatment	$\alpha^{AV} > \beta^{AV}$	.063	.000
A Lying Rate: AV vs. SB Treatment	$\alpha^{AV} > \alpha^{SB}$	.063	.002
A Acceptance Rate: AV vs. SB Treatment	$P_A^{AV} < P_A^{SB}$	.063	.005
B Acceptance Rate: AV vs. SB Treatment	$P_B^{AV} < P_B^{SB}$	.125	.096

Table 2: p-values for One-Sided Hypothesis Tests, $n = 4$

The above results are for paired tests, which typically generate lower $p-$values for within-subject designs since they control for subject-specific factors.[26] However, in our case, the treatment effects are so strong relative to any such factors that in almost every case the unpaired tests have equal or lower $p-$values than the paired tests. As seen in Table 1, for all but one hypothesis about false claim rates (α_b^{OI} vs. α^{AI}), all four rates for one treatment are ranked above all four rates for the other treatment, which is the strongest possible ranking and implies $p = .014$ for the unpaired Wilcoxon (Mann-Whitney) rank order test. All of the differences in acceptance rates also remain significant with equal or lower p-values in unpaired tests, with one exception: there

[26] Since we treat each session with 20 subjects as one data point, subject-specific factors are largely averaged out, but there may also be session-specific factors such as the evolution of play due to strategic interactions.

is still no significant tendency in the OI treatment for decision makers to appropriately discount recommendations in the opposite direction of biased experts.

4.3 Expert Lying Aversion

The model assumes heterogeneous lying aversion, but as shown in Proposition 2, the comparative static predictions of the model do not depend on the exact distribution of lying costs. The experiment is not designed to estimate the distribution of lying costs since it does not include treatments that systematically vary the incentive to lie. Figure 4 shows the lying frequencies for the 40 different experts, 10 in each of the four sessions. The experts are ordered according to their frequency of lying in the predicted direction, i.e., the percentage of times that the expert lies in the direction for which there is a material incentive to lie in equilibrium. This is a weighted average of $\alpha^{AI}, \alpha_b^{OI}, \beta_u^{OI}$, and α^{AV}, where the weights depend on the actual number of opportunities to lie, which vary randomly based on whether A or B was in fact the better action. To reduce noise, data for all 10 rounds is shown.

Figure 4 shows substantial variation in how frequently experts lie in the predicted direction. Of the 40 experts, 7 lie at least 80% of the time, 25 lie between 20% and 80% of the time, and 8 lie less than 20% of the time or not at all. With pure cheap talk, the equilibrium and best response is to always lie in these cases, so the pattern appears roughly consistent with our assumption from the literature that a substantial fraction of experts are averse but not categorically opposed to lying, and that the strength of this aversion varies. Of course, some natural variation in lying rates should occur even by chance.[27]

Lying aversion in our experiment could reflect a true preference against lying, but may also be a reduced form for other considerations, such as an altruistic concern for the payoff to the decision maker (Hurkens and Kartik, 2009). Note that one expert in each of Sessions 1 and 2 lies consistently in a non-predicted direction, which hurts the decision maker without benefiting the expert materially.[28] Lying frequencies in a direction for which there is no material incentive to lie are also shown in Figure 4. These frequencies are a weighted average of $\alpha^{SB}, \beta^{SB}, \beta^{AI}, \beta_b^{OI}, \alpha_u^{OI}$, and β^{AV}. Such lying may confuse decision makers and lead to further effects on play in the session, but our statistical analysis is unaffected by such interdependence between rounds since it is based on session-level data.

An implication of heterogeneous lying aversion is that experts with lower lying costs should benefit financially in the AI, AV, and OI treatments from taking advantage of the extra credibility

[27] With homogeneous lying costs, in the mixed strategy equilibrium there would be heterogeneity in realized lying behavior, and with anonymity there could also be heterogeneity in the mixed strategy lying rate by each player.

[28] Such lying might reflect competitive or "nasty" preferences (Abbink and Sadrieh, 2009), or might just reflect expert confusion about the best strategy.

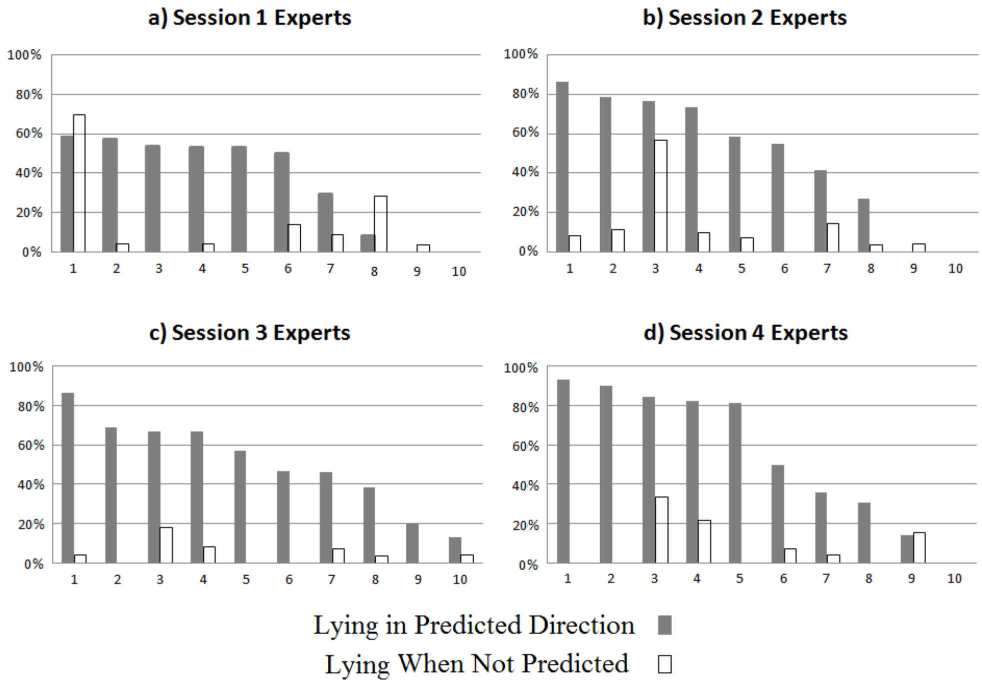


Figure 4: Overall Frequency of Lying by Each Expert

generated by experts with higher lying costs. We cannot observe lying aversion directly, but we expect a positive correlation between lying in the predicted direction and earnings. Regressing the observed lying rates against earnings, we find that over the three sessions, a 10% higher lying rate implies an extra \$0.40 in earnings for the expert.[29]

4.4 Decision Maker Best Responses

The decision maker can take a certain payoff of v_C from the outside option or accept the expert's recommendation and, depending on whether the expert lied or not, receive either zero or 10 (or 5 for action B in the AV treatment). A risk-neutral decision maker's best response acceptance rates from (3) are shown in the dark dashed lines in Figure 2 based on the empirical lying rates in each treatment. Expected monetary payoffs from deviating are small near the best response and increase in the square of the size of the deviation. Actual acceptance rates are on average 8% below these best response rates, which may reflect risk aversion. Overall, if (risk-neutral) decision makers best responded they would have made \$30.06 on average (net of the show-up payment)

[29] This effect is significant at the 1% level, though the conditions for OLS are not satisfied in our environment. We exclude rounds in which an expert lied in the opposite direction from that predicted. Similar results hold if we exclude any experts who lied in the wrong direction from the predicted direction at least once. To reduce noise, these calculations are for the average earnings over all rounds in a session rather than for the randomly selected payment round only.

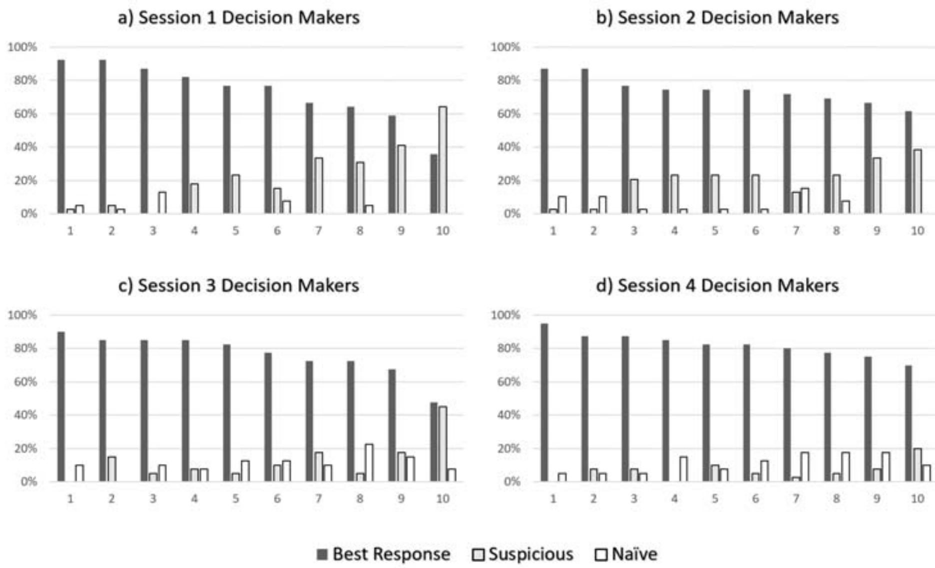


Figure 5: Overall Frequency of Behavior by Each Decision Maker

but instead made \$27.42 on average.[30] For comparison, if they had always ignored the experts and chosen C, decision makers would have made \$20 on average.[31] If the experts never lied and the decision makers always selected the best response to such honesty, decision makers would have made \$37.50 on average.[32]

Cases where acceptance rates diverge the most from the general pattern may reflect failure by decision makers to properly account for expert behavior. First, in the AI treatment, the lying rates of 41% for A and 13% for B imply from (3) that the expected value of A when it is recommended is $(1-.13)/(1-.13+.41)=.68$, implying a best response acceptance rate of 68% for A. In contrast, the actual rate is only 45%. Such strong discounting of A recommendations suggests that decision makers underestimate how much communication is possible from a biased expert. Second, in the OI treatment, the combined lying rates for biased and unbiased experts of 33% for A and 35% for B imply from (3) that the expected value of B when it is recommended is $(1-.35)/(1-.35+.33)=.66$, so the best response acceptance rate for B is 66%. However, the actual acceptance rate is 72%. Insufficient discounting of B recommendations, especially when

[30] These calculations are based on the last five rounds of each treatment, in which the decision makers use the lying rates for their own sessions to calculate P_A and P_B. Actual payoffs, which allowed for an equal chance of any round (not just the last five) being chosen in each session as the payoff round, were \$27.92.

[31] In the pure cheap talk equilibrium with honest recommendations in the SB treatment and lying in the other treatments, decision makers would have made 10+7.50+7.50+6.75=\$31.75 and experts would have made (1/2)8+(1/2)7.50 + (1/2)7.75 + (1/2)8.00=\$15.63.

[32] In this case, experts would have received $8+(10+5)/2+(10+5+8+8)/4+(8+8/2)/2=$ \$29. 25 on average rather than \$22.02, and would have suffered no utility loss from lying. Hence, as is common in cheap talk games, lack of trust hurts both the sender and receiver.

risk aversion appears to induce excess discounting in every other situation, is consistent with the inference from Table 1 and from the learning patterns in Figures 2(c) and 3(c) that some decision makers may not recognize how opaque incentives give even unbiased experts an incentive to lie. However, as discussed above, more data would be required to draw any firm conclusions.

Figure 5 orders the 10 decision makers in each session according to their best response rates over all four treatments, using data for all 10 rounds to reduce noise. Ignoring a recommendation even when it offers the highest expected value among the three choices is categorized as "suspicious" (or risk-averse), and following a recommendation even when it does not offer the highest expected value is categorized as "naive" (or risk-loving). The observed variation in behavior might reflect actual preference differences or different experiences as the participants figure out the game and learn about unobservables such as the distribution of expert lying aversion. To see whether such learning is a factor, we look at all rounds where a decision maker accepted a recommendation and compare their behavior in the next round after being lied to or told the truth in the previous round. We find that after being lied to, the "naive" action frequency decreases to 10% from 6%, and the "suspicious" action frequency increases from 8% to 15%.[33]

5 Conclusion

This paper combines insights from the literature on cheap talk recommendations into a simple, easily testable model. These insights were originally developed under varying assumptions, but they are sufficiently general that their main implications continue to hold. The model shows how recommendations can be persuasive even when the expert always wants the decision maker to take an action rather than no action, how limited asymmetries in incentives and values can distort communication but need not preclude it, and how lack of transparency about the expert's incentives can lead even an expert with unbiased incentives to offer a biased recommendation. When experts are lying averse, there is a unique equilibrium with testable comparative static predictions that do not depend on the exact distribution of lying costs.

In the first experimental tests of these predictions from the literature, we find that for every hypothesis regarding expert behavior we can reject the null hypothesis of no change in the hypothesized direction. The false claim rates by experts change as predicted overall and in every session. Of particular interest is that when incentives are opaque we find that biased experts are more likely to lie, and that unbiased experts appear to recognize that they are more persuasive if they sometimes lie in order to avoid the recommendation favored by a biased expert. For the hypotheses regarding decision maker behavior, the acceptance rates also change as predicted and

[33] However these effects are quite noisy due in part to variation in v_C and there is no significant pattern across session averages.

the changes are also statistically significant except for the Opaque Incentives treatment. Within the limited sample in the experiment, decision makers do not consistently discount the recommendation that is the opposite of that favored by the biased expert. Whether this represents a general failure of decision makers to understand how lack of transparency warps the incentives of even unbiased experts, or whether it is just a small sample artifact of our experiment, is an open question.

Since biased incentives make communication less reliable, and lack of transparency further undermines communication, these results provide theoretical and empirical support for policy measures that try to eliminate biases or at least make them more transparent. However, there are two important caveats. First, as shown by Inderst and Ottaviani (2012), disclosure requirements can lead to endogenous changes in incentives that affect the equilibrium of the subsequent communication game, and there may also be other endogenous changes in the payoff or information structure in response to disclosure requirements or other measures. Second, as seen from the theoretical results by Che, Dessein and Kartik (2013) that our experiment supports, an expert will often inefficiently pander to the decision maker even when the problems of biased incentives and lack of transparency are both solved.

6 Appendix A – Proofs

Proof of Persuasiveness Hypothesis: Since (5) holds by assumption the equilibrium is given by (3) and (7). Half the time A is better, in which case there is no lying, and half the time B is better, in which case there is lying toward A with probability α, so the probability that A is taken is $\frac{1}{2}P_A + \frac{1}{2}\alpha P_A = a/2$ and the probability that B is taken is $\frac{1}{2}(1-\alpha)P_B = \frac{1}{2}(1-\alpha)b$. Hence the overall probability that A or B is taken is $P = (a + b - \alpha b)/2$. From (7), α is strictly smaller with lying costs than with pure cheap talk, so substituting from (4),

$$P > \frac{a + b - (\pi_A a/\pi_B b - 1)b}{2}. \tag{8}$$

Note that $(a + b - (\pi_A a/\pi_B b - 1)b)/2 - a/2 = (2\pi_B b - \pi_A a)/(2\pi_B) \geq 0$ where the inequality follows from the recommendation condition in (5), so $P > a/2$. Also note that $(a+b-(\pi_A a/\pi_B b-1)b)/2 - b/2 = (\pi_B a - \pi_A a + \pi_B b)/(2\pi_B) \geq 0$ where the inequality follows from the acceptance condition in (5), so $P > b/2$.

Without communication, the decision maker will either take the outside option C or take the action A or B offering the highest expected payoff where $E[v_A] = a/2$ and $E[v_B] = b/2$. So without communication $P = \max\{a/2, b/2\}$, which is lower than with communication.

Regarding expert payoffs, with communication from (7) if B is better then expert types with a lying cost lower than d_A lie and receive a higher payoff than $\pi_B b$, while expert types with a lying cost above d_A tell the truth and receive $\pi_B b$. If A is better the expert receives $\pi_A a$. So, since $\pi_A a > \pi_B b$ by assumption, the lowest payoff for any expert type is $\pi_B b$. Without communication the expert receives $\max\{\pi_A a/2, \pi_B b/2\} = \pi_A a/2$, so communication benefits every expert type weakly and some types strictly if $\pi_B b \geq \pi_A a/2$, as required by (5). ∎

Proof of Discounting Hypothesis: Totally differentiating (7), $d\alpha/d\pi_A = a/(\pi_B b + d_A + \alpha d'_A) > 0$ and $d\alpha/d\pi_B = -\alpha b/(\pi_B b + d_A + \alpha d'_A) < 0$ where the derivative d'_A exists and is strictly positive by the assumption that G has full support with no mass points. So an increase in π_A or decrease in π_B implies there is strictly more lying and hence from (3) that P_A is strictly lower. ∎

Proof of Pandering Hypothesis: Totally differentiating (7), $d\alpha/da = \pi_A/(\pi_B b + d_A + \alpha d'_A) > 0$ and $d\alpha/db = -\alpha\pi_B/(\pi_B b + d_A + \alpha d'_A) < 0$ so an increase in a or decrease in b implies there is strictly more lying toward A. For an increase in a this implies from (3) that, in addition to the direct positive effect on P_A, there is an indirect negative effect from the rise in α. The direct effect must strictly dominate since from (6) α only increases if P_A increases. For a decrease in b there is the strictly negative direct effect on P_B, and a strictly negative indirect effect on P_A from the rise in α. ∎

Proof of Proposition 3 (Opaque Incentives): Using the assumption that $a = b = v > 0$, the expected value of the recommended actions are

$$\begin{aligned} E[v_A|m_A] &= \Pr[v_A > v_B|m_A]v = \frac{\gamma(1-\beta_b)+(1-\gamma)(1-\beta_u)}{\gamma(1-\beta_b)+(1-\gamma)(1-\beta_u)+\gamma\alpha_b+(1-\gamma)\alpha_u}v \quad (9) \\ E[v_B|m_B] &= \Pr[v_B > v_A|m_B]v = \frac{\gamma(1-\alpha_b)+(1-\gamma)(1-\alpha_u)}{\gamma(1-\alpha_b)+(1-\gamma)(1-\alpha_u)+\gamma\beta_b+(1-\gamma)\beta_u}v \quad (10) \end{aligned}$$

and the expected values of the unrecommended actions are $E[v_B|m_A] = (1 - \Pr[v_A > v_B|m_A])v$ and $E[v_A|m_B] = (1 - \Pr[v_B > v_A|m_B])v$.

(i) For $\gamma < 1/2$ suppose $\alpha_b = 1$, $\beta_b = 0$, $\alpha_u = 0$, $\beta_u < 1$, so the biased expert has a strict incentive to lie toward A, but the unbiased expert is made indifferent to lying toward B. Note that $E[v_A|m_A] > E[v_B|m_A]$ and $E[v_B|m_B] > E[v_B|m_A]$ so the acceptance probabilities are $P_A = E[v_A|m_A]$ and $P_B = E[v_B|m_B]$. So $\pi_u P_A = \pi_u P_B$ and $\pi_A P_A > \pi_B P_B$ if there exists a $\beta_u < 1$ such that $P_A = P_B$, or substituting,

$$\frac{\gamma+(1-\gamma)(1-\beta_u)}{2\gamma+(1-\gamma)(1-\beta_u)} = \frac{1}{1+\beta_u}. \quad (11)$$

Solving, $\beta_u = \gamma/(1-\gamma)$ so the equilibrium exists.

(ii) In the candidate uninformative equilibrium $\alpha_u = 0, \beta_u = 1$ and $\alpha_b = 1, \beta_b = 0$, implying $E[v_A|m_A] = E[v_B|m_A]$ and $E[v_A|m_B] = E[v_B|m_B]$. For $v_C < v$ the decision maker is indifferent between following the recommendation or taking the other action, so suppose they follow it. The unbiased type is indifferent between messages so loses nothing from always reporting m_B, while the biased type strictly prefers sending m_A. So the equilibrium exists.

Informative cheap talk requires $\Pr[m_A], \Pr[m_B] > 0$ and $E[(v_A, v_B)|m_A] \neq E[(v_A, v_B)|m_B]$. By the law of iterated expectations, it cannot be that $E[v_A|m_i] < E[v_B|m_i]$ or $E[v_B|m_i] < E[v_A|m_i]$ for all i, or that $E[v_A|m_i] \neq E[v_B|m_i]$ and $E[v_A|m_j] = E[v_B|m_j]$ for any i, j, which leaves two possibilities.

Suppose $E[v_A|m_A] > E[v_B|m_A]$ and $E[v_B|m_B] > E[v_A|m_B]$. First, suppose in equilibrium $\pi_u P_A = \pi_u P_B$ so $\pi_A P_A > \pi_B P_B$, which implies $\alpha_b = 1$,$\beta_b = 0$. Then from (9) and (10), $\pi_u P_A = \pi_u P_B$ requires for $\gamma = 1/2$ either $\alpha_u = 0, \beta_u = 1$ or $\alpha_u = 1, \beta_u = 0$. The former implies $E[v_A|m_A] = E[v_B|m_A]$ and $E[v_A|m_B] = E[v_B|m_B]$, a contradiction, and the latter implies only m_A is sent. Second, suppose in equilibrium $\pi_A P_A = \pi_B P_B$ so $\pi_u P_A < \pi_u P_B$, which implies $\beta_u = 1$ and $\alpha_u = 0$. Then from (9) and (10), $\pi_A P_A = \pi_B P_B$ requires for $\gamma = 1/2$ that $\alpha_b = 0$ and $\beta_b = 1$, which implies $E[v_A|m_A] = E[v_B|m_A]$ and $E[v_A|m_B] = E[v_B|m_B]$, a contradiction. Finally, suppose in equilibrium neither $\pi_u P_A = \pi_u P_B$ nor $\pi_A P_A = \pi_B P_B$. Then both types will send either the same message or opposing messages. In the former case one message is never sent and in the latter case $E[v|m_A] = E[v|m_B]$.

Suppose $E[v_A|m_A] < E[v_B|m_A]$ and $E[v_B|m_B] < E[v_A|m_B]$. Then messages imply the opposite of their literal meanings but otherwise the above analysis is the same.

(iii) For $\gamma > 1/2$ suppose $\alpha_b < 1$, $\beta_b = 0$, $\alpha_u = 0$, $\beta_u = 1$ so instead the biased expert is made indifferent, while the unbiased expert has a strict incentive to lie. Again $E[v_A|m_A] > E[v_B|m_A]$ and $E[v_B|m_B] > E[v_B|m_A]$, so the acceptance probabilities are $P_A = E[v_A|m_A]$ and $P_B = E[v_B|m_B]$. The equilibrium exists if, letting $r = \pi_A/\pi_B$, there is an $\alpha_b < 1$ such that $E[v_A|m_A] = E[v_B|m_B]$ or

$$\frac{r}{1+\alpha_b} = \frac{\gamma\,(1-\alpha_b) + (1-\gamma)}{\gamma\,(1-\alpha_b) + 2(1-\gamma)}. \tag{12}$$

For any given $\gamma \in (1/2, 1)$ consider the implicit function $\alpha_b(r)$ that solves (12). For $r = 1$, note $\alpha_b\,(1) = (1-\gamma)/\gamma$ and $d\alpha_b/dr = 1/\,(2\gamma - 1)$ so the function exists and is continuous in a neighborhood of $r = 1$. Hence for $r < 1$ sufficiently close to 1 there is an $\alpha_b \in (0, 1)$ that solves (12), so the equilibrium exists.

(iv) For each of the two actions, each of the two expert types can sometimes lie toward the action or never lie, so there are 16 cases. By elimination we want to show that only $\alpha_b > 0$, $\beta_b = 0$, $\alpha_u = 0$, $\beta_u > 0$ is possible in equilibrium.

If $\alpha_b > 0$ the biased types are lying toward A so it must be that $\pi_A P_A > \pi_B P_B$ and if $\beta_b > 0$ it must be that $\pi_A P_A < \pi_B P_B$, so any case with both $\alpha_b, \beta_b > 0$ is not possible. Similarly any case with $\alpha_u, \beta_u > 0$ is not possible.

If $\alpha_b = \alpha_u = 0$ then it must be that $\pi_A P_A \leq \pi_B P_B$ and $\pi_u P_A \leq \pi_u P_B$. From (9), $\Pr[v_A > v_B|m_A] = 1$ and $\Pr[v_B > v_A|m_B] = 1/(1+\gamma\beta_b + (1-\gamma)\beta_u) > 1/2$, so $E[v_A|m_A] > E[v_B|m_A]$ and $E[v_B|m_B] > E[v_A|m_B]$ and the acceptance probabilities are $P_A = v$ and $P_B = v/(1+\gamma\beta_b + (1-\gamma)\beta_u)$. This implies $P_A \geq P_B$, a contradiction with $\pi_A P_A \leq \pi_B P_B$ since $\pi_A > \pi_B$, so any case with $\alpha_b = \alpha_u = 0$ is not possible.

If $\beta_b = \beta_u = 0$ then it must be that $\pi_A P_A \geq \pi_B P_B$ and $\pi_u P_A \geq \pi_u P_B$. From (10), $\Pr[v_A > v_B|m_A] = 1/(1+\gamma\alpha_b + (1-\gamma)a_u)$ and $\Pr[v_B > v_A|m_B] = v$, so $E[v_A|m_A] > E[v_B|m_A]$ and $E[v_B|m_B] > E[v_A|m_B]$ and the acceptance probabilities are $P_A = v/(1+\gamma\alpha_b + (1-\gamma)a_u)$ and $P_B = v$. This implies $P_A < P_B$, a contradiction with $\pi_u P_A \geq \pi_u P_B$ since $\pi_u > 0$, so any case with $\beta_b = \beta_u = 0$ is not possible.

Therefore we are left with two cases, $\alpha_b = 0, \beta_b > 0, \alpha_u > 0, \beta_u = 0$ and $\alpha_b > 0, \beta_b = 0, \alpha_u = 0, \beta_u > 0$, both of which imply $E[v_A|m_A] > E[v_B|m_A]$ and $E[v_B|m_B] > E[v_A|m_B]$ so again $P_A = E[v_A|m_A]$ and $P_B = E[v_B|m_B]$. The former case implies that the financial benefit is higher for biased types to recommend B and for unbiased types to recommend A, $\pi_A P_A < \pi_B P_B$ and $\pi_u P_A > \pi_u P_B$, which cannot both hold for $\pi_A > \pi_B$, so only the latter case is possible. ■

Proof of Political Correctness Hypothesis: From Proposition 3(iv) we know that $\alpha_b > 0, \beta_b = 0, \alpha_u = 0, \beta_u > 0$ so the equilibrium indifference conditions are

$$\pi_A P_A - d_A = \pi_B P_B \qquad \text{for } v_B > v_A \tag{13}$$

$$\pi_u P_A = \pi_u P_B - d_B \qquad \text{for } v_A > v_B \tag{14}$$

for the biased and unbiased types respectively where $d_A = G^{-1}(\alpha_b)$ and $d_B = G^{-1}(\beta_u)$.

Regarding existence of equilibria, for any given β_u the LHS of (13) is greater than the RHS starting at $\alpha_b = 0$, so either there is an $\alpha_b \in (0,1)$ that solves (13) or $\alpha_b = 1$. Similarly for any given α_b the RHS of (14) is greater than or equal to the LHS starting at $\beta_u = 0$, so either there is a $\beta_u \in (0,1)$ that solves (13) or $\beta_u = 1$. Considering the corresponding implicit functions $\alpha_b(\beta_u)$ and $\beta_u(\alpha_b)$, from (13) and (14),

$$\frac{d\alpha_b}{d\beta_u} = -\frac{\pi_A \frac{\partial P_A}{\partial \beta_u} - \pi_B \frac{\partial P_B}{\partial \beta_u}}{\pi_A \frac{\partial P_A}{\partial \alpha_b} - \pi_B \frac{\partial P_B}{\partial \alpha_b} - \frac{\partial d_A}{\partial \alpha_b}} > 0 \tag{15}$$

$$\frac{d\beta_u}{d\alpha_b} = -\frac{\pi_u \frac{\partial P_A}{\partial \alpha_b} - \pi_u \frac{\partial P_B}{\partial \alpha_b}}{\pi_u \frac{\partial P_A}{\partial \beta_u} - \pi_u \frac{\partial P_B}{\partial \beta_u} + \frac{\partial d_B}{\partial \beta_u}} > 0 \tag{16}$$

where the inequalities follows since $\partial P_B/\partial\alpha_b < \partial P_A/\partial\alpha_b < 0$ and $\partial P_B/\partial\beta_u < \partial P_A/\partial\beta_u < 0$ from (9) and (10), since $\pi_A > \pi_B$, and since d_A and d_B are increasing functions. Hence the solutions are unique and the implicit functions exist and are continuous on $[0,1)$ and, by the continuous extension theorem, on $[0,1]$. Therefore they intersect at least once in $[0,1]^2$ by the Brouwer fixed point theorem, so an equilibrium with lying by both types exists.

Multiple such equilibria may exist, but we can still use the equilibrium requirement $\alpha_b > 0, \beta_b = 0, \alpha_u = 0, \beta_u > 0$ to compare equilibrium behavior with the case of transparent incentives. For an unbiased expert there is no lying when incentives are transparent, so there is more lying with opaque incentives. For a biased expert there is an incentive to lie in favor of A when incentives are transparent, so the question is whether lying is higher with opaque incentives. From (13), the incentive to lie toward A is increasing in P_A and decreasing in P_B. The latter must be lower with opaque incentives since there is no lying in favor of B when incentives are transparent, so P_A must be smaller with opaque incentives if the lying rate is smaller. But from (9) and (3) this requires $(2-\beta_u)/(2-\beta_u+\alpha_b) > 1/(1+\alpha)$ or $\alpha_b + \alpha\beta_u > 2\alpha$ where α is the lying rate with transparent incentives. This in turn requires $\alpha_b > \alpha$, so lying by biased experts must still be higher with opaque incentives.

Regarding acceptance probabilities, compared to when the expert is known to be unbiased, the increases in lying in both directions imply P_A and P_B both decrease. Compared to when the expert is known to be biased, P_B falls due to lying toward B by unbiased types, but the impact on P_A is ambiguous since biased types lie more toward A but with some chance the expert is unbiased and never lies toward A. ■

7 Appendix B – Instructions and Procedures

Subjects are led into the computer room and told to sit in front of any of 20 terminals. Instructions are provided on the terminals for the consultants and clients as the experiment progresses, and summaries are read aloud to ensure that information is common knowledge. A screenshot of the main introductory screen is in Appendix C and a summary is read aloud after subjects are seated.

After all subjects have clicked through the introductory instructions on their own terminals they see a screen which assigns them their roles for the rest of the experiment and provides instructions for the first Symmetric Baseline treatment (or "first series" as described to subjects). Screenshots of these instructions for both the consultant and client are included in Appendix C. The following summary is read aloud.

> **First Series [Symmetric Baseline Treatment]:** The consultant has two projects - Project A and Project B. One is a good project worth \$10 to the client and the other is a bad project worth \$0 to the client. Each round the computer randomly assigns one project to be good and tells the consultant. The client does not know which project is good and which is bad.
>
> The client has his/her own project - Project C. Each round Project C is randomly assigned by the computer to be worth any value between \$0.00 and \$10.00. Any such value is equally likely. The computer tells the client how much Project C is worth, but the consultant does not know.
>
> The consultant will give a recommendation to the client via the computer. The recommendation will be "I recommend Project A" or "I recommend Project B".
>
> After getting the recommendation, the client will make a decision. The client earns the value of the project that is chosen. If the client chooses Project A or Project B the consultant will earn \$8 in that round. However, if the client chooses his/her own Project C instead, the consultant will earn \$0. One round from this series will be randomly chosen at the end as the round you are actually paid for.

After subjects click through the instructions on their screen, in each round consultants see the values of A and B and make their recommendations, and clients see the recommendations and the values of V_C and make their choices. Screenshots for these recommendations and choices are also included in Appendix C. The realized values and choices are then revealed to each consultant-client pair at the end of each round. Subjects only see the realized values and choices for their own pairing in that round. After the 10 rounds are completed subjects see a summary of their actions and payoffs in each round. They then see a screen with instructions for the next Asymmetric Incentives treatment ("second series") and the following summary is read aloud.

> **Second Series [Asymmetric Incentives Treatment]:** Everything is the same as the first series, except if the client chooses Project A the consultant will earn \$10 and if the client chooses Project B the consultant will earn \$5.

After the Asymmetric Incentives treatment is completed in the same manner as the Symmetric Baseline treatment, the Opaque Incentives treatment ("third series") begins. This is the most complicated treatment but its description is facilitated by its being a mix of the first two treatments. Each consultant is assigned the incentives of the first or the second treatment with equal chance. The assignment is block random and is for the duration of the treatment. The consultants know their incentives, but the clients do not know whether the consultant in any given round is biased or not. The following summary of the treatment is read aloud while subjects see a screen with the detailed description.

> **Third Series [Opaque Incentives Treatment]:** Everything is the same as before, except the Computer will randomly assign the consultant's payoff scheme. Half of the consultants will earn \$8 if either Project A or B is chosen, and half of consultants will earn \$10 if project A is chosen but only \$5 if project B is chosen. The consultant knows his/her payoff scheme but the client does not know which payoff scheme the Computer assigned the consultant.

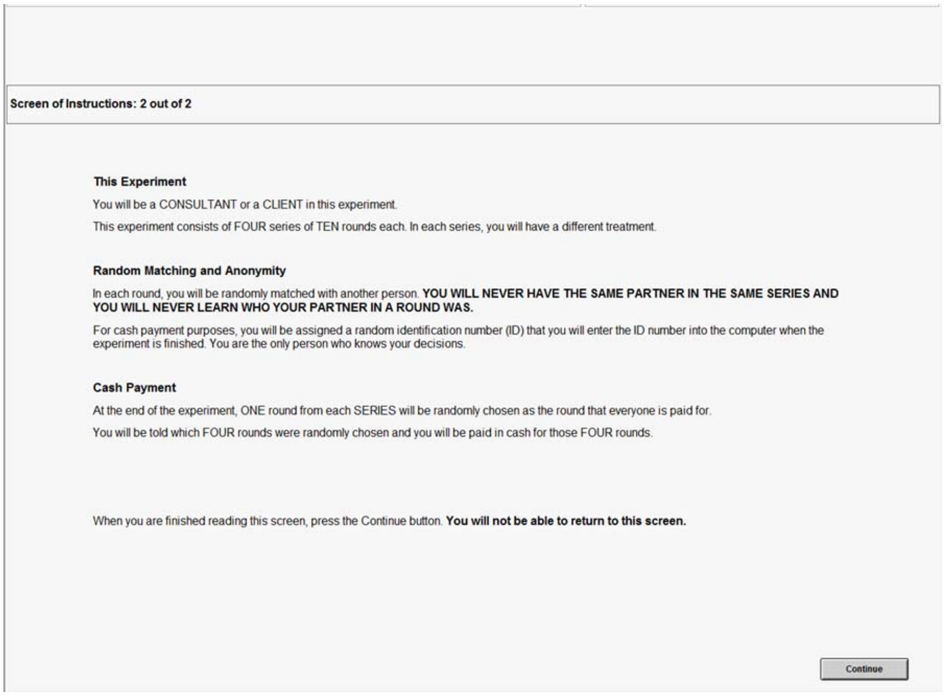
After the Opaque Incentives treatment is completed, the Asymmetric Values treatment ("fourth series") begins. The following summary of the treatment is read aloud while subjects see a screen with the detailed description.

> **Fourth Series [Asymmetric Values Treatment]:** Everything is the same as the first series, including the consultant's payoff scheme of earning \$8 if either Project A or B is chosen, except the value of Project B to the client if it is good is only \$5 instead of \$10. If project A is good its value to the client is still \$10. A bad project is still worth \$0 to the client.

After the Asymmetric Values treatment is completed, subjects see a screen summarizing their actions and payoffs across all 40 rounds. On this screen they are also told what rounds have been randomly chosen as the rounds they will be paid for. They enter into the computer a number 1-20 that has been randomly placed next to the terminal and they are called out in sequence by this number to be paid by the experimenter. They do not see the payments for any other subjects, they sign for their payment next to their number without a printed name, and the record of their signature is preserved separately by the university's bookkeeping department. Other than the signature, no personally identifying records are maintained.

8 Appendix C – Sample Screenshots

Main introductory screen:



Screen of Instructions: 2 out of 2

This Experiment

You will be a CONSULTANT or a CLIENT in this experiment.

This experiment consists of FOUR series of TEN rounds each. In each series, you will have a different treatment.

Random Matching and Anonymity

In each round, you will be randomly matched with another person. **YOU WILL NEVER HAVE THE SAME PARTNER IN THE SAME SERIES AND YOU WILL NEVER LEARN WHO YOUR PARTNER IN A ROUND WAS.**

For cash payment purposes, you will be assigned a random identification number (ID) that you will enter the ID number into the computer when the experiment is finished. You are the only person who knows your decisions.

Cash Payment

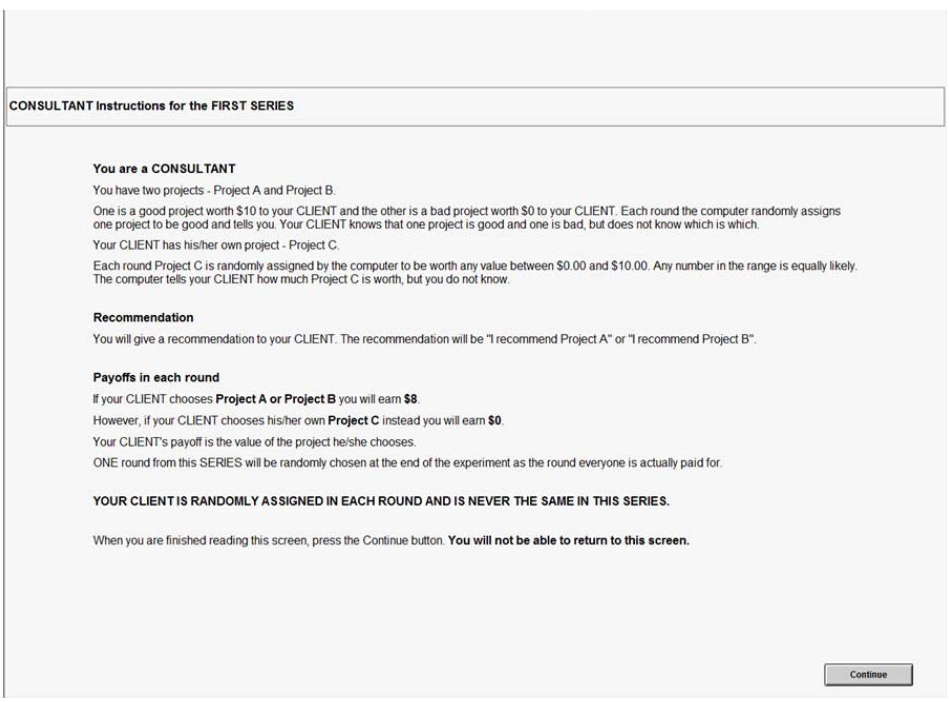
At the end of the experiment, ONE round from each SERIES will be randomly chosen as the round that everyone is paid for.

You will be told which FOUR rounds were randomly chosen and you will be paid in cash for those FOUR rounds.

When you are finished reading this screen, press the Continue button. **You will not be able to return to this screen.**

Continue

Consultant instructions for Symmetric Baseline treatment:



CONSULTANT Instructions for the FIRST SERIES

You are a CONSULTANT

You have two projects - Project A and Project B.

One is a good project worth $10 to your CLIENT and the other is a bad project worth $0 to your CLIENT. Each round the computer randomly assigns one project to be good and tells you. Your CLIENT knows that one project is good and one is bad, but does not know which is which.

Your CLIENT has his/her own project - Project C.

Each round Project C is randomly assigned by the computer to be worth any value between $0.00 and $10.00. Any number in the range is equally likely. The computer tells your CLIENT how much Project C is worth, but you do not know.

Recommendation

You will give a recommendation to your CLIENT. The recommendation will be "I recommend Project A" or "I recommend Project B".

Payoffs in each round

If your CLIENT chooses **Project A or Project B** you will earn **$8**.

However, if your CLIENT chooses his/her own **Project C** instead you will earn **$0**.

Your CLIENT's payoff is the value of the project he/she chooses.

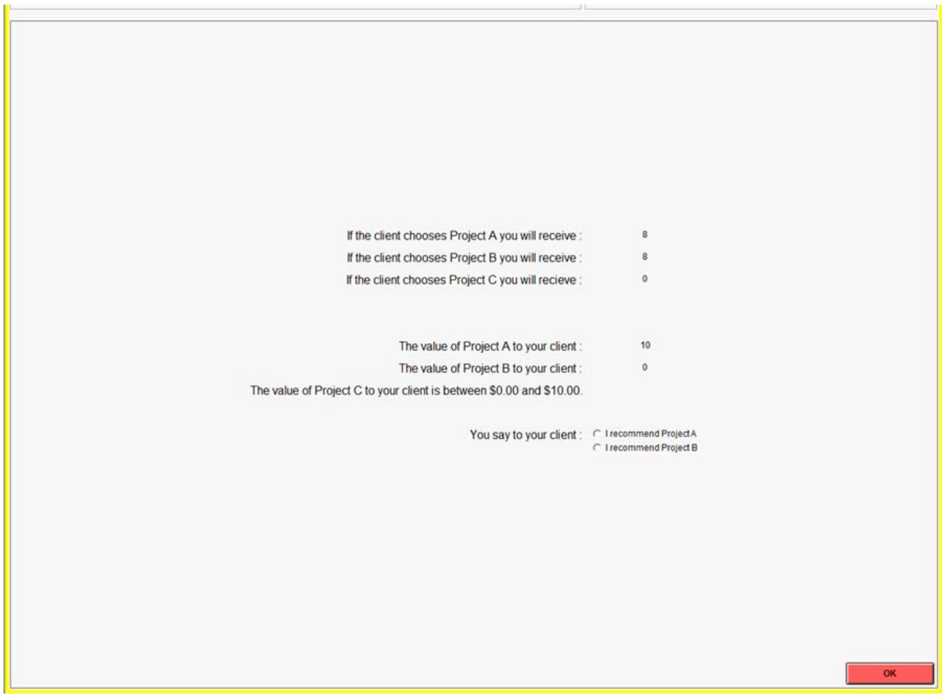
ONE round from this SERIES will be randomly chosen at the end of the experiment as the round everyone is actually paid for.

YOUR CLIENT IS RANDOMLY ASSIGNED IN EACH ROUND AND IS NEVER THE SAME IN THIS SERIES.

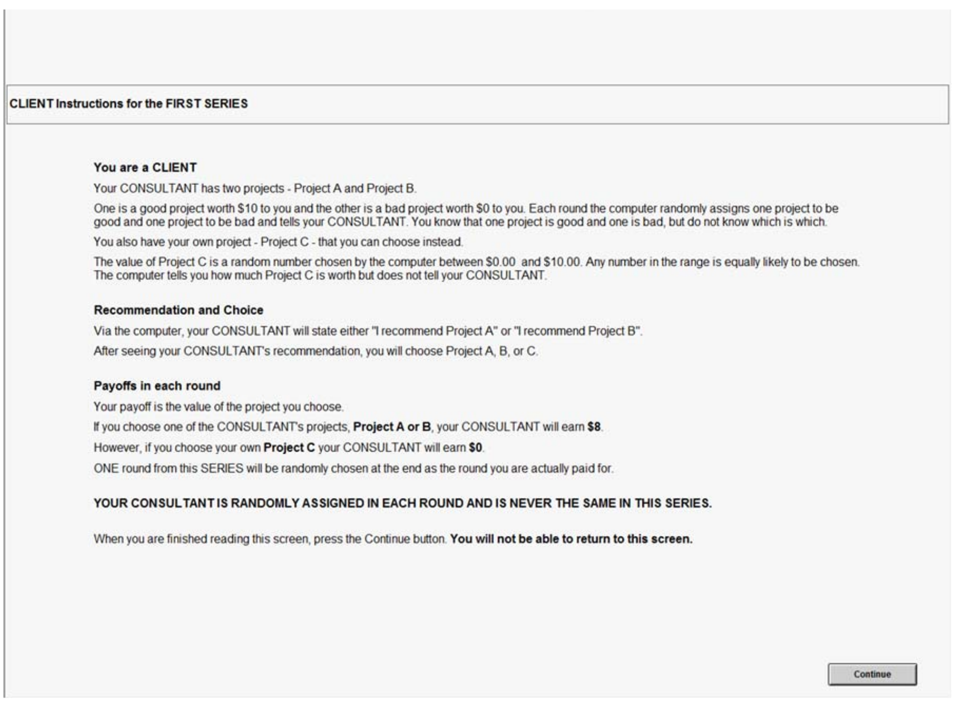
When you are finished reading this screen, press the Continue button. **You will not be able to return to this screen.**

Continue

Consultant decision screen for Symmetric Baseline treatment:



Client instructions for Symmetric Baseline treatment:



CLIENT Instructions for the FIRST SERIES

You are a CLIENT

Your CONSULTANT has two projects - Project A and Project B.

One is a good project worth $10 to you and the other is a bad project worth $0 to you. Each round the computer randomly assigns one project to be good and one project to be bad and tells your CONSULTANT. You know that one project is good and one is bad, but do not know which is which.

You also have your own project - Project C - that you can choose instead.

The value of Project C is a random number chosen by the computer between $0.00 and $10.00. Any number in the range is equally likely to be chosen. The computer tells you how much Project C is worth but does not tell your CONSULTANT.

Recommendation and Choice

Via the computer, your CONSULTANT will state either "I recommend Project A" or "I recommend Project B".

After seeing your CONSULTANT's recommendation, you will choose Project A, B, or C.

Payoffs in each round

Your payoff is the value of the project you choose.

If you choose one of the CONSULTANT's projects, **Project A or B**, your CONSULTANT will earn **$8**.

However, if you choose your own **Project C** your CONSULTANT will earn **$0**.

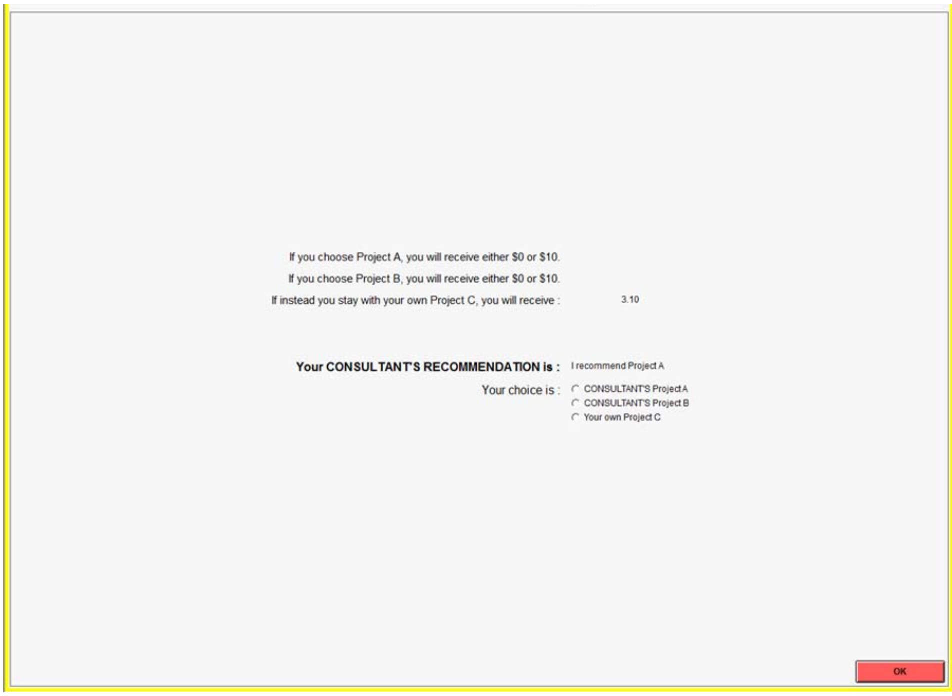
ONE round from this SERIES will be randomly chosen at the end as the round you are actually paid for.

YOUR CONSULTANT IS RANDOMLY ASSIGNED IN EACH ROUND AND IS NEVER THE SAME IN THIS SERIES.

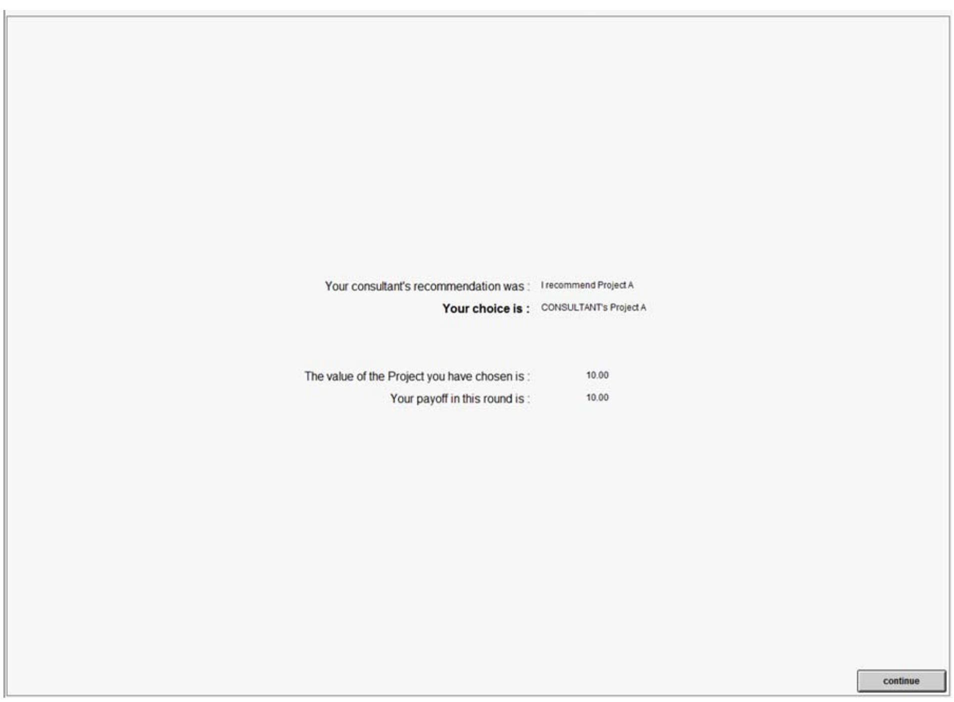
When you are finished reading this screen, press the Continue button. **You will not be able to return to this screen.**

Continue

Client decision screen for Symmetric Baseline treatment:



Client payoff screen for Symmetric Baseline treatment:



References

[1] Abbink, Klaus and Abdolkarim Sadrieh. 2009. The pleasure of being nasty. *Economics Letters* 105(3): 306–308.

[2] Abeler, Johannes, Daniele Nosenzo, and Collin Raymond. 2016. Preferences for truth-telling. Working paper.

[3] Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas D. Evanoff. 2014. Predatory lending and the subprime crisis. *Journal of Financial Economics* 113(1): 29–52.

[4] Armstrong, Mark and Jidong Zhou. 2011. Paying for prominence. *Economic Journal* 121(556): 368–395.

[5] Battaglini, Marco. 2002. Multiple referrals and multidimensional cheap talk. *Econometrica* 70(4): 1379–1401.

[6] 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): 1323–1340.

[7] Blume, Andreas, Douglas V. DeJong, Yong-Gwan Kim, and Geoffrey B. Sprinkle. 2001. Evolution of communication with partial common interest. *Games and Economic Behavior* 37: 79–120.

[8] Bolton, Patrick, Xavier Freixas, and Joel Shapiro. 2007. Conflicts of interest, information provision, and competition in the financial services industry. *Journal of Financial Economics* 85: 297–330.

[9] Cai, Hongbin, and Joseph Tao-Yi Wang. 2006. Overcommunication in strategic information transmission games. *Games and Economic Behavior* 56(1): 7–36.

[10] Cain, Daylian M., George Loewenstein, and Don A. Moore. 2005. The dirt on coming clean: perverse effects of disclosing conflicts of interest. *Journal of Legal Studies* 34(1): 1–25.

[11] Chakraborty, Archishman and Rick Harbaugh. 2007. Comparative cheap talk. *Journal of Economic Theory* 132(1): 70–94.

[12] Chakraborty, Archishman and Rick Harbaugh. 2010. Persuasion by cheap talk. *American Economic Review* 100(5): 2361–2382.

[13] Chakraborty, Archishman and Rick Harbaugh. 2014. Persuasive puffery. *Marketing Science* 33(3): 382–400.

[14] Che, Yeon-Koo, Wooter Dessein, and Navin Kartik. 2013. Pandering to persuade. *American Economic Review* 103(1): 47–79.

[15] Chiba, Saori and Kaiwen Leong. 2015. An example of conflicts of interest as pandering disincentives. *Economics Letters* 131: 20–23.

[16] Chung, Wonsuk. 2012. Product recommendations by a biased seller. Working paper.

[17] Crawford, Vincent. 2003. Lying for strategic advantage: Rational and boundedly rational misrepresentations of intentions. *American Economic Review* 93(1): 133–149.

[18] Crawford, Vincent P. and Joel Sobel. 1982. Strategic information transmission. *Econometrica* 50(6): 1431–1450.

[19] Darby, Michael R. and Edi Karni. 1973. Free competition and the optimal amount of fraud. *Journal of Law and Economics* 16(1): 67–88.

[20] De Jaegher, Kris and Marc Jegers. 2001. The physician-patient relationship as a game of strategic information transmission. *Health Economics* 10: 651–658.

[21] DellaVigna, Stefano and Ethan Kaplan. 2007. The Fox News effect: Media bias and voting. *Quarterly Journal of Economics* 122(3): 1187–1234.

[22] Dessein, Wouter. 2002. Authority and communication in organizations. *Review of Economic Studies* 69(4): 811–838.

[23] Dickhaut, John, Kevin McCabe, and Arijit Mukherji. 1995. An experimental study of strategic information transmission. *Economic Theory* 6(3): 389–403.

[24] Diehl, Christoph and Christoph Kuzmics. 2018. The (non-)robustness of influential cheap talk equilibria when the sender's preferences are state independent. Working paper.

[25] Dulleck, Uwe, Rudolf Kerschbamer, and Matthias Sutter. 2011. The economics of credence goods: an experiment on the role of liability, verifiability, reputation, and competition. *American Economic Review 101* (2): 526–555.

[26] Egan, Mark, Gregor Matvos, and Amit Seru, 2016. The market for financial advisor misconduct. *Journal of Political Economy*, forthcoming.

[27] Fischbacher, Urs. 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2): 171–178.

[28] Gibson, Rajna, Carmen Tanner, and Alexander F. Wagner. 2013. Preferences for truthfulness: heterogeneity among and within individuals. *American Economic Review* 103(1): 532–548.

[29] Gneezy, Uri. 2005. Deception: the role of consequences. *American Economic Review* 95(1): 384–394.

[30] Gneezy, Uri, Agne Kajackaite, and Joel Sobel. 2018. Lying aversion and the size of the lie. *American Economic Review*, 108(2): 419–453.

[31] Hurkens, Sjaak and Navin Kartik. 2009. Would I lie to you? On social preferences and lying aversion. *Experimental Economics* 12(2): 180–192.

[32] Iizuka, Toshiaki. 2012. Physician agency and generic pharmaceuticals. *American Economic Review* 102(6): 2826–2858.

[33] Inderst, Roman and Marco Ottaviani. 2012. Competition through commissions and kickbacks. *American Economic Review* 102(2): 780–809.

[34] Kartik, Navin. 2009. Strategic communication with lying costs. *Review of Economic Studies* 76(4): 1359–1395.

[35] Kartik, Navin, Marco Ottaviani, and Francesco Squintani. 2007. Credulity, lies, and costly talk. *Journal of Economic Theory* 134(1): 93–116.

[36] Lafky, Jonathan and Alistair J. Wilson. 2015. Quality vs. quantity in information transmission: theory and evidence. Working paper.

[37] Lai, Ernest K., Wooyoung Lim, and Joseph Tao-Yi Wang. 2015. An experimental analysis of multidimensional cheap talk. *Games and Economic Behavior* 91: 114–144.

[38] Li, Ming and Kristóf Madarász. 2008. When mandatory disclosure hurts: expert advice and conflicting interests. *Journal of Economic The ory* 139(1): 47–74.

[39] Morgan, John and Phillip C. Stocken. 2003. An analysis of stock recommendations. *RAND Journal of Economics* 34(1): 183–203.

[40] Morris, Stephen. 2001. Political correctness. *Journal of Political Economy* 109(2): 231–265.

[41] Nagel, Rosemarie. 1995. Unraveling in guessing games: an experimental study. *American Economic Review* 85(5): 1313–1326.

[42] Pitchik, Carolyn and Andrew Schotter. 1987. Honesty in a model of strategic information transmission. *American Economic Review* 77(5): 1032–1036.99

[43] Sánchez-Pagés, Santiago and Marc Vorsatz. 2007. An experimental study of truth-telling in a sender–receiver game. *Games and Economic Behavior* 61(1): 86–112.

[44] Sobel, Joel. 1985. A theory of credibility. *Review of Economic Studies* 52(4): 557–573.

[45] Sobel, Joel. 2016. Lying and deception in games. Working paper.

[46] Stahl, Dale O. and Paul W. Wilson. 1995. On players' models of other players: theory and experimental evidence. *Games and Economic Behavior* 10(1): 218–254.

[47] Sutter, Matthias. 2009. Deception through telling the truth?! Experimental evidence from individuals and teams. *Economic Journal* 119(534): 47–60.

[48] Thaler, Richard H. and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness.* Yale University Press.

[49] Vespa, Emanuel and Alistair Wilson. 2016. Communication with multiple senders: an experiment. *Quantitative Economics* 7(1):1–36.

[50] Wang, Joseph Tao-yi, Michael Spezio, and Colin F. Camerer. 2010. Pinocchio's pupil: using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review* 100(3): 984–1007.