Manipulative Disclosure

Claudio Mezzetti

University of Melbourne
Department of Economics
Parkville, VIC 3010, Australia
cmez@unimelb.edu.au

PRELIMINARY AND INCOMPLETE

March 2, 2012

Abstract

We study verifiable information disclosure by an informed expert when the direction of the expert's bias is unkown to the decision maker. We show that the expert is able to manipulate the decision maker. A positive measure of expert types induces the decision maker to choose their most favorite outcome (full manipulation). All the other types are able to engage in partial manipulation. The decision maker's choice coincides with her first best outcome only when the expert has zero bias, or when her ideal choice is a boundary point and the expert would prefer a more extreme outcome. An increase in the average bias, or a reduction in the variance of the state of the world, benefits the decision maker and hurts the expert.

Journal of Economic Literature Classification Numbers: D44, D82.

Keywords: disclosure, information manipulation

1 Introduction

When they do not know all the relevant information, decision makers must rely on the advice of experts. Sometimes the experts have no direct connection with the different alternatives, as when one moves to a new city and a colleague who has lived there for a long time suggests the best restaurants. Other times experts have stakes in the decision, as the sale representatives in a department store. Trusting an expert with no stake in the decision is often sensible. It is also reasonable to expect that good advice will be forthcoming, if one is engaged in a long-term relationship with the expert. Matters are not so clear when one is in a one-off relationship with an expert who may have stakes in the decision. In such circumstances, when should one trust the advice?

There is an extended economic literature devoted to the study of this problem. For the purpose of this paper, it is useful to distinguish between two strands: cheap talk and disclosure of verifiable information. In the cheap talk literature, the informed agent simply talks. The messages sent have no verifiable meaning. Anything that expert A says could be said by expert B, irrespective of the underlying information the two have. This paper focuses on the disclosure of verifiable information. In many economic, social, and political settings, disclosure of verifiable information is the norm. If rating agency A has discovered hidden liabilities in the balance sheet of company C, it can fully or partially report them. If there are no such liabilities in the balance sheet of company D, then rating agency B cannot mimic the report of A. Media companies can decide what news to report, but fabricating them is (typically) not an option. Scientists working for a biotechnology firm may or may not publish the results of their research, but they cannot publish research they have not undertaken. Experts that possess verifiable information can decide to disclose all or part of it, but they cannot manufacture and disclose facts that are not true.

The standard assumption in the literature on disclosure of verifiable information is that the preferences of the expert are well known. A buyer, for example, knows that the salesman is trying to sell as much as possible of a given product. The buyer can then safely assume that all positive information about the quality of the product will be disclosed, because it will induce the buyer to buy more. The privacy of the salesman information unravels; all information will be disclosed and the buyer will be able to select her first best alternative

(Grossman, 1981, and Milgrom, 1981).

The assumption that the preferences of the expert are fully known nicely fit many situations, but not all. Consider a rating agency discovering bad news about a large company; it will have an incentive to investigate thoroughly and report the bad news in order to preserve its reputation, but it will also have an incentive to gloss over the bad news in order to appease the large company and get its business again in the future. A biotechnology firm will have an incentive to let its scientists publish all preliminary research in order to gain financing, but it also has an incentive not to publish to make it more difficult for competitors to imitate. Even in the case of a salesman, it may not always be clear what his incentives are. When shopping for a washing machine, for example, a buyer may not know whether the salesman stands to gain more by selling brand A or brand B. In general, when information is not unidimensional it is particularly difficult to be sure about the precise direction of the expert's preferences.

So, what if the decision maker, be it a single agent or "the market," doesn't fully know the precise direction of an expert's bias? The goal of this paper is to shed light on this question.

The paper takes a minimalist modelling approach, by simply changing one assumption in an otherwise standard model. The model used is the work-horse model of the literature on cheap talk, in which the expert knows the state of the world, which coincides with the first best alternative for the decision maker. Rather than cheap talk messages, the expert must disclose verifiable information. The fundamental difference with the existing literature is that the decision maker does not know the direction of the expert's bias. The expert may want to distort choices up or down.

Many new insights arise. Full disclosure is never an equilibrium. Quite to the contray, almost all expert types are able to manipulate the decision maker and induce the choice of an alternative that is close to their favorite outcome. Indeed, a positive measure of expert types will be able to obtain their first best outcome. We call this full manipulation. Expert types who observe a non-extreme state are all able to manipulate the decision maker. Those with non-extreme biases will achieve full manipulation, while those with extreme biases will partially manipulate. The only experts that fully disclose are the ones whose preferences

are perfectly aligned with the decision maker, and the ones that have observed an extreme state and would prefer an even more extreme outcome. These is no room for these types to distort the decision in a direction that is favorable to them.

The intuition for why manipulation occurs is simple and instructive. Because the direction of the expert's bias is uncertain, the decision maker cannot fully decode partial disclosures that do not fully reveal the state. A partial disclosure is consistent with an expert that has observed a lower state but would prefer a higher choice, and with an expert that has observed a higher state but would prefer a lower choice. The compromise choice of the decision maker ends up being exactly what the expert wants.

We show that, if given a choice, experts would want to maintain uncertainty about their biases and incentives. Decision makers, on the other hand should not worry about getting advice from a strongly biased expert, as long as the direction of the bias is known. An increase in the size of the expert's mean bias benefits the decision maker and hurts the expert. This is because disclosures from more highly biased experts are easier to read.

The paper proceeds as follows. The remaining part of the introduction discusses related literature. Section 2 introduces the model. Section 3 shows that any equilibrium involves full manipulation by a positive measure of expert types. Section 4 studies equilibrium in a simplified version of the model. Section 5 contains comparative statics results. Section 6 derives equilibrium in the general model. Section 7 concludes.

Literature Review. The starting papers in the literature on disclosure of verifiable information are Grossman (1981) and Milgrom (1981). They assume that the bias of the expert is common knowledge. Their main insight is that the decision maker ought to be highly skeptical. When evidence is incomplete, the decision maker assumes that any missing information is likely to be unfavorable to the expert. As a consequence, experts will want to reveal all favorable information. This unravel any attempt to hide information and leads to full disclosure. A few papers have introduced some uncertainty in the expert's preferences, but none have gone as far as this paper in demonstrating the potential for full manipulation by the expert when the bias direction is unknown. Seidmann and Winter (1997) allow uncertainty over the expert's preferences, but they focus on the conditions under which there is an equilibrium with full disclosure. In Shin (1994), full disclosure may fail because

of uncertainty over whether the expert knows the true state. In such a case the decision maker's skepticism is tempered. Evidence may be incomplete because the expert does not have all the information. In Wolinsky (2003), the expert's bias is unknown, but it can only take two values. The expert can fully report, or under-report favorable information. In equilibrium, the biased expert with favorable information above a threshold fully discloses, while all other expert types play a mixed strategy. In Dziuda (2010), it is uncertain whether the expert is honest or biased in favor of one of two alternatives. Her focus is showing that a biased expert may also disclose unfavourable information.

Kamenica and Gentzkow (2009) and Rayo and Segal (2010) study models in which the informed party is able to commit to a disclosure policy before receiving private information. Rayo and Segal is closer to this paper, as they assume, like we do, that the sender's information is two-dimensional. Their focus is very different, however. Their lead example is an internet platform that is both informed about the profitability of different ads (or prospects) and their relevance to users. They show that the optimal policy is partial information disclosure, which can be interpreted as bundling prospects that are relevant to users (but less profitable) with irrelevant (but profitable) prospects.

The literature on cheap talk is vast and cheap talk has quite different consequences from disclosure of verifiable information. Still, we should mention two related cheap talk papers in which the preferences of the privately informed party (the sender) are not perfectly known to the receiver. Dimitrakas and Sarafidis (2005) show that even if the bias of the sender is uncertain, with cheap talk equilibrium is partitional. In Li and Madarász (2008), the sender has private information about her bias and transmits payoff relevant information to the receiver through cheap talk messages. They show that, in many circumstances, mandating that the sender discloses her bias reduces the sender's ability to transmit information to the receiver and reduces welfare for both parties.

2 The Model

There are two players: a decision maker, DM, and an expert, EX. EX knows the true value of the state of the world $\omega \in [0,1]$; DM does not. In the first stage, EX decides how

much verifiable information about the state of the world to disclose. After observing EX's disclosure, DM chooses an alternative. The set of possible alternatives is the real line.

We model the state of the world as the proportion of heads that results from a continuum of pairwise exchangeable coin tosses. Let the index $z \in [0,1]$ indicate the continuum of tosses (random variables) and let the indicator function w be $w(z) = 1$ when the outcome of the toss z is head and $w(z) = 0$ when the outcome is tail, then the value of the state of the world is the Lebesgue integral

$$\omega = \int_{[0,1]} w(z)dz. \tag{1}$$

The distribution of the continuum of coin tosses is unknown a priori to all players; hence, the value of ω is also unknown. Let $F(\omega)$ be the prior distribution of the state ω. We assume that F admits a positive density $f(\omega)$ at all points in $[0,1]$. EX observes the realization of a countably infinite Monte Carlo draws from the continuum of tosses. Thus, by the law of large numbers (Hammond and Sun, 2003), EX's posterior probability about the state of the world puts unit mass on its true value. DM knows what tosses EX has observed, but does not observe their realization. After observing the tosses, and as a function of their realization, the expert decides the proportion $h(w)$ of heads and $t(w)$ of tails she discloses to DM. Disclosures are verifiable. The decision maker chooses an alternative after having observed the expert's disclosure.

The measure of observed heads is $h(\omega)\,\omega$, while $t(\omega)(1-\omega)$ is the measure of observed tails. Thus, the disclosure of the tosses by the expert tells DM that $\omega \in [h(\omega)\,\omega, 1 - t(\omega)(1-\omega)]$; that is, the disclosure truncates the support of ω to an interval $[a,b]$ containing the true state ω. Conversely, any interval $[a,b]$ containing the true state could be the support of DM's posterior resulting from an expert's disclosure policy. Thus, we can work with the reduced form model in which EX knows the true state, and her disclosure policy consists of verifiably showing DM an interval $I = [a,b]$ to which the true state ω belongs.

DM's posterior distribution after observing $[a,b]$ and without using the expert strategy to update is $F(\omega) = 0$ for $\omega \leq a$, $F(\omega) = 1$ for $\omega \geq b$ and $F(\omega) = \frac{F(\omega)-F(a)}{F(b)-F(a)}$ for $\omega \in [a,b]$. Naturally, if DM knew that the expert's policy is to reveal a proportion p of heads, or q of tails, then upon observing $p\omega$ heads, or $q\omega$ tails, DM could infer the true value of ω. In general, using EX's equilibrium strategy allows DM's to further refine her beliefs.

We assume that the decision maker's goal is to choose the alternative that minimizes the Euclidean distance from the state of the world. That is, we take the state of the world to represent the decision maker's ideal choice. Thus, if x is the choice and ω is the state, then DM's payoff is

$$V(x, \omega) = -(x - \omega)^2.$$

We could adopt a more general formulation; what we will use is that DM's optimal choice is the alternative which equals the expected value of the state, conditional on the information directly disclosed by the expert and the expert's equilibrium strategy.

The expert's preferences need not be perfectly aligned with DM's preferences. The simplest way to model this is to assume that the expert has a bias β and that her payoff depends on the distance between the chosen alternative x and $\omega + \beta$:

$$U(x, \omega, \beta) = -(x - \omega - \beta)^2.$$

The ideal choice of the expert diverges from the ideal choice of DM, state ω, by the bias β.

Our main point of departure from the literature on disclosure of verifiable information is that we assume that the bias β is unknown to DM; β is private information of the expert and viewed as a random variable with support $[\beta_L, \beta_H]$ by DM. As we shall see, the interesting case is when $\beta_L < 0 < \beta_H$; in such a case DM is uncertain about the direction (as opposed to just the size) of EX's bias. We allow β to be correlated with ω and denote with $G(\beta|\omega)$ its conditional distribution. We assume that G admits a positive density g for all β and ω.

An expert's type is a pair (ω, β) of a state and a bias; let $\mathcal{T} = \{(\omega, \beta) : \omega \in [0, 1], \beta \in [\beta_L, \beta_H]\}$ be the set of possible types of the expert. Let $\mathcal{C}$ be the set of closed subintervals of the unit interval. A strategy for the expert is a function $\sigma_{EX} : \mathcal{T} \to \mathcal{C}$. A strategy for DM is a function $\sigma_{DM} : \mathcal{C} \to \mathbb{R}$.

We are interested in the (pure) perfect Bayesian equilibria (PBE) of the game. A PBE is a triple $\langle \sigma_{EX}, \sigma_{DM}, \mu \rangle$, where σ_{EX} and σ_{DM} are the players' strategies and μ is a map that associates to each $I \in \mathcal{C}$ a probability density over $\mathcal{T}$, representing DM's posterior beliefs about EX's type after EX's disclosure of I.

Let $\sigma_{EX}^{-1}(I) = \{(\omega, \beta) \in \mathcal{T} : \sigma_{EX}(\omega, \beta) = I\}$ be the inverse image of I; $\sigma_{EX}^{-1}(I)$ is the set of EX's types that disclose interval I. To be a PBE the triple $\langle \sigma_{EX}, \sigma_{DM}, \mu \rangle$ must

satisfy the following conditions:

$$\mu(\omega,\beta|I)=\begin{cases}\frac{f(\omega)g(\beta|\omega)}{\int_{(\widetilde{\omega},\widetilde{\beta})\in\sigma_{EX}^{-1}(I)}f(\widetilde{\omega})g(\widetilde{\beta}|\widetilde{\omega})d\widetilde{\omega}d\widetilde{\beta}} & \text{if} \quad \omega\in I \quad \text{and} \quad \sigma_{EX}^{-1}(I)\neq\varnothing \\ 0 & \text{if} \quad \omega\notin I\end{cases} \tag{2}$$

$$\sigma_{DM}(I)=E_{\mu}[\omega|I]=:\int_{(\omega,\beta)\in\sigma_{EX}^{-1}(I)}\omega\mu(\omega,\beta|I)\,d\omega d\beta \tag{3}$$

$$\sigma_{EX}(\omega,\beta)\in\arg\min_{\{I\in C:\ \omega\in I\}}(\sigma_{DM}(I)-\omega-\beta)^2. \tag{4}$$

Condition (2) says that on the equilibrium path DM's posterior beliefs about EX's type are consistent with EX's strategy and put zero mass on states outside the disclosed interval.[1] Condition (3) says that DM's equilibrium strategy is to choose the expected value of the state conditional on his posterior beliefs. Condition (4) requires that EX chooses the disclosure interval containing the state ω that maximizes her payoff, given DM's strategy.

Define the composite outcome map $\alpha:\mathcal{T}\to\mathbb{R}$ as $\alpha=\sigma_{DM}\circ\sigma_{EX}$; then, $\alpha(\omega,\beta)=\sigma_{DM}(\sigma_{EX}(\omega,\beta))$ is the alternative chosen in equilibrium by DM when EX's type is (ω,β). It is the norm for disclosure games to have multiple equilibria that induce the same outcome map. For this reason, in discussing equilibrium we will focus on the equilibrium map and will not present all possible outcome equivalent equilibria that generate it.

It is instructive to end this section with the benchmark case in which the direction of the bias is common knowledge; that is, β_L and β_H have the same sign. Without loss of generality, take $0\leq\beta_L\leq\beta_H$. It is then commonly known that EX would like to push DM's choice upward. Given this, it is natural for DM to believe that EX has shown all tosses that result in head; in other words, given any disclosed interval $[a,b]$, DM puts all probability mass on $\omega=a$ and hence chooses $\sigma_{DM}([a,b])=a$. Given these beliefs, it is a best reply for EX to reveal all tosses that resulted in head: $\sigma_{EX}(\omega,\beta)=[\omega,b]$ with $b\geq\omega$.[2] The equilibrium outcome is full disclosure, the same that results when the bias is fully known (e.g., see Milgrom, 1981 and Grossman, 1981); DM always obtain the information needed

[1] Thus, the only restriction on beliefs that follow disclosure of an interval I that is not on the equilibrium path is that all mass be on I.

[2] Note that there are multiple, outcome equivalent, equilibria, that only vary in the upper bound of the interval $[\omega,b]$.

to selects his first best alternative. Knowing the direction of EX's bias is as good for DM as knowing the true value of the bias.

Proposition 1 *If $0 \leq \beta_L \leq \beta_H$, then the unique equilibrium outcome map is $\alpha(\omega, \beta) = \omega$ for all $(\omega, \beta) \in \mathcal{T}$; the expert's disclosure fully reveals the true state of the world and the decision maker achieves her first-best outcome.*

Proof. First, note that in all equilibria it must be $\alpha(\omega, \beta) \geq \omega$ for all (ω, β). If it were $\alpha(\omega, \beta) < \omega$ for some type (ω, β), then type (ω, β) could profitably deviate by disclosing $[\omega, \omega]$, which would certainly lead DM to chooses $\sigma_{DM}([\omega, \omega]) = \omega$. Second, suppose that $\alpha(\omega', \beta') = x > \omega'$ for some (ω', β'). Then, there must be a disclosure I' that is made by EX's type (ω', β') and also by at least another type (ω'', β'') with $\omega'' > x$, because it must be $x = E_\mu[\omega | I']$. But then type (ω'', β'') would gain by deviating and disclosing $[\omega'', \omega'']$. ■

In the remainder of the paper we assume $\beta_L < 0 < \beta_H$, so that the bias direction is uncertain.

3 Full Manipulation

We say that the equilibrium outcome map α involves full manipulation by expert type (ω, β), or, alternatively, EX type (ω, β) fully manipulates DM, if $\alpha(\omega, \beta) = \omega + \beta$; that is, if EX type (ω, β) induces DM to choose her first best outcome. In this section, we prove that in any equilibrium of the disclosure game a positive mass of EX's types fully manipulates DM.

First, we need to derive some general properties of the equilibrium outcome map α. The first lemma is obvious: unbiased expert types induce the decision maker to choose the alternative that is first best optimal for both players. When the interests of EX and DM are perfectly aligned, the expert could always disclose all coin tosses, thus making sure that the decision maker picks the first best outcome.

Lemma 1 *For all ω, it is $\alpha(\omega, 0) = \omega$.*

Proof. Suppose, to the contrary, that $\alpha(\omega, 0) = x \neq \omega$. By (2), disclosing interval $[\omega, \omega]$ induces DM to choose ω, and is thus a profitable deviation for type $(\omega, 0)$, a contradiction.

■

The next lemma, whose proof is left to the appendix, shows that the equilibrium outcome $\alpha(\omega, \beta)$ must be increasing in β, because types with the same state ω can mimic each other and disclose the same interval. The outcome map α must be continuous at points $(\omega, 0)$ where EX's bias is zero; in addition, at a continuity point α must either be constant or select EX's first best outcome.

Lemma 2 *For all ω, the map $\alpha(\omega, \beta)$ is: (i) weakly increasing and hence continuous a.e. in β; (ii) either continuous from the left or from the right at a discontinuity point β; (iii) either constant or equal to $\omega + \beta$ in an interval where it is continuous. (iv) continuous at all points (ω, β) with $\beta = 0$.*

We are now ready to prove the main result of this section: a positive measure of EX's types achieve their first best outcome. The intuition for the proof of the result is the following. Suppose interval $[a, b]$ is disclosed. DM will choose an outcome $x \in [a, b]$ because any state outside the interval has zero probability of being the true state. Then, all EX's types that observe a state $\omega \in [a, b]$ and have x as their first best outcome (i.e., types such that $\omega + \beta = x$) are able to fully manipulate and make sure that x is chosen by DM. In particular, the set of types that are able to fully manipulate include types in an interval containing type $(x, 0)$. The continuity of the outcome map $\alpha(\omega, \beta)$ at $(x, 0)$ will enable us to conclude that for x' sufficiently close to x it is also the case that an interval of types containing $(x', 0)$ is able to fully manipulate DM.

Proposition 2 *In all equilibria, a positive measure of EX's types (ω, β) fully manipulates DM; that is, for a positive measure of types $\alpha(\omega, \beta) = \omega + \beta$.*

Proof. The proof is by contradiction. Suppose, to the contrary, that at most a zero measure of EX's types (ω, β) induce their first best outcome $\omega + \beta$. By Lemma 2, $\alpha(\omega, \beta)$ is continuous in an open rectangle around any point $(\omega_1, 0)$. Let $\varepsilon > 0$ and $\omega_1 \in (\varepsilon, 1 - \varepsilon)$. Take the open rectangle of types $R = \{(\omega, \beta) : \omega_1 - \varepsilon < \omega < \omega_1 + \varepsilon, -\varepsilon < \beta < \varepsilon\}$. If only a zero measure set of types (ω, β) in R are such that $\alpha(\omega, \beta) = \omega + \beta$, then by Lemmas 1 and 2 it must be $\alpha(\omega, \beta) = \omega$ for all types in R. Consider the interval $I = [\omega_1 - \varepsilon, \omega_1 + \varepsilon]$;

disclosing interval I must induce an outcome $\omega_2 \in [\omega_1 - \varepsilon, \omega_1 + \varepsilon]$. But then disclosing interval I is a profitable deviation for all types $(\omega, \beta) \in R$ with $\omega + \beta = \omega_2$, a contradiction. ■

It is important to emphasize that full manipulation by the expert is quite a different outcome from the standard full disclosure outcome that obtains when there is no uncertainty about the direction of the expert's bias. In essence, in our model partial disclosures (i.e., disclosures of a non degenerate interval) cannot be fully decoded by the decision maker. Disclosure $[a, b]$ could be made by an EX type that has observed a state as low as a and has a positive bias, or by a type that has observed a state as high as b and has a negative bias. No matter what alternative x the decision maker picks after such a disclosure, there will be EX types that have x as their first best choice and that can pool and disclose $[a, b]$. Full manipulation is made possible by this ability to pool and disclose the same proportion of coin tosses by expert's types with positive and negative bias that have a given outcome x as their first best choice.

Before presenting the second substantive result of this section, we now derive some additional properties of the equilibrium outcome map α. By Lemma 2(*iv*), the outcome map $\alpha(\omega, \beta)$ must be continuous at zero bias points (i.e., when $\beta = 0$). This suggests that it is natural to look at equilibrium maps $\alpha(\omega, \beta)$ that are continuous everywhere in β. By Lemma 2(*iii*), it must then be either $\alpha(\omega, \beta) = \omega + \beta$ and the expert gets it's favorite's outcome, or $\alpha(\omega, \beta)$ is constant in an interval around β. The next lemma shows that $\alpha(\omega, \beta)$ can only be constant in an interval including one of the boundary points β_L and β_H. Furthermore, if $\alpha(\omega, \beta)$ is continuous in β, then it must be monotone in ω. The proof is in the appendix.

Lemma 3 *If $\alpha(\omega, \beta)$ is continuous in β, then it must be: (i) $\alpha(\omega, \beta) = \omega + \beta$ in an interval $[\beta_1, \beta_2]$ with $\beta_L \leq \beta_1 \leq 0 \leq \beta_2 \leq \beta_H$, (ii) $\alpha(\omega, \beta) = \omega + \beta_1$ if $\beta \leq \beta_1$, and (iii) $\alpha(\omega, \beta) = \omega + \beta_2$ if $\beta \geq \beta_2$. (iv) If $\alpha(\omega, \beta)$ is continuous in β. Then it is is weakly increasing in ω.*

The second main result of this section shows that at the boundary points of the state space the equilibrium outcome map must satisfy certain boundary constraints. More pre-

cisely, EX types that observe state $\omega = 0$ and have a negative bias and types that observe state $\omega = 1$ and have a positive bias must fully disclose the state in equilibrium. The intuition is straightforward. Types that observe state $\omega = 0$ and have a negative bias prefer alternative $x = 0$ to any positive alternative. Their ideal outcome is some $x < 0$, but DM would never choose a negative outcome. Thus, among the outcomes that EX can feasibly induce DM to choose, EX's preferences are aligned with DM's. It is thus optimal for EX to disclose interval $[0, 0]$. Similarly, in the set $[0, 1]$ of "feasible" alternatives, the preferences of EX types that observe state $\omega = 1$ and have a positive bias are perfectly aligned with DM's preferences.

Proposition 3 *In all equilibria, there is full disclosure by boundary EX types that either observe the lowest state and have a negative bias, or observe the highest state and have a positive bias; that is* $\alpha(0, \beta) = 0$ *if* $\beta \leq 0$ *and* $\alpha(1, \beta) = 1$ *if* $\beta \geq 0$.

Proof. Consider EX's type $(0, \beta)$ with $\beta \leq 0$. Disclosing interval $[0, 0]$ is possible for this type and induces DM to select $x = 0$. DM never selects $x < 0$ because it is known that $\omega \in [0, 1]$ and EX prefers $x = 0$ to any positive outcome. Similarly, disclosing interval $[1, 1]$ is possible for EX's type $(1, \beta)$ with $\beta \geq 0$ and induces DM to select $x = 1$, which is preferred by EX to any $x < 1$. No $x > 1$ is ever selected by DM. ■

4 The Uniform Model

To present the main results and insights in the most transparent way, in this section we simplify the model by assuming that the expert's bias and the state of the world are independent, uniformly distributed, random variables. We will return to the general model in Section 6.

4.1 The Unbounded Case

In this sub-section, we make an additional modification of the model that simplifies the structure of the equilibrium.

As shown by Proposition 3, at the boundary of the state space types whose ideal decision is outside the $[0, 1]$ interval fully disclose the state. The boundedness of the set of states

of the world also constrain the equilibrium outcome for interior states that are close to the boundary. In order to gain intuition for the amount of manipulation the expert is able to achieve, and to focus on the equilibrium configuration in the interior of the state space, in this sub-section we modify the model so as to eliminate the boundary effects. We assume that when the proportion of heads that results from the coin tosses is η, the state of the world is $\omega = \varphi(\eta)$ where $\varphi : [0,1] \to \mathbb{R}$ is a function from the unit interval into the real numbers.[3] Thus, in this section definition (1) is replaced by

$$\omega = \varphi\left(\int_{[0,1]} w(z)dz\right).$$

In the next sub-section, we will go back to analyze the model with a bounded state space, where φ is the identity map.

We begin with the additional assumption that $-\beta_L = \beta_H$. This implies that the expected bias is zero. More importantly, it implies that the expected value of the state of all EX's types that have x as their favorite choice is x, $E[\omega|\omega + \beta = x] = x$.

Suppose that the two "twin" EX's types $(x - \beta, \beta)$ and $(x + \beta, -\beta)$ disclose the same interval (e.g., they disclose $[x - \beta, x + \beta]$). Suppose no other type discloses this interval. It is then immediate that DM's best response to such a disclosure is to choose outcome x, because x is the average value of the state. It is also immediate that the two twin EX types have no profitable deviation, because they induce DM to choose their first best outcome. Finally, note that since there is no boundary in our example and $\beta_L = -\beta_H$, for any EX type (ω, β) there is a twin type $(\omega + 2\beta, -\beta)$. Thus, all EX types may just pool with their twin and fully manipulate the decision maker. This argument is formalized in the following proposition.

Proposition 4 *In the unbounded uniform model, if $-\beta_L = \beta_H$, then $\alpha(\omega, \beta) = \omega + \beta$ for all EX types; the decision always coincides with the expert's optimal (first-best) decision.*

Proof. We claim that the following equilibrium strategies implement the outcome map $\alpha(\omega, \beta) = \omega + \beta$: $\sigma_{EX}(\omega, \beta) = [\omega - |\beta|, \omega + |\beta|]$; $\sigma_{DM}([x, y]) = \frac{x+y}{2}$. Note that: (1) Message $[x, y]$ is sent in equilibrium only by EX's types $\left(x, \frac{y-x}{2}\right)$ and $\left(y, \frac{x-y}{2}\right)$. Hence it is

[3] When combined with the assumption that the state is uniformly distributed, this implies that ω has an improper, diffuse distribution over the real line.

$E[\omega|[x,y]] = \frac{x+y}{2}$ and σ_{DM} is a best response for DM. (2) Given DM's strategy, by following σ_{EX} all types of expert obtains their first-best choice, hence they are playing a best response. ■

Figure 1 describes equilibrium in this case. All EX types on a negative 45 degree line have the same ideal outcome and are able to induce DM to choose it. In this extreme version of the example, all expert types fully manipulate the decision maker. Note that there is also an equilibrium with the same outcome map (full manipulation) in which all EX types that have the same ideal choice x pool and disclose the same interval (e.g., they disclose interval $[x - \beta_H, x + \beta_H]$).

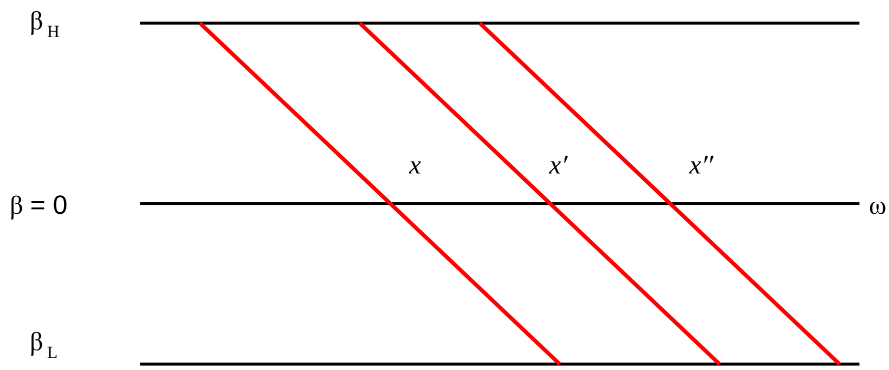

Figure 1: Unbounded Uniform Model, $-\beta_L = \beta_H$.

Suppose now that $-\beta_L < \beta_H$; that is, there are more EX types with a positive than a negative bias. The average bias is positive and the expected value of the state of all EX types that have x as their first best choice is less than x, $E[\omega|\omega + \beta = x] < x$. In this case, not all types have a twin with whom they can pool. More precisely, types with a high positive bias do not have a twin and they will only be able to partially manipulate the outcome by pooling with types that have a smaller ideal choice than them. Thus, a positive measure of expert types get their first-best outcome, while all other types get less than their first-best choice.

Before formally presenting the equilibrium in this case, it is useful to define the parameter β^* :

$$\beta^* = \frac{\beta_H - \sqrt{\beta_H^2 - \left(2\sqrt{2} - 2\right)\beta_L^2}}{2 - \sqrt{2}}$$

Proposition 5 *In the unbounded uniform model, if $-\beta_L < \beta_H$, then the following is an equilibrium outcome map:*

$$\alpha(\omega, \beta) = \begin{cases} \omega + \beta & \text{if } \beta \leq \beta^* \\ \\ \omega + \beta^* & \text{if } \beta > \beta^* \end{cases}$$

All expert types (ω, β) whose bias is not greater than β^*, or, equivalently, whose preferred choice is either below or above the true state ω by at most β^*, induce the DM to take their favorite decision. Strongly upward biased expert types with $\beta > \beta^*$ induce the decision maker to choose $\omega + \beta^*$. Since $\beta^* > 0$, all experts with a downward bias, and all experts with an upward bias less than β^*, are able to induce the DM to choose their favorite decision. Thus, we can think of β^* as a measure of the set of upward biased experts that obtain what they want.

Figure 2 describes the equilibrium outcome. The set of types that are able to induce the same choice x (the x-decision set) includes the types on the portion of the 45 degrees line that starts at type $(x - \beta^*, \beta^*)$ and ends at type $(x - \beta_L, \beta_L)$, and the set of types on the vertical line starting at type $(x - \beta^*, \beta^*)$ and ending at type $(x - \beta^*, \beta_H)$. Suppose all x-decision types disclose interval $[x - \beta^*, x - \beta_L]$. (There are many other outcome equivalent equilibria.) The critical value of the bias $\beta^* > 0$ is computed so that the expected value of the state on the x-decision set is equal to x. Thus, choosing x is indeed a best response for DM. To see that no EX type has a profitable deviation, note that types on the diagonal portion of the x-decision set obtain their first best outcome. Hence only types on the vertical portion could possibly want to deviate, if they could induce a higher choice than x. The only way to induce a choice higher than x is to disclose an interval with a left boundary higher than $x - \beta^*$, which is impossible for the types of the vertical portion of the x-decision set, since $x - \beta^*$ is the true state that they observe.

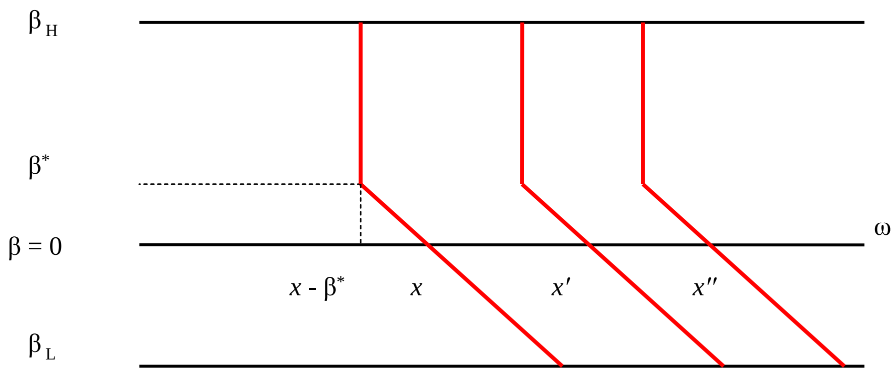

Figure 2: Unbounded Uniform Model, $-\beta_L < \beta_H$.

The case $\beta_L < -\beta_H$ is just the mirror image of the case just analyzed. Now $E[\omega|\omega + \beta = x] > x$ and equilibrium is described in Figure 3. The x-decision set includes the types on the portion of the 45 degrees line that starts at type $(x - \beta_H, \beta_H)$ and ends at type $(x - \beta^{**}, \beta^{**})$ and the set of types on the vertical line starting at type $(x - \beta^{**}, \beta^{**})$ and ending at type $(x - \beta^{**}, \beta_L)$. Now it is types with a high negative bias (below β^{**}) that can only partially manipulate the decision maker.[4]

In the remainder of the paper, we assume that $-\beta_L \leq \beta_H$ in order to cut down on the number of cases to consider; the analysis when $|\beta_L| > |\beta_H|$ mirrors the one we provide.

[4]It is

$$\beta^{**} = \frac{\beta_L + \sqrt{\beta_L^2 - (2\sqrt{2} - 2)\,\beta_H^2}}{2 - \sqrt{2}} < 0.$$

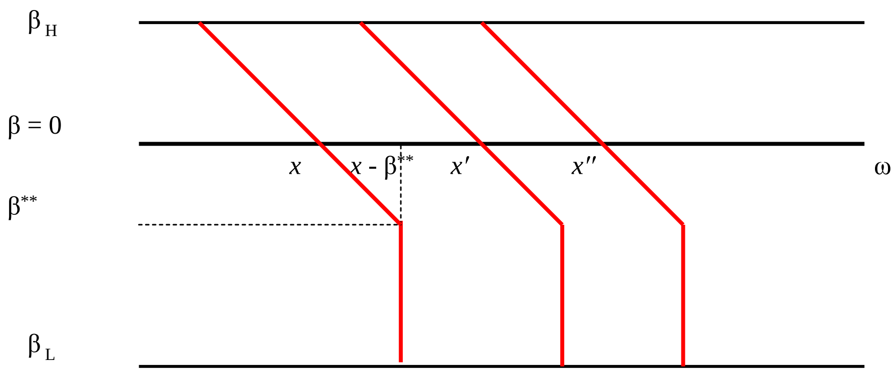

Figure 3: Unbounded Uniform Model, $-\beta_L > \beta_H$.

4.2 The Bounded Case

We now return to the bounded model with state of the world $\omega \in [0,1]$. As we saw in Proposition 3 boundary types whose ideal decisions are outside the state space fully disclose. Boundary effects also play a role for types close to the boundary.

In order to describe the equilibrium map, we need to define the following two functions of the state of the world:

$$\rho(\omega) = \frac{-\beta_L - \sqrt{\beta_L^2 - \left(2\sqrt{2}-2\right)\left(\min\left\{\omega, -\beta_L\right\}\right)^2}}{2-\sqrt{2}}$$

$$\lambda(\omega) = \frac{\beta_H - \sqrt{\beta_H^2 - \left(2\sqrt{2}-2\right)\left(1-\max\left\{\omega, 1+\beta_L\right\}\right)^2}}{2-\sqrt{2}}$$

The equilibrium is depicted in Figure 4. There is a cut-off outcome ω^* that is the choice induced by a positive measure of EX types with a positive bias (and an observed state less than ω^*) and a positive measure of types with negative bias (and an observed state higher than ω^*). In Figure 4 this is represented by the two shaded areas. Types on the 45 degrees line joining the two areas, which includes type $(\omega^*, 0)$, also induce DM to choose ω^*.

The choice of any outcome ω to the right of the cut-off ω^* is induced by types on the 45 degree line going from type $(\omega - \lambda(\omega), \lambda(\omega))$ to type $(\omega - \beta_L, \beta_L)$ and by types on the

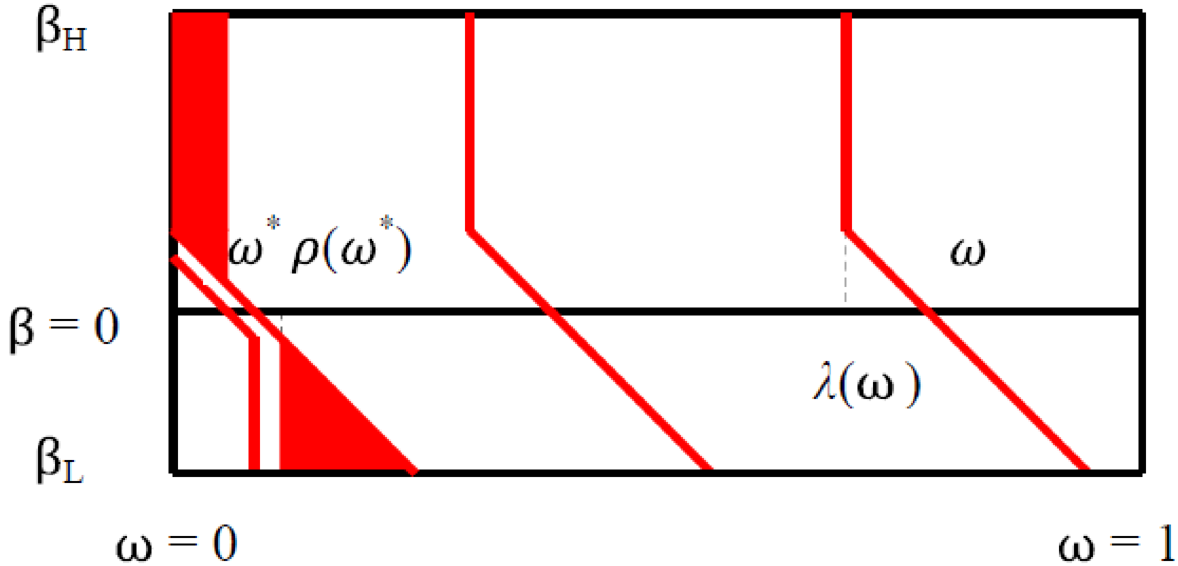

Figure 1: Figure 4: Bounded State Space $-\beta_L < \beta_H$.

vertical segment going from $(\omega - \lambda(\omega), \lambda(\omega))$ to $(\omega - \lambda(\omega), \beta_H)$.

Note that for $\omega \leq 1 + \beta_L$, it is $\lambda(\omega) = \beta^*$. This says that for types to the right of ω^* and sufficiently far away from the right boundary $\omega = 1$, equilibrium takes the same form as in the case of an unbounded state space (see Figure 2). Note also that $\lambda(\omega)$ is strictly decreasing in ω for $\omega > 1 + \beta_L$. This says that the ability of experts with a positive bias to manipulate DM gets smaller as as the state ω gets closer to the right boundary. In the limit $\lambda(1) = 0$ and all types with positive bias and observed state $\omega = 1$ fully disclose.

The equilibrium map to the left of ω^* resembles the equilibrium in Figure 3. Any outcomes ω to the left of the cut-off ω^* is induced by types on the 45 degree line going from type $(0, \omega)$ to type $(\omega + \rho(\omega), -\rho(\omega))$ and by types on the vertical segment going from $(\omega + \rho(\omega), -\rho(\omega))$ to $(\omega + \rho(\omega), \beta_L)$.

Observe that $\rho(\omega)$ is strictly increasing in ω for $\omega < -\beta_L$. This says that the ability of experts with a negative bias to manipulate DM gets smaller as as the state ω gets closer to the left boundary. In the limit $\rho(0) = 0$ and all types with negative bias and observed state $\omega = 0$ fully disclose.

The cut-off value ω^* is defined so that it is equal to the expected value of the state

in the shaded areas shown in Figure 4. Consider the 45 degree line that goes through type $(\omega^*, 0)$. The first shaded area is the one above the 45 degree line, starting from type $(\omega^* - \lambda(\omega^*), \lambda(\omega^*))$. The second is the area below the 45 degree line starting from type $(\omega^* + \rho(\omega^*), -\rho(\omega^*))$. Formally, ω^* is implicitly defined by the equation $\psi(\omega^*) = 0$, where the function ψ is given by:

$$\psi(\omega) = \int_0^{\max\{0,\omega-\lambda(\omega)\}} (x-\omega)(\beta_H - \omega + x)\, dx + \int_{\min\{1,\omega+\rho(\omega)\}}^{\min\{1,\omega-\beta_L\}} (x-\omega)(-\beta_L - x + \omega)\, dx$$

Lemma **??** in the appendix shows that ω^* exists and is uniquely defined.

We are now ready to formally present the equilibrium characterization.

Proposition 6 *In the bounded uniform model, if $-\beta_L < \beta_H$, then the following is an equilibrium outcome map:*

$$\alpha(\omega,\beta) = \begin{cases} \omega+\beta & \text{if } \omega+\beta \geq \omega^* \text{ and } \beta \leq \lambda(\omega+\beta) \\ \omega+\lambda(\omega) & \text{if } \omega \geq \max\{\omega^*-\beta, \omega^*-\lambda(\omega^*)\} \text{ and } \beta > \lambda(\omega+\beta) \\ \omega^* & \text{if } \omega^*-\lambda(\omega^*) \geq \omega \geq \omega^*-\beta, \text{ and } \beta > \lambda(\omega+\beta) \\ \omega^* & \text{if } \omega^*-\beta \geq \omega \geq \omega^*+\rho(\omega^*) \text{ and } \beta < -\rho(\omega+\beta) \\ \omega+\beta & \text{if } \omega+\beta \leq \omega^* \text{ and } \beta \geq -\rho(\omega+\beta) \\ \omega+\rho(\omega) & \text{if } \omega \leq \min\{\omega^*-\beta, \omega^*+\rho(\omega^*)\} \text{ and } \beta < -\rho(\omega+\beta) \end{cases}$$

If verifiable disclosure of the coin tosses was not feasible and the expert were only able to send cheap talk messages to the decision maker, equilibrium would be quite different. Indeed, it is simple to show that in the model with bounded state ($\omega \in [0,1]$) equilibrium would be partitional, as in the standard cheap talk model, with all expert types that have the same preferences (i.e., such that $\omega + \beta$ is the same) sending the same message (see Dimitrakas and Sarafidis, 2005). More precisely, one can find messages $m_1, m_2, ..., m_n$ with $n \geq 1$ such that all EX types with $x_{j-1} \leq \omega + \beta \leq x_j$ send message m_j, with $1 \leq j \leq n$

and $x_0 = 0$. In such an equilibrium, only a zero measure of expert types is able to fully manipulate DM. The only exception is the unbounded uniform model with $-\beta_L = \beta_H$. In that case there is a cheap talk equilibrium that replicates the full disclosure outcome of full manipulation by all EX's types. With that exception, the expert is surprisingly better off (in an ex-ante sense) if she is able to disclose verifiable information than if she must send cheap talk messages. It is also typically the case that the decision maker is better off if verifiable disclosure is available.

5 Comparative Statics

In this section, we derive comparative static results for the uniform model. We start by looking at how the measure of types that fully manipulates DM varies with the parameters. Then we look at how the expected losses of DM and EX vary. All proofs are in the appendix.

Proposition 7 *In the uniform model, the measure of expert types that fully manipulate the decision maker decreases with the size of the expert's mean bias and increases with the variance (or size of the range) of the state of the world.*

The intuition for effect of the mean bias can be more easily gained from the case with unbounded state. For a fixed size $\beta_H - \beta_L$ of the bias support, if the size of the mean bias $\frac{\beta_H+\beta_L}{2}$ is sufficiently large, then the direction of the bias is known (because β_L and β_H have the same sign) and the equilibrium outcome is full disclosure. On the other hand, if the mean bias is zero we are in the case described in Figure 1 (i.e., $-\beta_L = \beta_H$), and all expert's types fully manipulate DM. The proposition shows that as the mean bias changes we move monotonically from one extreme to the other.

An increase in the variance of the state of the world diminishes the importance of the boundaries. Since the measure of types that fully manipulate DM is smaller near the boundaries, an increase in the variance of the state increases the measure of the intermediate types and hence increases the measure of EX types that achieve full manipulation.

The next proposition shows that the expected losses of EX and DM move in opposite directions as the mean bias and the variance of the state change.

Proposition 8 *In the uniform model, the ex-ante expected loss of the expert increases with the size of the expert's mean bias, and decreases with the variance in the state. The expected loss of the decision maker decreases with the expert's mean bias, and it increases with the variance in the state.*

Propositions 7 and 8 make it clear that experts should strive to look ex-ante unbiased, while decision makers should look for experts with a large mean-bias size. This is because disclosures from experts with a large mean-bias size are easier to read; their preferences and incentives are more transparent than the preferences of experts with equal bias variance but a more uncertain direction of their bias. The propositions also show that the expert is in a better position to manipulate when the decision maker is more uncertain about the optimal choice (e.g., because he is less familiar with the problem).

6 Equilibrium in the General Model

In this section, we return to the general model and provide conditions under which equilibrium exists. Proposition 6 and the uniform case with bounded state space will guide us in constructing the equilibrium.

Consider the inverse α^{-1} of the map α. The iso-decision set $\alpha^{-1}(x)$ is the set of EX's types that induce outcome x :

$$\alpha^{-1}(x) = \{(\omega, \beta) : \alpha(\omega, \beta) = x\}.$$

The equilibrium of the disclosure game can be easily understood by partitioning EX's type space $\mathcal{T}$ into three regions. First, there exists a set of critical outcomes $[x_L^*, x_R^*]$ such that $x^* \in [x_L^*, x_R^*]$ is induced by all EX's types for which x^* is the first best outcome; that is, those satisfying $\omega+\beta = x$. Except for special examples (e.g., when ω and β are independent, uniform, random variables and $\beta_L = -\beta_H < 1/2$), this set is a singleton, $x_L^* = x_R^* = x^*$, and x^* is also the DM's choice induced by a positive mass of types who prefer a higher outcome and a positive mass of types that prefer a lower outcome. This region does not appear in the unbounded-state, uniform example of the previous section when $\beta_L \neq -\beta_H$.

The second region includes EX types that have observed a sufficiently low state and that induce the choice of an outcome $x \in [0, x^*)$. In this region $E[\omega|\omega + \beta = x] > x$ and, as

a result, the iso-decision sets look like the iso-decision sets in Figure 3. More precisely, the iso-decision set $\alpha^{-1}(x)$ is the union of two sets. The first is the (top-diagonal) set of EX's types for which x is the first best outcome and that have observed a state below $\rho(x) \geq x$ (and hence have a bias β above $x - \rho(x)$). The second is the (bottom-vertical) set of EX's types that have observed state $\omega = \rho(x)$ and have bias $\beta < x - \rho(x)$. The ideal choice of types in this second set is an outcome lower than x, but they are unable to achieve it, and pooling with types that have x as their first best outcome is the best they can do.

The third region includes all types that induce the choice of an outcome $x \in (x^*, 1]$. In this region $E[\omega|\omega + \beta = x] < x$ and the iso-decision sets look like the iso-decision sets in Figure 2. As in the second region, the iso-decision set $\alpha^{-1}(x)$ is the union of two sets. The first is the (bottom-diagonal) set of EX's types for which x is the first best outcome and that have observed a state above $\lambda(x) \leq x$ (and hence have a bias β below $x - \lambda(x)$). The second is the (top-vertical) set of EX's types that have observed state $\omega = \lambda(x)$ and have bias $\beta > x - \lambda(x)$. The ideal choice of types in this set is an outcome higher than x.

We now formally define the top and bottom, diagonal and vertical sets and the functions $\rho(x)$ and $\lambda(x)$.

Let $z \geq x$; the top-diagonal set of types $D_T(z;x)$ is the set of EX's types who have x as first best outcome and who observe a state $\omega \leq z$ (and hence have bias greater than or equal to $x - z$):

$$D_T(z;x) = \{(\omega, \beta) : \omega = x - \beta,\ \beta \geq x - z\}.$$

The bottom-vertical set of types $V_B(z;x)$ is the set of EX's types who observe state z and have a bias lower than $x - z$:

$$V_B(z;x) = \{(\omega, \beta) : \omega = z,\ \beta < x - z\}.$$

The union of the two sets is:

$$DV(z;x) = D_T(z;x) \cup V_B(z;x).$$

Let $z < x$; the bottom-diagonal set of types $D_B(z;x)$ is the set of EX's types who have x as first best outcome and whose bias is not higher than $x - z$ (hence they observe a state $\omega \geq z$):

$$D_B(z;x) = \{(\omega, \beta) : \omega = x - \beta,\ \beta \leq x - z\}.$$

The top-vertical set of types $V_T(z;x)$ is the set of EX's types who observe state z and have a bias higher than $x - z$:

$$V_T(z;x) = \{(\omega, \beta) : \omega = z,\ \beta > x - z\}.$$

The union is the set:

$$VD\,(z;x) = D_T\,(z;x) \cup V_B\,(z;x)\,.$$

In the second region, when $x < x^*$, the equilibrium outcome map is characterized by an increasing function $\rho : [0, x^*] \to [0,1]$: for any $x \in [0, x^*)$ the set of EX's types inducing outcome x is $T_{DV}\,(\rho\,(x)\,, x)$; that is, $\alpha^{-1}\,(x) = T_{DV}\,(\rho\,(x)\,, x)\,.$

In the third region, when $x > x^*$, the equilibrium outcome map is characterized by an increasing function $\lambda : [x^*, 1] \to [0,1]$: for any $x \in (x^*, 1]$ the set of EX's types inducing outcome x is $T_{VD}\,(\lambda\,(x)\,, x)$; that is, $\alpha^{-1}\,(x) = T_{VD}\,(\lambda\,(x)\,, x)\,.$ The functions ρ and λ are implicitly defined as the solutions of the following equalities:

$$\rho(x) \quad = \quad \min\left\{z : E\left[\omega | \omega \in DV\,(z;x)\right] - x = 0\right\} \tag{5}$$

$$\lambda\,(x) \quad = \quad \max\left\{z : E\left[\omega | \omega \in VD\,(z;x)\right] - x = 0\right\}. \tag{6}$$

In the proof of Proposition 9 we show that, under appropriate conditions, the functions exists and are uniquely defined. To state the needed conditions, we need to define two additional functions.

For all $x \in [0,1]$ and $z \in [x, 1]$ define:

$$\mathcal{H}(z;x) = E\left[\omega | \omega \in DV\,(z;x)\right] - x.$$

For all $x \in [0,1]$ and $z \in [0, x]$ define:

$$\mathcal{K}(z;x) = E\left[\omega | \omega \in VD\,(z;x)\right] - x.$$

Note the following properties of the functions $\mathcal{H}$ and $\mathcal{K}$:

(P1) if $\beta_H \leq x \leq 1 + \beta_L$, then $\mathcal{K}(0;x) = \mathcal{H}(1;x) = E\left[\omega | \omega \in \omega + \beta = x\right];$

(P2) $\mathcal{H}(0;0) = 0$ and, for all $x > 0$, $\mathcal{H}(x;x) < 0$;

(P3) $\mathcal{K}(1;1) = 0$ and, for all $x < 1$, $\mathcal{K}(x;x) > 0$;

(P4) if $z, z' \in [x - \beta_L, 1]$, then $\mathcal{H}\,(z, x) = \mathcal{H}\,(z', x)$;

(P5) if $z, z' \in [0, x - \beta_H]$, then $\mathcal{K}(z, x) = \mathcal{K}(z', x)$.

$\mathcal{K}$ and $\mathcal{H}$ are continuous functions; we make the following single-crossing assumptions about them:

A1 (Single crossing of $\mathcal{H}$ in z) At any point $(z; x)$ such that $\mathcal{H}(z; x) = 0$, $\mathcal{H}(z; x)$ is a strictly increasing function of z, unless $z \geq x - \beta_L$ (in which case it stays zero as z increases).

A2 (Single crossing of $\mathcal{H}$ in x) At any point $(z; x)$ such that $\mathcal{H}(z; x) = 0$, $\mathcal{H}(z; x)$ is a strictly decreasing function of x, unless $z > x - \beta_L$ (in which case it stays zero as x increases).

A3 (Single crossing of $\mathcal{K}$ in z) At any point $(z; x)$ such that $\mathcal{K}(z; x) = 0$, $\mathcal{K}(z; x)$ is a strictly increasing function of z, unless $z < x - \beta_H$ (in which case it stays zero as z increases).

A4 (Single crossing of $\mathcal{K}$ in x) At any point $(z; x)$ such that $\mathcal{K}(z; x) = 0$, $\mathcal{H}(z; x)$ is a strictly decreasing function of x, unless $z \leq x - \beta_H$ (in which case it stays zero as x increases).

Assumptions A1 and A3 are fairly weak. Assumptions A2 and A4 are more restrictive. A1-A4 impose monotonic structure on the problem. All assumptions are satisfied if ω and β are independent, uniformly distributed, random variables.

We are now ready to describe formally the equilibrium.

Proposition 9 *Under assumptions A1-A4, there exists an alternative x^* such that: (1) If $x < x^*$, then all expert types (ω, β) in the diagonal-vertical set $DV(\rho(x), x)$ induce the decision maker to choose alternative x, $\alpha(\omega, \beta) = x$ if and only if $(\omega, \beta) \in DV(\rho(x), x)$; (2) If $x > x^*$, then all expert types (ω, β) in the vertical-diagonal set $VD(\lambda(x), x)$ induce the decision maker to choose alternative x, $\alpha(\omega, \beta) = x$ if and only if $(\omega, \beta) \in VD(\lambda(x), x)$; (3) Outcome x^* is the choice induced by all the remaining expert types, $\alpha(\omega, \beta) = x^*$ if and only if $(\omega, \beta) \notin \cup_{x<x^*} DV(\rho(x), x)$ and $(\omega, \beta) \notin \cup_{x>x^*} VD(\lambda(x), x)$.*

7 Conclusions

We have introduced uncertainty in the direction of the expert's bias in the standard model of disclosure of verifiable information. With an uncertain bias, information unravelling fails

and full disclosure is not an equilibrium. Experts with positive bias observing a low state pool with experts with a negative bias observing a high state. Manipulation is pervasive and a positive measure of expert types are able to obtain what they want (i.e., their first best outcome).

An increase in the familiarity with the problem helps the decision maker. More interestingly, the size of an expert's bias is less important to a decision maker than knowledge of the bias direction. Experts that are known to be strongly biased in one direction can be easily read and their disclosures decoded. Thus, experts should strive to be poker faced. They should try to conceal which way they would like to push the outcome.

A number of promising extensions of the basic idea in this paper are worth pursuing in future research. First, the value of keeping a poker face may shape the composition of expert partnerships and explain, for example, the value of diversity. If partners come from diverse backgrounds and experiences, it will naturally be more difficult for a client to discern the direction of the organization's bias. As a suggestive example, it is broadly consistent with the theory in this paper that in 2003 the first Bush adminstration sent Colin Powell, and not Dick Cheney, to present to the United Nations Security Council evidence about the existence of weapons of mass destruction in Saddam's Iraq.

Second, when the decision maker is an active agent (and not the market, or a mass of customers) then it is often possible to seek a second option. Intuitively, a second opinion ought to be valuable, even with a priori identical experts, because there is always a chance the new expert has a different bias. Indeed, if it is known that the decision maker will seek a second opinion, the first expert will want to change its disclosure policy. More careful analysis is needed to understand what would happen. For example, it is not clear whether the decision maker should consult two different experts simultaneously or sequentially. It is also not clear whether the decision maker should disclose the information revealed by an expert she has previously visited.

Third, manipulation and in particular market manipulation, often involves taking actions, as opposed to disclosing information. For example, transactions which create an artificial security price are regarded as market manipulation by the Securities Exchange Act of 1934. In general, taking actions involves both elements of pure information disclo-

sure and elements of costly signaling. The difference between the two hinges on what is observable to the market and on the cost of divulging the information. The basic insight of this paper ought to go through, however. If the market is not certain whether the manipulator is trying to push price up or down, market manipulation is likely to have a good chance of succeeding, at least partially.

Finally, it is important to study ways for decision makers to improve reliability of the disclosure process. In this regard, it is instructive to look at the measures taken by the editors of several top medical journals. Because data manipulation, or "fudging the data", is thought to be common, they have decided to stop publishing drug research sponsored by pharmaceutical companies unless the research was registered from its beginning in a public database.[5] In essence, this is a way to put constraints on the diclosures available to the companies.

[5] Note that while it is perfectly sensible to be skeptical of sponsored research, it is not obvious that the individual researcher should be always assumed to have no integrity. Thus, it is reasonable to say that the exact direction of the bias is at least to some extent unknown.

Appendix

In this appendix, we provide the missing proofs of the results in the main body of the paper.

Proof of Lemma 2. (i) Type (ω, β) can always mimic type $\left(\omega, \beta'\right)$ by disclosing the same interval. Hence it must be:

$$-\left[\alpha\left(\omega, \beta\right)-\omega-\beta\right]^2 \geq -\left[\alpha\left(\omega, \beta'\right)-\omega-\beta\right]^2,$$

or

$$2\left(\omega+\beta\right)\alpha\left(\omega, \beta\right)-\alpha\left(\omega, \beta\right)^2 \geq 2\left(\omega+\beta\right)\alpha\left(\omega, \beta'\right)-\alpha\left(\omega, \beta'\right)^2,$$

which can be written as

$$2\left(\omega+\beta\right)\left[\alpha\left(\omega, \beta\right)-\alpha\left(\omega, \beta'\right)\right] \geq \alpha\left(\omega, \beta\right)^2-\alpha\left(\omega, \beta'\right)^2. \tag{7}$$

Similarly, since type $\left(\omega, \beta'\right)$ can mimic type (ω, β)

$$2\left(\omega+\beta'\right)\left[\alpha\left(\omega, \beta'\right)-\alpha\left(\omega, \beta\right)\right] \geq \alpha\left(\omega, \beta'\right)^2-\alpha\left(\omega, \beta\right)^2 \tag{8}$$

and hence

$$\left(\beta-\beta'\right)\left[\alpha\left(\omega, \beta\right)-\alpha\left(\omega, \beta'\right)\right] \geq 0.$$

This shows that $\alpha(\omega, \beta)$ is weakly increasing in β and hence continuous a.e.

(ii) Suppose β is a point at which $\alpha(\omega, \beta)$ is discontinuous both on the left and the right. Then, it is $\lim_{\beta_1 \uparrow \beta} \alpha\left(\omega, \beta_1\right) < \alpha(\omega, \beta) < \lim_{\beta_2 \downarrow \beta} \alpha\left(\omega, \beta_2\right)$. By (7) and (8) we have

$$2\left(\omega+\beta\right)\left[\alpha\left(\omega, \beta\right)-\alpha\left(\omega, \beta_j\right)\right] \geq \alpha\left(\omega, \beta\right)^2-\alpha\left(\omega, \beta_j\right)^2 \geq 2\left(\omega+\beta_j\right)\left[\alpha\left(\omega, \beta\right)-\alpha\left(\omega, \beta_j\right)\right]. \tag{9}$$

Taking limits, letting β_1 and β_2 converge to β and applying the sandwich theorem for sequences yields

$$2\left(\omega+\beta\right)=\alpha\left(\omega, \beta\right)+\lim_{\beta_1 \uparrow \beta} \alpha\left(\omega, \beta_1\right)=\alpha\left(\omega, \beta\right)+\lim_{\beta_2 \downarrow \beta} \alpha\left(\omega, \beta_2\right),$$

a contradiction.

(iii) Suppose $\alpha(\omega, \beta)$ is continuous in the interval $[\beta_1, \beta]$. First note that the inequalities in (9) are satisfied as equalities if $\alpha(\omega, \beta) = \alpha(\omega, \beta_1)$. Suppose $\alpha(\omega, \beta) \neq \alpha(\omega, \beta_1)$. By (1) it is $\alpha(\omega, \beta) > \alpha(\omega, \beta_1)$, and we can write (9) as

$$2(\omega + \beta) \geq \alpha(\omega, \beta) + \alpha(\omega, \beta_1) \geq 2(\omega + \beta_1).$$

Taking limits as β_1 converges to β, we obtain $\alpha(\omega, \beta) = \omega + \beta$. The argument when $\alpha(\omega, \beta)$ is continuous in the interval $[\beta, \beta_2]$ is analogous.

(iv) Suppose to the contrary that $\alpha(\omega, \beta)$ is discontinuous at some type $(\omega, 0)$. Then, there exists a sequence $\{\omega_n^0, \beta_n^0\}_{n=1}^{\infty} \to (\omega, 0)$ such that $\lim_{n\to\infty} \alpha(\omega_n^0, \beta_n^0) = x \neq \omega$. We can choose this sequence so that $\beta_n^0 \neq 0$ for all n. Suppose $x < \omega$. (Note that this implies that it must be $\omega > x \geq 0$.) Consider the sequence $\{\omega_n^1, \beta_n^1\}_{n=1}^{\infty}$ defined by $\omega_n^1 = \omega_n^0$, and $\beta_n^1 = -|\beta_n^0| < 0$. By the monotonicity of α in β, it is $\lim_{n\to\infty} \alpha(\omega_n^1, \beta_n^1) = y \leq x < \omega$. By construction, it is $\lim_{n\to\infty} \left[\omega_n^1 + 2\beta_n^1 - \alpha(\omega_n^1, \beta_n^1)\right] = \omega - y > 0$. Hence, there must exist N such that for all $n > N$ it is $\omega_n^1 + 2\beta_n^1 - \alpha\left(\omega_n^1, \beta_n^1\right) > 0$. Now observe that EX's type $\left(\omega_n^1, \beta_n^1\right)$ may disclose the interval $\left[\omega_n^1, \omega_n^1\right]$ and induce the choice of outcome ω_n^1. For $n > N$ such a disclosure is a profitable deviation, since $0 < \omega_n^1 - \left(\omega_n^1 + \beta_n^1\right) = -\beta_n^1 < \left(\omega_n^1 + \beta_n^1\right) - \alpha\left(\omega_n^1, \beta_n^1\right)$. This shows that it cannot be $\lim_{n\to\infty} \alpha(\omega_n, \beta_n) = x < \omega$; the proof that it cannot be $\lim_{n\to\infty} \alpha(\omega_n, \beta_n) = x > \omega$ is analogous. ■

Proof of Lemma 3. (i)-(iii) By Lemma 2, either $\alpha(\omega, \beta)$ is constant or is equal to $\omega + \beta$. Suppose $\alpha(\omega, \beta)$ is constant in an interval $\left[\beta', \beta''\right]$ with $\beta_L < \beta' < \beta'' < \beta_H$, while it is increasing to the left of β' and to the right of β''. By Lemma 2, it must be $\alpha(\omega, \beta') = \omega + \beta'$ and $\alpha(\omega, \beta'') = \omega + \beta'' = \omega + \beta'$. This contradicts the assumption that $\alpha(\omega, \beta)$ is constant in the interval $\left[\beta', \beta''\right]$. Thus, it can only be constant in an interval including one of the endpoints of the bias support.

(iv) For $\beta = 0$ the statement is true, since $\alpha(\omega, 0) = \omega$ for all ω. Fix $\beta > 0$. If $\alpha(\omega, \beta) = \omega$, then $\alpha(\omega, \beta)$ is strictly increasing to the right of ω, since it must be $\alpha(\omega', \beta) \geq \omega'$ for all $\omega' > \omega$. Suppose $\alpha(\omega, \beta) = x > \omega$. Then, type (ω, β) must disclose an interval $[a, b]$ with $a \leq \omega$ and $b \geq x$. Hence, type (ω', β) with $\omega < \omega' < b$ could mimic (disclose the same

interval as) type (ω,β). For mimicking not to be profitable it must be

$$\begin{aligned}\left[\omega'+\beta-\alpha(\omega',\beta)\right]^2 &\leq \left[\omega'+\beta-x\right]^2, \text{ or}\\ 2\left(\omega'+\beta\right)\left[x-\alpha(\omega',\beta)\right] &\leq \left[x+\alpha(\omega',\beta)\right]\left[x-\alpha(\omega',\beta)\right], \text{ or}\\ x &\leq \alpha(\omega',\beta),\end{aligned}$$

where the third inequality holds because, by $\beta>0$ and Lemma 3, $\omega'+\beta>\omega+\beta\geq\alpha(\omega,\beta)=x$ and $\omega'+\beta\geq\alpha(\omega',\beta)$, which implies $2\left(\omega'+\beta\right)\geq\left[x+\alpha(\omega',\beta)\right]>0$. This shows that for $\beta>0$ the map $\alpha(\omega,\beta)$ is increasing in ω. The proof for $\beta<0$ is analogous. ■

Proof of Proposition 5. We claim that the following equilibrium strategies implement the equilibrium outcome map:

$$\sigma_{EX}(\omega,\theta)=\begin{cases}[\omega+\beta-\beta^*,\omega+\beta-\beta_L] & \text{if} \quad \beta\leq\beta^*\\ \\ [\omega,\omega+\beta^*-\beta_L] & \text{if} \quad \beta>\beta^*\end{cases}$$

$$\sigma_{DM}([x,y])=\min\{x+\beta^*,y\}.$$

Given the decision strategy $\sigma_{DM}(\cdot)$ of the DM, all expert types $(\omega,\beta)\in\mathcal{T}$ with $\beta\leq\beta^*$ obtain their first best choice by using σ_{EX}. Types $(\omega,\beta)\in\mathcal{T}$ with $\beta>\beta^*$ must sent a message $[x,y]$ with $x\leq\omega$. Since their ideal choice is $\omega+\beta>\omega+\beta^*$, it is a best response for them to send a message $[\omega,y]$, as the message prescribed by strategy σ_{EX}.

Given the expert's message strategy σ_{EX} and any equilibrium message $[x,x+\beta^*-\beta_L]$, the DM computes the expected value of the state of the world to be:

$$\begin{aligned}&E[\omega|[x,x+\beta^*-\beta_L]]\\ =\;& x\frac{(\beta_H-\beta^*)}{(\beta_H-\beta^*)+(-\beta_L+\beta^*)\sqrt{2}}+\frac{2x-\beta_L+\beta^*}{2}\cdot\frac{(-\beta_L+\beta^*)\sqrt{2}}{(\beta_H-\beta^*)+(-\beta_L+\beta^*)\sqrt{2}}\\ =\;& \frac{x\left(\beta_H-\beta^*\right)\sqrt{2}+\left(2x-\beta_L+\beta^*\right)\left(-\beta_L+\beta^*\right)}{\sqrt{2}\left(\beta_H-\beta^*\right)+2\left(-\beta_L+\beta^*\right)}\\ =\;& x+\frac{\left(-\beta_L+\beta^*\right)^2}{\sqrt{2}\left(\beta_H-\beta^*\right)+2\left(-\beta_L+\beta^*\right)}.\end{aligned}$$

Thus,

$$E[\omega|[x,x+\beta^*-\beta_L]]=x+\beta^*$$

if and only if

$$\beta^* = \frac{(-\beta_L + \beta^*)^2}{\sqrt{2}\,(\beta_H - \beta^*) + 2\,(-\beta_L + \beta^*)},$$

or, equivalently,

$$\sqrt{2}\,(\beta_H - \beta^*)\,\beta^* + 2\,(-\beta_L + \beta^*)\,\beta^* = (-\beta_L + \beta^*)^2\,,$$

or,

$$\sqrt{2}\,(\beta_H - \beta^*)\,\beta^* + (-\beta_L + \beta^*)\,(\beta^* + \beta_L) = 0,$$

or,

$$\sqrt{2}\,(\beta_H - \beta^*)\,\beta^* + \beta^{*2} - \beta_L^2 = 0,$$

or,

$$\left(\sqrt{2} - 2\right)\beta^{*2} + 2\beta_H\beta^* - \sqrt{2}\beta_L^2 = 0,$$

which gives

$$\beta^* = \frac{\beta_H - \sqrt{\beta_H^2 - \left(2\sqrt{2} - 2\right)\beta_L^2}}{2 - \sqrt{2}}.$$

It follows that it is a best response for DM to follow strategy σ_{DM}. ■

Proof of Proposition 9. We begin by showing that the function $\rho(x)$ is well defined and increasing in an interval $[0, x_H]$ and that the function $\lambda(x)$ is well defined and increasing in an interval $[x_K, 1]$. Then we show that it is $x_K \leq x_H$.

Consider the function $\mathcal{K}$ first. Note that $\mathcal{K}(0; x) < 0$ for x sufficiently close to 1 and $\mathcal{K}(0; 0) > 0$. By continuity of $\mathcal{K}$ there exists x such that $\mathcal{K}(0; x) = 0$. Let x_K be the smallest such value; then $\lambda(x_K)$ as defined in (6) exists. By A4, it is $\mathcal{K}(0; x) \leq 0$ for $x \geq x_K$; since $\mathcal{K}(x; x) > 0$ and $\mathcal{K}$ is continuous, $\lambda(x)$ exists for all $x \geq x_K$. A3 and A4 imply that $\lambda(x)$ is strictly increasing in x.

Now consider the function $\mathcal{H}$. $\mathcal{H}(1; x) > 0$ for x sufficiently close to 0 and $\mathcal{H}(1; 1) < 0$. By continuity of $\mathcal{H}$ there exists x such that $\mathcal{H}(1; x) = 0$. Let x_H be the largest such value; then $\rho(x_H)$ as defined in (5) exists. By A2, it is $\mathcal{H}(1; x) \geq 0$ for $x \leq x_H$; since $\mathcal{H}(x; x) < 0$ and $\mathcal{H}$ is continuous, $\rho(x)$ exists for all $x \leq x_H$. A1 and A2 imply that $\rho(x)$ is strictly increasing in x.

Now observe that it must be $x_K \leq x_H$. This is because $\mathcal{H}(1;x) \geq \mathcal{K}(0;x)$ for all x. In particular, $\mathcal{H}(1;x_K) \geq \mathcal{K}(0;x_K) = 0$. Hence $\rho(x_K)$ exists (since $\mathcal{H}(x_K;x_K) < 0$). There are two possible cases. Case 1: $x_K < x_H$ and $\mathcal{H}(1;x_K) > \mathcal{K}(0;x_K) = 0$. Case 2: $\mathcal{H}(1;x) = \mathcal{K}(0;x) = 0$ for all $x \in [x_K, x_H]$ and $x_K \leq x_H$. Because it is less cumbersome to describe, in both cases we will construct an equilibrium in which all types on the same iso-decision set disclose the same interval. There are other outcome equivalent, and perhaps more plausible, equilibria (e.g., there is an equilibrium in which all types with a small bias on an iso-decision set pool with their "twin" and a set of types with large positive bias pool with a set of types with large negative bias).

Consider Case 2 first. The equilibrium we construct is depicted in Figure 6. Note that in this case for all $x \in [x_K, x_H]$ it is $DV(\rho(x), x) = VD(\lambda(x), x)$, $\rho(x) = x - \beta_L$ and $\lambda(x) = x - \beta_H$. Let ρ^{-1} and λ^{-1} be the inverses of the functions ρ and λ. Take any $x^* \in [x_K, x_H]$ and define the following strategy by DM:

$$\sigma_{DM}([a,b]) = \begin{cases} \max\left\{\rho^{-1}(b), a\right\} & \text{if} \quad b \leq x^* - \beta_L \\ \min\left\{\lambda^{-1}(a), b\right\} & \text{if} \quad b \geq x^* - \beta_L \end{cases}$$

Such a strategy is optimal for DM if for all $x \leq x^*$ types that belong to the set $DV(\rho(x), x)$ disclose the interval $[\max\{0, x - \beta_H\}, \rho(x)]$ and for all $x > x^*$ types that belong to the set $VD(\lambda(x), x)$ disclose the interval $[\lambda(x), \min\{1, x - \beta_L\}]$. Given such a DM's strategy, it is optimal for all types on $DV(\rho(x), x)$ with $x \leq x^*$ to disclose $[\max\{0, x - \beta_H\}, \rho(x)]$ and for all types on $VD(\lambda(x), x)$ with $x > x^*$ to disclose $[\lambda(x), \min\{1, x - \beta_L\}]$. To see this, note that the only types on $DV(\rho(x), x)$ that do not fully manipulate the outcome have observed state $\rho(x)$ and their ideal outcome is smaller than x. They have no profitable deviation, since any disclosure that is feasible to them must include $\rho(x)$ and hence, given DM's strategy, leads to an outcome at least as large as x. Similarly, the only types on $VD(\lambda(x), x)$ that do not fully manipulate the outcome have observed state $\lambda(x)$ and their ideal outcome is higher than x. However, any feasible disclosure must include $\lambda(x)$ and leads to an outcome at least as small as x.

Now consider Case 1. The equilibrium we construct is depicted in figures 4 and 5. First,

we define x^* implicitly as the solution of the following equation

$$\int_{\max\{0,x^*-\beta_H\}}^{\lambda(x^*)} \int_{x^*-\omega}^{\beta_H} (x^* - \omega) f(\omega) g(\beta|\omega) d\beta d\omega = \int_{\rho(x^*)}^{\min\{1,x^*-\beta_L\}} \int_{\beta_L}^{x^*-\omega} (\omega - x^*) f(\omega) g(\beta|\omega) d\beta d\omega. \tag{10}$$

Note that for $x^* = x_K$, $\lambda(x^*) = \max\{0, x^* - \beta_H\}$ and the left hand side is zero while the right hand side is positive. For $x^* = x_H$ the left hand side is positive while the right hand side is zero, since $\rho(x^*) = \min\{1, x^* - \beta_L\}$. Since both sides are continuous in x^*, a solution exists. Note that the expression on the left hand side is proportional to the expected difference between x^* and ω belonging to the upper shaded area in Figures 4 and 5, while the expression on the right hand side is proportional to the expected difference between ω belonging to the lower shaded area and x^*. EX's types in these two areas pool and disclose the same interval. Define DM's strategy as follows:

$$\sigma_{DM}([a,b]) = \begin{cases} \max\left\{\rho^{-1}(b), a\right\} & \text{if} \quad b < \rho^{-1}(x^*) \\ \max\{x^*, a\} & \text{if} \quad \rho^{-1}(x^*) \le b \le x^* - \beta_L \\ \min\left\{\lambda^{-1}(a), b\right\} & \text{if} \quad b > x^* - \beta_L \end{cases}$$

The EX's strategy is the following: for all $x < x^*$ types that belong to the set $DV(\rho(x), x)$ disclose the interval $[\max\{0, x - \beta_H\}, \rho(x)]$; for all $x > x^*$ types that belong to the set $VD(\lambda(x), x)$ disclose the interval $[\lambda(x), \min\{1, x - \beta_L\}]$; all other types disclose the interval $[\max\{0, x^* - \beta_H\}, \min\{1, x^* - \beta_L\}]$. The proof that it is optimal for the DM and EX to follow these strategies is analogous to the proof of Case 2. The only additional steps that are needed are the following. First, by (10) the DM's expected value of the state for the set of types that disclose interval $[\max\{0, x^* - \beta_H\}, \min\{1, x^* - \beta_L\}]$ is x^*. Second, EX types in this set have no profitable deviation from inducing choice x^*. The types with a positive bias would prefer a higher alternative, but all their feasible disclosures $[a, b]$ have $a < x^*$. The types with a negative bias would prefer a lower alternative, but all their feasible disclosures $[a, b]$ have $b \ge \rho^{-1}(x^*)$. ■

References

[1] Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50(6): 1431-1451.

[2] Dimitrakas, Vassilios and Yianis Sarafidis. 2005. "Advice from an Expert with Unknown Motives." Working Paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=797844

[3] Dziuda, Wioletta. 2010. "Strategic Argumentation." Working Paper, Kellogg School of Management, Northwestern University.

[4] Grossman, Sanford. 1981. "The Informational Role of Warranties and Private Disclosure about Product Quality." Journal of Law and Economics, 24(3): 461–83.

[5] P. J. Hammond, Peter J. and Yeneng Sun. 2003. " Monte Carlo Simulation of Macroeconomic Risk with a Continuum of Agents: The Symmetric Case." *Economic Theory*, 21: 743-766.

[6] Kamenica, Emir, and Matthew Gentzkow. 2009. "Bayesian Persuasion." Working Paper no. 15540, NBER, Cambridge, MA.

[7] Li, Ming, and Kristóf Madarász. 2008. "When Mandatory Disclosure Hurts: Expert Advice and Conflicting Interests." *Journal of Economic Theory*, 139(1): 47-74.

[8] Milgrom, Paul. 1981. "Good News and Bad News: Representation Theorems and Applications." *Bell Journal of Economics*, 12(2): 380–91.

[9] Rayo, Luis, and Ilya Segal. 2010. "Optimal Information Disclosure." *Journal of Political Economy*, 118(5):

[10] Seidmann, Daniel and Eyal Winter. 1997. "Strategic Information with Verifiable Messages." *Econometrica*, 65: 163-169.

[11] Shin, Hyun Song. 1994. "News Management and the Value of Firms." *Rand Journal of Economics*, 25(1): 58–71.

[12] Wolinsky, Asher. 2003. "Information Transmission when the Sender's Preferences are Uncertain." *Games and Economic Behavior*, 42(2): 319-326.