Well Informed Intermediaries in Strategic Communication*

Wei Li†

May, 2009

Abstract

This paper studies how a sender with private information can influence the decisionmaker through well informed intermediaries. Both the sender and the intermediary may be independently objective or biased: with the objective type assumed to pass on the most accurate information while the biased type wanting to push a particular agenda but also to appear objective. Although using one’s own information is a sign of objectivity, the biased intermediary selectively incorporates the sender’s information to push his agenda, and his truth-telling incentives always decrease in those of the biased sender’s. Hence measures raising the sender’s reputation cost worsen the intermediary’s distortion and may make the decisionmaker strictly worse off. In contrast, the biased sender’s and the intermediary’s truth-telling incentives are strategic complements if they report simultaneously.

JEL classification: C70, D82, M31, M38

Keywords: strategic communication through intermediaries, indirect marketing, reputational concerns.

---

*I am grateful for the insightful comments from Glenn Ellison, Hao Li, R. Preston McAfee, Marco Ottaviani, Edward Schlee, Jean Tirole, and the seminar participants at ASU, UCR, Fuqua School of Business, Duke University, Johns Hopkins University, Marshall School of Business, UBC, USC, the 2007 Duke/Northwestern/Texas IO conference, and the 83th Western Economic Association Annual Conference.

†University of California, Riverside. Email: wei.li@ucr.edu. Phone: 951-827-7809. Fax: 951-827-5685.
I. INTRODUCTION

Recommendations from well informed experts are often important to the final users of such information. In health care, for example, patients often follow the prescriptions and other advice from their physicians. Because of the importance of experts’ recommendations, it is increasingly often for businesses to attempt to influence their customers through expert opinion. Some businesses even pay experts or bloggers to promote their image or products while pretending to be independent reviewers.\(^1\) In the medical and health care industry, serious questions have been raised concerning pharmaceutical companies who promote their drugs through physicians and medical researchers.\(^2\) These companies exert influence through providing physicians with the company’s own information about their products (“detailing”) and other gifts; through funding research and giving perks such as paid consultant positions; and through “ghostwriting” of journal articles where the purported academic authors have done little of the actual research.\(^3\)

These influence activities have become more prevalent in recent years. It is estimated that drug companies spend approximately $19 billion annually for marketing to doctors.\(^4\) In particular, the amount of money spent “detailing” physicians has increased from $3 to $4.8 billion from 1996 to 2000.\(^5\) Nearly 75% of physicians in a national poll said the information they received from pharmaceutical representatives was “very” useful (15%) or “somewhat” useful (59%).\(^6\) Studies such as Avorn, Chen, and Hartley (1982) and Watkins, Moore, Harvey, Carthy, Robinson, and Brawn (2003) also indicate that these influence activities are effective in changing physicians’ prescription behaviors, and consequently affecting patients’ welfare.\(^7\)

---


\(^3\) One recent example concerns an Annals of Internal Medicine article on Merck’s “Advantage” trial of Vioxx, which omitted some trial participants’ deaths. The article’s first author Jeffrey Lisse, a rheumatologist at the University of Arizona, said that Merck actually wrote the report, and that “Merck designed the trial, paid for the trial, ran the trial.....Merck came to me after the study was completed.” New York Times. “Evidence in Vioxx Suits Shows Intervention by Merck Officials”, 24 April 2005.

\(^4\) For details, please see “Health industry practices that create conflicts of interest: a policy proposal for academic medical centers,” published in the Journal of the American Medical Association (Brennan, Rothman, Blank, Blumenthal, Chimonas, Cohen, Golden, Kassirer, Kimball, Naughton, and Smelser 2006).

\(^5\) See IMS Health Inc. and Competitive Media Reporting, as reported by Kaiser Family Foundation, 2001.

\(^6\) More than half (55%) of a group of “high-prescribing” doctors surveyed by the industry data tracking group ImpactRx said that drug representatives serve as their primary source of information about newly approved drugs. See “Getting Doctors to Say Yes to Drugs: The Cost and Quality Impact of Drug Company Marketing to Physicians” by the BlueCross BlueShield Association. See http://www.bcbs.com/betterknowledge/cost/getting-doctors-to-say-yes.html.

\(^7\) Another example concerns General Motors Corp., which found itself spending $52 million in 2001 for prescriptions that physicians wrote for Prilosec, even though a subsequent analysis found that 91% of the patients receiving prescriptions had no prior diagnosis of the problem.
This paper presents a model of strategic communication through a well informed intermediary and examines its effect on the final users of such information. As is often the case in medical and health care industry, the final user only has access to the intermediary’s recommendations, who may have been influenced by a sender. This differs from many existing papers analyzing how a sender influences the final user directly by manipulating the information he sends. Consider the medical example above, where a pharmaceutical company may promote its drugs through physicians. How are the physicians influenced by the company, especially if their information disagree? How do the pharmaceutical company’s truth-telling incentives interact with those of the physician’s? What is the net impact on the patients? This paper addresses these questions by investigating the strategic interactions of the sender and the intermediary. It then applies these insights to study the effectiveness of policy measures aiming at improving reporting accuracy, and the potential implications for disclosure laws and professional ethical rules.

In this paper, a sender receives a private signal about the state of the world and sends a message to an intermediary who also has an independent, private signal. The intermediary in turn sends a recommendation to the decisionmaker. Because this paper focuses on a well informed intermediary who has expertise such as physicians; or has experience in a market for credence goods, the intermediary’s signal is assumed to be more accurate than the sender’s. The decisionmaker, who does not observe the sender’s message, takes an action based solely on the intermediary’s recommendation. Next, the true state becomes observable, and the decisionmaker forms her opinion of how truthful the sender and the intermediary have been. The sender and the intermediary may be independently one of two types: objective or biased. An objective agent is assumed to pass on the most accurate information he has, while a biased agent, who has reputational concerns, wants a particular action taken. Thus a biased agent faces a tradeoff between reporting truthfully to appear objective and distorting to induce the decisionmaker to take the action he prefers.

In equilibrium, both biased sender and biased intermediary are shown to report truthfully if their information supports their bias, but distorts with some positive probability otherwise. The first result of this paper is that the more truthful a biased sender is, the more a biased intermediary distorts. Consequently, small changes in the

---

8 See for example, Crawford and Sobel (1982), Austen-Smith (1990), Dewatripont and Tirole (1999), Morris (2001) and Ottaviani and Sorensen (2006a).
9 In a companion piece, Li (2009) considers the case when the intermediary has little information of his own and acts as a pure intermediary. The strategic interactions between the intermediary and the sender, as well as policy implications, differ significantly due to the lack of information aggregation considerations. This is further discussed in Section V.
10 Throughout this paper, the decisionmaker is female.
sender’s reporting accuracy — whose signal is relatively uninformative — may have a strong impact on the biased intermediary’s reporting accuracy. The reason that the biased intermediary’s truth-telling incentives always decrease in those of the sender’s is that, on one hand, the sender’s message affects the biased intermediary’s benefit from distortion, which is the change in the decisionmaker’s action induced by his recommendations. If the sender becomes more truthful, then the biased intermediary’s recommendation becomes more credible. Thus the decisionmaker’s action responds more to the intermediary’s recommendation, increasing his incentive to lie. On the other hand, the biased intermediary is concerned about any reputation cost he incurs if he distorts. A more truthful message from the sender changes how biased intermediary’s reputation depends on his recommendation through a reputation sensitivity effect. Since the objective intermediary always reports his signal, independence from the sender’s influence is the sign of objectivity: the intermediary cannot blame any mistake on being misled by the wrong source. If the sender’s message becomes more truthful, the decisionmaker assigns the intermediary less credit for a correct recommendation because he may have followed the more accurate message; and less blame for a wrong recommendation because any mistake is more likely due to a wrong signal. Since the biased intermediary has more influence on the decisionmaker at a lower reputation cost, he distorts more if the sender reports more truthfully.

A biased sender never reports truthfully if he can influence the intermediary. Moreover, when the intermediary has high reputational concerns, his own reputational cost from distortion decreases in the intermediary’s truth telling, because of a similar reputation sensitivity effect. In this case, the biased agents’ truth-telling incentives become strategic substitutes: if one reports more truthfully, the other distorts more. In particular, a biased sender may distort completely, regardless of how high his reputational concerns are, if the biased intermediary is very concerned about his reputation.

The result that the biased intermediary may distort more if the sender reports more truthfully implies that the decisionmaker may value experts who base their recommendations solely on their own signal. Moreover, she may consider policy measures to increase the reputational cost of the biased agents to improve truth telling. The second result of this paper is that not all such measures are equal: making it more costly for the sender to lie, for instance by strengthening regulations over the pharmaceutical industry, may make the decisionmaker strictly worse off. This occurs when both biased agents have low levels of reputational concerns, in which case the sender’s message is not credible and unlikely to be followed by the intermediary. If the biased sender reports more truthfully under the new policy, the biased intermediary lies more against his own signal. This
means that the decisionmaker may receive a recommendation against both agents’ true signals with a higher probability, and thus she makes worse decisions. She should target the intermediary instead, for instance by strengthening medical board review process and monitoring disclosure of industry ties of the physicians and researchers. The resulting gain in truth telling from the intermediary, whose signal is more accurate, outweighs any indirect effect on the sender.

In comparison, simultaneous reporting is considered where the decisionmaker receives two messages, one from each agent, before taking an action. Here each biased agent pays his own reputation cost because he is evaluated based on his message and the later observed true state. Therefore the presence of another message only affects a biased agent’s (expected) distortion benefit, which is shown to decrease in the other agent’s truth telling. Intuitively, if one biased agent reports more truthfully, it increases the probability that the other agent’s distorted message is contradicted, in which case the decisionmaker is less likely to change her action than if there is a concurring message. Overall, distortion becomes less effective while one’s reputation cost is unaffected, thus each biased agent lies less. The third result of this paper is that biased agents’ truth-telling probabilities are strategic complements under simultaneous reporting. Thus the decisionmaker may want to encourage simultaneous reporting, especially if both agents have high levels of reputational concerns. In this case, any policy measure targeting either the industry or the physicians has a strictly positive effect.11

The present paper focuses on how a sender can influence an intermediary by providing information, which alters the intermediary’s confidence in his own signal. Durbin and Iyer (2008) consider the case where intermediaries (advisors) may be bribed by an uninformed and biased third party to support its bias. They show that if the advisors have reputational concerns, a bribe may be necessary for the advisor to report his true signal if it happens to favor the biased third party. In a similar vertical structure but with a different focus, Inderst and Ottaviani (2008) study the incentive problems faced by an intermediary (such as a sales agent) if he has to perform two tasks: one for the seller and the other for the buyer. Given this inherent conflict of interest, they study the design of compensation schemes for the intermediary.

Many existing papers such as Scharfstein and Stein (1990), Prendergast (1993), Prendergast and Stole (1996), Prat (2005) and Ottaviani and Sorensen (2006b) have studied the truth-telling incentives of experts driven by reputational concerns. In the context of multiple experts making reports sequentially, Scharfstein and

---

11 Levy (2007) considers a similar case in which multiple imperfectly informed agents in a committee submit individual reports, which are publicly observed. Thus each agent’s reputation depends only on his or her own report, not the recommendation of the committee.
Stein (1990) and Levy (2004) show that the second expert may bias his report, depending on the report from the first expert, to appear smart. The early report in these papers is public information, which simply changes the prior beliefs of both the second expert and the decisionmaker. As a result, the decisionmaker (investors or other evaluators) can filter out the early report’s influence on the second expert when she evaluates the experts — and thus the truth-telling incentives of the first expert do not interact with those of the second one. In contrast, the early report in the current paper is private information, and thus the sender and the intermediary’s truth-telling incentives (and their reputation) are necessarily linked, which gives rise to the aforementioned reputation sensitivity effect. Relatedly, Li (2007) considers a setting in which one single expert receives private signals of increasing precision and gives multiple reports. In her model, the sequencing of reports becomes a signal of how fast the expert is learning, which affects his reputation. The signaling effect of sequencing does not surface here because the decisionmaker only hears from the intermediary.

In the rest of the paper, Section II sets up the basic model and Section III analyzes the equilibria behavior of the biased agents. It also considers how the decisionmaker can encourage truthful reporting. Section IV considers the case of simultaneous reporting by the agents, and Section V discusses several main assumptions and concludes. All the proofs are contained in the Appendix.

II. MODEL

There are two players in this game, agent \( A \) and \( B \), and there is a decisionmaker \( C \). The state of the world is binary: \( \eta \in \{0, 1\} \). Each state occurs with equal probability. The game proceeds in three stages: information transmission, decision making, and evaluation. In the information transmission stage, agent \( A \) first observes a private signal \( s_A \in \{0, 1\} \), which is equal to the true state with probability \( p_A > 0.5 \); otherwise it is wrong. He then sends a message \( m_A \in \{0, 1\} \) to an intermediary, agent \( B \). Agent \( B \) observes a private signal \( s_B \in \{0, 1\} \) of his own, which is equal to the true state with probability \( p_B > p_A \). Agent \( B \) then sends a message \( m_B \in \{0, 1\} \) to the decisionmaker.\(^{12} \) Agent \( A \) and \( B \)’s signals are independent conditional on the state. In the decision making stage, \( C \) chooses an action \( a \in \mathbb{R} \) given message \( m_B \). In the evaluation stage, \( C \) first observes the true state \( \eta \) and then forms posterior beliefs about each agent’s type, to be described next. In

\(^{12} \) The binary message of \( B \) is best thought of as a simple, “yes-or-no” type of recommendation, which is the simplest way to illustrate the direction of biased \( B \)’s distortions given his information. Allowing \( B \) to convey both \( A \)’s message and his own signal is discussed further in Section V.A, and in Appendix B.
all three stages, agent $B$ and decisionmaker $C$ only observe the message sent directly to him (her). Moreover, all messages in this model are assumed to be private and unverifiable.

The decisionmaker’s payoff is represented by the quadratic loss function $-(a - \eta)^2$. Her optimal action is thus to choose $a$ equal to the probability she attaches to $\eta = 1$. An agent may be either objective (type $o$) or biased (type $b$). Each agent’s type is independently drawn from $\{o, b\}$: $\Pr(A = o) = \theta_A$, and $\Pr(A = b) = 1 - \theta_A$, with $\theta_A$ referred to as $A$’s prior objectivity. Similarly, agent $B$ is objective with probability $\theta_B$. An objective agent is assumed to report the most accurate information he has, while a biased agent always wants action $a = 1$ taken, regardless of the true state. However, he also wants to appear objective. Denote $C$’s posterior belief of agent $A$ and $B$ being objective, formed at the evaluation stage, as $\pi_A$ and $\pi_B$ respectively.\(^{13}\) Biased $A$ and $B$’s payoffs are assumed to be:

$$U_A = a + \alpha \pi_A \quad \text{and} \quad U_B = a + \beta \pi_B.$$  

The first half of a biased agent’s payoff function is $C$’s action: the higher is $a$, the better off a biased agent is. The second half is a reduced-form formulation, representing a biased agent’s reputational payoffs.\(^{14}\) Parameters $\alpha, \beta \in [0, \infty)$ are the weights biased $A$ and $B$ attach to their reputation. The game is illustrated in Figure 1.

In this game, a strategy of biased $A$ consists of two probabilities of reporting truthfully, one for each signal $s_A$. Analogously, a strategy of biased $B$ consists of four probabilities of reporting truthfully, one for each combination of his own signal and $A$’s message. This paper looks for perfect Bayesian equilibrium (PBE), in which given strategies of biased $A$ and $B$, decisionmaker $C$’s action $a$ at the decision-making stage maximizes her (expected) payoff, and her posterior beliefs at the evaluation stage $\pi_A, \pi_B$ satisfy Bayes’ rule.

In the medical industry example, the true state refers to the effectiveness of a drug: it may be “useless” (state 0) or “useful” (state 1). Agent $A$ is the manufacturer of this drug and agent $B$ may be either a physician

\(^{13}\) Biased $A$ may be concerned about his reputation with agent $B$, which may lead him to report more truthfully than in the current model if such reputational concerns are sufficiently high.

\(^{14}\) Morris (2001) and Li (2007) show that an agent’s reputational payoff may be convex in his posterior objectivity, reflecting the fact that information about the agent’s type itself may have value for the decisionmaker.
who makes a recommendation to the patients, or a medical researcher who studies the effectiveness of the drug. 

$C$ may be a patient who needs to decide how much to rely on this drug, or the medical community who learns about the drug from the researcher. Both $A$ and $B$ have information about the true state, but $B$’s signal is more informative, reflecting the fact that $B$ either has experience or expertise in evaluating a product. For example, a physician is better at figuring out whether a drug is useful for his patients.

An objective agent in this model always passes on the information he believes to be the most accurate. This can be justified either on the grounds of professional ethics or institutional goals.$^{15}$ Clearly, objective $A$ reports his only piece of information truthfully: $m_A = s_A$. Within the confines of this model, objective $B$’s recommendation is also very simple even though he has two pieces of information $m_A, s_B$:

**OBSERVATION 1:** *Objective $B$ always reports his own signal regardless of $A$’s message: $m_B = s_B$.*

Because $B$’s signal is a better source of information than that of $A$’s, whether $A$’s message confirms or contradicts it, his recommendation should simply reflect his signal. More precisely, $\Pr(\eta = s_B) > \Pr(\eta = m_A)$ for any message agent $A$ sends, and thus objective $B$ does not listen to $m_A$. This suggests that sometimes, if the loss in information is low, information aggregation is less important to the decisionmaker than hearing from objective $B$, a well informed expert. This is especially so if $A$’s private information is of low quality. Therefore it is assumed, throughout this paper, that $A$’s signal is very uninformative: $p_A$ is sufficiently close to 0.5.$^{16}$

Before turning to the analysis, it is useful to keep in mind a few other applications of the model. In an application to real estate business, agent $A$ is a mortgage broker (or a bank) and the state is the value of a property. Agent $B$ is a real estate appraiser who specializes in estimating the market value of a property. Biased $A$ wants a higher appraisal so as to boost the commissions from financing the purchase while biased $B$ wants to inflate the appraisal, perhaps for future referrals. Decisionmaker $C$ is the perspective homeowner who needs to decide what to bid for the property.$^{17}$

In an application to political arena, agent $A$ is a political action committee (PAC) who may be genuinely

---

15 For instance, Lahey Clinic, one of the major U.S. adult care hospitals writes: “Because good ethics begins with good medicine, the patient must receive accurate medical information and must understand it.”

16 All the results in this paper hold if in equilibrium, biased $B$ is willing to report $m_B = 1$ if his own signal supports it (see more details in Section III.A). One sufficient, though not necessary, condition for this to occur is that $A$’s signal quality is sufficiently low.

17 Biased appraisals have become a key issue in several recent lawsuits. For instance, New York’s attorney general announced a case against eAppraiseIT, a leading appraisal management firm, for caving in to pressure from Washington Mutual to use a list of “proven appraisers” who he claims inflated home appraisals. Associated Press, 1 November 2007.
concerned about a policy or biased toward certain special interests. The state is the potential impact of this policy. Agent $B$ is a legislator overseeing this policy related area and makes a recommendation about its impact. Biased $B$ may want to push for the policy but is also concerned about his political future. Decisionmaker $C$ is the legislature who decides how much to support this policy.

III. DISTORTION EQUILIBRIUM AND DECISIONMAKER’S WELFARE

A. Biased Intermediary’s Best Response

Objective $B$ reports $m_B = s_B$, but biased $B$ chooses $m_B$ to maximize, given his information $s_B$ and $m_A$:

$$E_U(m_B|m_A, s_B) = \Pr(\eta = 1|m_B) + \beta E_\eta[\Pr(B = o|m_B, \eta)|m_A, s_B].$$

The first part is $C$’s action given $B$’s recommendation, and the second his (expected) reputation, where the expectation is taken with respect to state $\eta$.

Next, suppose that biased $A$ and biased $B$ use the following types of strategy, referred to as a distortion strategy from now on. Biased $A$ reports truthfully if his signal favors his bias, but reports truthfully with some probability if his signal is against his bias: $m_A = 1$ if $s_A = 1$, but $m_A = 0$ with probability $x$ if $s_A = 0$. Similarly, biased $B$ reports $m_B = 1$ if $s_B = 1$. If $s_B = 0$, he reports $m_B = 0$ with probability $y_2$ if $A$’s message agrees with his signal ($m_A = 0$); but reports $m_B = 0$ with probability $y_1$ if $m_A = 1$. Note that $x, y_1,$ and $y_2$, chosen later to construct an equilibrium, are all truth-telling probabilities.

Given the distortion strategy, biased $B$ strictly benefits in terms of $C$’s action if he reports $m_B = 1$ instead of $m_B = 0$: $\Pr(\eta = 1|m_B = 1) - \Pr(\eta = 1|m_B = 0) > 0$. Intuitively, even if biased $B$ lies completely, $m_B = 1$ may result from signal $s_B = 1$, but $m_B = 0$ must result from $s_B = 0$, and thus it is less likely that $\eta = 1$ if $m_B = 0$. However, biased $B$ also incurs a reputation cost,

$$\sum_\eta \Pr(\eta|m_A, s_B) [\Pr(B = o|m_B = 0, \eta) - \Pr(B = o|m_B = 1, \eta)] > 0,$$

if he reports $m_B = 1$ instead of $m_B = 0$. Intuitively, regardless of state $\eta$, $m_B = 0$ is more likely to come from objective $B$ and is thus a better sign of $B$’s objectivity. It is worth noting that because objective $B$ never follows $m_A$, not being influenced by $m_A$ is a sign of objectivity. Thus biased $B$ may have an incentive to show his recommendation is independent. In particular, given the distortion strategy, biased $B$’s posterior objectivity
that were the case, assumption that support for his bias while against two signals is more likely to lead to the worst reputation (Pr(B = o|m_B = 0, η = 0) > Pr(B = o|m_B = 0, η = 1)): he is considered more objective despite a wrong recommendation. Intuitively, if m_B = 0 but η = 1, C thinks that it is likely that m_A = 1, but B did not follow m_A. This effect is particularly strong if A’s signal is very informative, because biased B is less confident about his own signal, and may want to appear independent by reporting m_B = 0 when m_A = 0, s_B = 1. The assumption that s_A is sufficiently informative ensures that no perverse equilibrium of this type arise.

If s_B = 1, biased B always reports m_B = 1. But if s_B = 0, then biased B’s best response to any given x is to choose probabilities y_1 and y_2 to maximize his expected payoff, incorporating how C chooses her actions given m_B; and how she updates her belief about B’s objectivity in the evaluation stage.

PROPOSITION 1: For any given x, biased B has a unique pair of best response (y_1, y_2) which satisfies the following properties: (1) y_1 and y_2 cannot both be strictly between 0 and 1; (2) y_2 ≥ y_1, and the inequality is strict if max\{y_1, y_2\} > 0; (3) If y_1 ∈ (0, 1), it decreases in x; similarly, if y_2 ∈ (0, 1), it decreases in x.

Although biased B’s distortion benefit does not vary with A’s message, his reputation cost does and thus his best response. A’s message, by affecting Pr(η|m_A, s_B), may strengthen or weaken biased B’s belief in his own signal. If s_B = 0, biased B’s reputation cost if m_A = 0 is higher than that if m_A = 1, because lying against two signals is more likely to lead to the worst reputation (Pr(B = o|m_B = 1, η = 0)). Therefore biased B cannot be indifferent between reporting truthfully and distorting both when m_A = 0 and when m_A = 1, or, y_1 and y_2 cannot be both strictly between 0 and 1. Moreover, biased B is more apt to lie if A’s message agrees with, rather than contradicts, his bias: y_2 > y_1 if max\{y_1, y_2\} > 0. This is because m_A = 1 lends some support for his bias while m_A = 0 is another strike against it. If s_B = 0 and biased B is indifferent between reporting m_B = 1 and m_B = 0 when m_A = 0; he must strictly prefers reporting m_B = 1 when m_A = 1: y_1 = 0, y_2 ∈ (0, 1]. If biased B never distorts when m_A = 0, he reports m_B = 0 with some probability strictly smaller than 1 when m_A = 1: y_1 ∈ (0, 1), y_2 = 1. However, biased B is never completely truthful (y_1 < 1). If that were the case, m_B is always credible, yet biased B pays no reputation cost because C rationally attributes any mistake to a faulty signal, which is impossible.

Note that biased B’s distortion benefit increases in his truth-telling probability while his reputation cost decreases in it. Intuitively, the more truthful he is, the more credible his recommendation is. Also, his reputation depends less on his recommendations, because any mistake is less likely to result from his distortion. Hence biased B has a unique best response to any given x. He may strictly prefers reporting m_B = 1 and thus
$y_1 = 0, y_2 = 0$. Otherwise, because truthful reporting is never a best response, and because of the above monotonicity, there exists a unique probability strictly between 0 and 1 such that he is indifferent between reporting $m_B = 0$ or $m_B = 1$.

Somewhat surprisingly, it is cheaper for biased $B$ to distort if $m_A$ is more truthful. More formally, biased $B$’s truth-telling probability, if strictly between 0 and 1, decreases in $\theta_A + (1 - \theta_A)x$, the perceived probability that agent $A$ reports $m_A = 0$. One may think that biased $B$ should lie less instead: since $C$ cannot observe $A$’s message, $A$ and $B$ should share the blame if $B$’s recommendation turns out to be wrong. And if $C$ attributes less blame to $A$, whose message is more truthful, it is more costly for biased $B$ to lie. In this model, however, any sign of being influenced is a sign of bias, thus biased $B$ cannot shift any blame to $A$. In particular, if $s_B = 0$, a more truthful message from biased $A$ has two effects on biased $B$. The first is a reputation sensitivity effect: how responsive biased $B$’s reputation is to a more truthful $m_A$, holding biased $B$’s belief about $m_A$’s credibility fixed. Since $y_2 > y_1$ if $y_2 > 0$, it is more likely for biased $B$ to report $m_B = 0$ if biased $A$ reports $m_A = 0$ with a higher probability. In this way, $m_B = 0$ becomes a less positive signal of independence (and hence $B$’s objectivity). Similarly, $m_B = 1$ becomes a less negative signal of $B$’s objectivity because the chance of being influenced by $m_A = 1$ is smaller. Hence biased $B$’s reputation is less sensitive to his recommendation. The second is an information aggregation effect: how likely biased $B$ believes that $\eta = 0$ given $m_A$, holding $C$’s evaluations of $B$’s posterior objectivity fixed. If $m_A$ is more truthful, $m_A = 1$ becomes more credible. Therefore biased $B$ is more likely to follow $A$’s message and report $m_B = 1$: everything being equal, he is less likely to receive the worst reputation. Both these effects reduce biased $B$’s reputation cost of distortion, and biased $B$’s benefit from distortion increases from the information aggregation effect. As a result, biased $B$ lies more if $x$ increases.

B. Distortion Equilibrium

This subsection focuses on the strategic interactions between biased $A$ and $B$. Given their distortion strategy, biased $A$ chooses a message $m_A$ to maximize his expected payoff:

$$E_{m_B} [\Pr(\eta = 1|m_B) + \alpha\mathbb{E}_\eta [\Pr(A = o|m_B, \eta)] | s_A, m_A].$$

Note that biased $A$ takes expectation not only with respect to the true state, but also with respect to $B$’s recommendation. Biased $A$ knows that if $s_B = 1$, $B$ always reports $m_B = 1$. Thus the pivotal event for biased
A, which determines his message choice, is if biased B truthfully reports \( m_B = 0 \) when \( s_B = 0 \). Biased A’s message only matters if biased B changes his report because of \( m_A \). Let \( \kappa \equiv (1 - \theta_B)(y_2 - y_1) \). If \( s_A = 0 \), then the benefit from distortion if biased A reports \( m_A = 1 \) instead of \( m_A = 0 \) is:

\[
\kappa \Pr(s_B = 0|s_A = 0)[\Pr(\eta = 1|m_B = 1) - \Pr(\eta = 1|m_B = 0)],
\]

which is biased B’s distortion benefit, multiplied by the probability that biased B changes \( m_B \) because of \( m_A \). Intuitively, if B is known to be objective (\( \theta_B = 1 \)); or if biased B always distorts (\( y_1 = y_2 = 0 \), \( m_A \) has no impact on C’s action (\( \kappa = 0 \)). Moreover, the larger is the gap between \( y_2 \) and \( y_1 \), the stronger is A’s (indirect) influence on C’s action. Similarly, if \( s_A = 0 \), biased A’s reputation cost if he distorts is:

\[
\kappa \Pr(s_B = 0|s_A = 0) \sum \Pr(\eta|s_A = 0, s_B = 0)[\Pr(A = o|m_B = 0, \eta) - \Pr(A = o|m_B = 1, \eta)],
\]

which only depends on the event that both their signals are against their bias. Otherwise, biased A’s reputation does not depend on his own message.

A distortion equilibrium is one in which both biased agents use distortion strategies. Since a distortion equilibrium consists of three probabilities: \( x, y_1, y_2 \), biased A’s best response to biased B’s truth telling depends on both \( y_1 \) and \( y_2 \); and biased B’s best response \((y_1, y_2)\) depends on biased A’s truth-telling probability \( x \). The next result summarizes the biased agents’ equilibrium behavior.

**Proposition 2:** Every equilibrium is a distortion equilibrium. Further, there exist cutoff values \( \beta^c, \beta^s, \beta^w \) satisfying \( \beta^c < \beta^s < \beta^w \) such that: (1) a complete distortion equilibrium exists if \( \beta \leq \beta^c \), with \( y_1 = 0, y_2 = 0 \) and \( x \in [0, 1] \); (2) a strong distortion equilibrium exists if \( \beta \in [\beta^c, \beta^s] \), with \( y_1 = 0, y_2 \in (0, 1) \) and \( x \in [0, 1] \) such that \( x = 0 \) if \( \beta \) is sufficiently close to \( \beta^c \); (3) either a strong or a weak distortion equilibrium exists, or both exist, if \( \beta \in (\beta^s, \beta^w] \); (4) a weak distortion equilibrium exists if \( \beta > \beta^w \), with \( y_2 = 1, y_1 \in (0, 1) \) and \( x \in [0, 1] \) such that \( x = 0 \) if \( \beta \) is sufficiently high.

Biased B with very low reputational concerns (\( \beta \leq \beta^c \)) always reports \( m_B = 1 \) because his benefit from distortion strictly exceeds his reputation cost. In this case, biased A has no influence on the decisionmaker: his message affects neither C’s action nor his own reputation. Consequently, he is free to choose any truth-telling probability \( x \in [0, 1] \).\(^{18}\) Biased A only has influence, in which case he never reports completely truthfully, if biased B may alter his recommendation because of \( m_A \) (\( y_1 \neq y_2 \)).

\(^{18}\) The cutoff \( \beta^c \) is independent of \( x \). Instead, it depends on B’s characteristics and A’s signal quality, which affects B’s private estimate of the true state if \( m_A = 0, s_B = 0 \).
If biased $B$ has low reputational concerns ($\beta \in [\beta^s, \beta^c]$), he wants to distort with some probability even if both signals are against his bias.\footnote{At the cutoff $\beta^c$, $y_1 = 0, y_2 = 1$ is biased $B$’s best response to $x = 0$ if $m_A = 0, s_B = 0$.} In a strong distortion equilibrium, biased $A$’s reputation cost \textit{increases}, rather than decreases, in biased $B$’s truth-telling probability $y_2$. Intuitively, this is because here biased $B$ (primarily) follows $A$’s message if $s_B = 0$. Thus if $m_B = 1$, $C$ assigns some blame to $A$ for (possibly) misleading biased $B$. As a result, biased $A$ has more to lose if he distorts.

In a weak distortion equilibrium, biased $B$ has sufficiently high reputational concerns that he does not want to distort if both signals are against his bias; he only distorts if $m_A = 1$: $y_1 > 0$ if $\beta > \beta^w$.\footnote{At the cutoff $\beta^w$, $y_1 = 0, y_2 = 1$ is biased $B$’s best response to $x = 1$ if $m_A = 1, s_B = 0$.} In this case, biased $A$’s reputation cost decreases in biased $B$’s truth-telling probability $y_1$ due to the reputation sensitivity effect. If $y_1$ increases, $C$ knows that biased $B$ is less influenced by $A$’s message, thus her posterior belief of $A$’s objectivity varies less with $m_B$. This also explains why, if $y_1$ is sufficiently close to 1, biased $A$ always reports $m_A = 1$. Since biased $B$ reports (almost) truthfully, his recommendation is credible. Therefore, biased $A$ strictly benefits from distortion while his reputation cost approaches zero. That is, even if biased $A$ is very concerned about his reputation, he may nonetheless distort to exert indirect influence. Further, if positive, $x$ and $y_1$ are strategic substitutes because each biased agent’s distortion benefit strictly increases in the other’s truth telling; and his reputation cost strictly decreases in it. Hence if biased $A$ reports more truthfully, biased $B$ distorts more and vice versa.

C. Decisionmaker $C$: Information Loss and Policy Remedy

If experts are influenced by sources with inferior and potentially distorted information, their recommendations may exacerbate the decisionmaker’s loss from taking wrong actions.\footnote{The settlement of the $185$ million class action lawsuit against Bristol-Myers Squibb in January 2006 shows that they paid physicians to exaggerate the benefits of their drug for patients with high blood pressure and heart failure. These physicians also failed to report publicly on substantial numbers of life-threatening drug complications which they knew to exist. See, for instance, \textit{Boston Globe}. “How Drug Lobbyists Influence Doctors” by Jerome P. Kassirer, 13 February 2006.} Before receiving $m_B$, decisionmaker $C$’s ex ante expected payoff is:

$$EU_C = -\mathbb{E}_{\eta}\mathbb{E}_{m_B} [(\Pr(\eta = 1|m_B) - \eta)^2|m_B] = -0.5 \sum_{m_B} \Pr(\eta = 1|m_B) (1 - \Pr(\eta = 1|m_B)),$$

which depends only on her action given $B$’s message: $a = \Pr(\eta = 1|m_B)$. In particular, it increases in the probability that $B$’s recommendation is correct: $\Pr(\eta = 1|m_B = 1)$ and $\Pr(\eta = 0|m_B = 0)$. Because biased
$B$ may be influenced by $m_A$, everything else being equal, $C$’s ex ante expected payoff increases in $x$, since she benefits indirectly from a more accurate $m_A$, which is passed on by biased $B$ with some probability. Similarly, it increases in $y_1$ in a weak distortion equilibrium and $y_2$ in a strong distortion equilibrium, because biased $B$ reports his true, and better, signal with a higher probability.

Given that $C$’s expected payoff increases in the probability that $m_B$ is correct, one natural question is whether the decisionmaker benefits if $A$’s information quality increases. That is, do patients benefit if the physicians have access to better information from other sources?

**PROPOSITION 3:** For any given $x$, $C$’s ex ante expected payoff decreases in $p_A$ if $\beta$ is sufficiently large.

Proposition 3 suggests why sometimes the decisionmaker may want well informed experts to give independent recommendations instead of listening to weak (even if honest) sources of information. The direct effect of a more informative message from $A$ on $C$ is positive because as $p_A$ increases, both $m_B = 1$ and $m_B = 0$ become more credible, which reduces $C$’s loss from taking the wrong actions. The indirect effect through $B$, however, is negative and outweighs the positive direct effect if $\beta$ is sufficiently large. This is because biased $B$ may distort more, not less, because of the information aggregation effect if $A$’s signal quality increases. Recall from Proposition 2 that $y_1$ is close to 1 if $\beta$ is sufficiently high. In this case, the direct effect is negligible because biased $B$ is already reporting very truthfully. As a result, $C$ is worse off.

One may think that $C$ should increase truth telling by making it more costly for biased agents to distort. For instance, she may raise the reputation cost of the intermediary by changing $\beta$; or that of the source by changing $\alpha$.\footnote{Even though only measures affecting the reputation cost of the agents are considered here, monetary fines or any cost required to implement such measures can be easily incorporated.} In practice, such measures may take the form of more rigorous medical board reviews; or stronger disclosure rules by the pharmaceutical companies. However, the strategic interactions between biased agents suggest that the net effect of such measures must also be considered.

**PROPOSITION 4:** (1) A biased agent’s truth-telling probability, if positive, increases in his own weight on reputation. (2) In a strong distortion equilibrium, $y_2$ decreases in $\alpha$. If further $\beta$ is larger and sufficiently close to $\beta^c$, then $C$’s expected payoff increases in $\beta$ and decreases in $\alpha$. (3) In a weak distortion equilibrium, $y_1$ decreases in $\alpha$ and $x$ decreases in $\beta$. If further $\beta$ is sufficiently high, $C$’s expected payoff increases in $\beta$ and is unaffected by $\alpha$.\footnote{Even though only measures affecting the reputation cost of the agents are considered here, monetary fines or any cost required to implement such measures can be easily incorporated.}
Surprisingly, Proposition 4 shows that, increasing the reputation cost of the biased source — pharmaceutical companies in the opening example — may make the patients worse off. To see this, observe that in a strong distortion equilibrium, if $\alpha$ increases, biased $A$ reports more truthfully because it is more costly for him to lie, which leads biased $B$ to distort more as shown in Proposition 1. Further, if $\beta$ is sufficiently low, $y_2$ is close to 0 and thus biased $A$ has little influence on $m_B$. As a result, $C$ gains little from $A$’s more truthful reporting; but biased $B$ lies against both signals with a higher probability. Thus the negative effect from the fall in $y_2$ dominates. Intuitively, a small increase in $A$’s reputational concerns may worsen the intermediary’s incentives when it counts the most. As an example, Figure 2 depicts $C$’s ex ante expected payoff as a function of $\beta$ when $
abla_A = \nabla_B = 0.5, p_A = 0.7$ and $p_B = 0.95$. Note that for $\beta \in [0.72, 1.01]$, $C$’s payoff is lower if $\alpha = 1$ than $\alpha = 2$.

FIGURE 2: Comparing $C$’s ex ante expected payoffs at $\alpha = 1$ and $\alpha = 2$.

If $C$ increases $\beta$ instead, biased $B$ always reports more truthfully. Moreover, $C$’s payoff strictly increases if $\beta$ is sufficiently close to $\beta^c$, or when $\beta$ is sufficiently high. In the medical industry example, strengthening medical board review process or disclosure rules improve the credibility of the medical profession not only because the physicians are under less industry influence, but also because drug companies may become more
truthful in revealing side effects. To see this, observe that in the former case, as \( y_2 \) increases, the increase in biased \( A \)’s reputation cost due to the blame sharing effect exceeds the gain in his benefit from distortion. As a result, \( x \) increases in \( \beta \) (if positive). In the latter case, a weak distortion equilibrium exists in which \( x \) and \( y_1 \) are strategic substitutes if \( x > 0 \). However, Proposition 2 shows that if \( \beta \) is sufficiently high, biased \( A \) always reports \( m_A = 1 \). Thus increasing \( \beta \) makes \( C \) strictly better off.

IV. SIMULTANEOUS REPORTING

Businesses sometimes exert influence indirectly through experts, and sometimes try to reach the consumers directly through advertising campaigns. Simultaneous reporting serves as a natural benchmark against the main model, since both agents have informative signals and their messages may be useful to the decisionmaker.

Suppose instead of communicating through \( B \), biased \( A \) sends a message directly to \( C \). Decisionmaker \( C \) now receives \( m_A \) and \( m_B \) before taking an action. All the other assumptions remain. Biased \( A \) then chooses \( m_A \), given his signal, to maximize his expected payoff:

\[
E_{m_B}[\Pr(\eta = 1|m_A, m_B)|s_A] + E_\eta[\Pr(A = 0|m_A, \eta)|s_A];
\]

and biased \( B \) chooses \( m_B \) similarly. Note that the first part of biased \( A \)’s expected payoff depends on both \( A \) and \( B \)’s messages, because the presence of multiple messages affects the tradeoff a biased agent faces primarily by changing the marginal impact of his message on \( C \)’s action, and thus his (expected) distortion benefit. In contrast, a biased agent’s reputation cost, the second part of his expected payoff, depends solely on \( m_A \). Because \( A \) and \( B \)’s signals are independent conditional on the true state observed by \( C \) at the evaluation stage, \( m_B \) imposes no additional discipline on biased \( A \) through \( A \)’s reputational concerns.\(^{23}\)

Under simultaneous reporting, then, all the strategic interactions between the biased agents enter through their messages’ influence on \( C \)’s action. The following result characterizes their equilibrium behavior.

PROPOSITION 5: (1) There exists a distortion equilibrium in which biased \( i = A, B \) reports \( m_i = 1 \) if \( s_i = 1 \). If \( s_i = 0 \), biased \( i \) always reports \( m_i = 1 \) if his weight on reputation \( \alpha \) or \( \beta \) is sufficiently low; and reports

\(^{23}\) Additional discipline is present if \( A \) and \( B \)’s signals are correlated. This effect surfaces in Chan, Li, and Suen (2007), who study grade inflation and show that the market judges how likely one school has distorted in terms of grade inflation based on the other school’s grades if their students’ quality is correlated. Even if the signals are independent conditional on the state, Gerardi, McLean, and Postlewaite (2008) show that a decisionmaker who receives reports from multiple partially informed sources can extract more truthful reports by exploiting the correlation of the sources’ signals with the true state.
m_i = 0 with probability x_i > 0 if his weight on reputation is sufficiently high. (2) If positive, x_A and x_B are strategic complements, and C’s ex ante expected payoff increases in \( \alpha \) and in \( \beta \).

It is worth noting that, under simultaneous reporting, each biased agent’s (expected) distortion benefit decreases in the other’s truth telling. For instance, if \( s_A = 0 \) and biased \( B \) is more truthful (\( x_B \) increases), then biased \( A \) has a smaller marginal impact on \( C \)’s action if he reports \( m_A = 1 \) instead of \( m_A = 0 \). There are two cases. First, if \( m_B = 1 \), then as \( x_B \) increases, \( m_B = 1 \) is more credible and \( C \) believes more strongly that the state is 1. Thus \( m_A = 1 \) induces a smaller additional change in her action. Second, if \( m_B = 0 \), then the marginal impact of \( A \)’s message does not vary with \( x_B \) because \( C \) knows that \( s_B = 0 \) in this case. But \( m_B = 0 \) makes \( C \) believe that \( \eta = 0 \) is more likely, and thus \( m_A = 1 \) becomes less credible than if \( m_B = 1 \). Since \( s_A = 0 \), biased \( A \) knows that \( B \) is more likely to receive \( s_B = 0 \) and to report \( m_B = 0 \) as biased \( B \) becomes more truthful. Since biased \( A \) is less effective at changing \( C \)’s action in either case, his marginal impact on \( C \) falls in \( x_B \), but his reputation cost is unaffected. Consequently, biased \( A \) distorts less than he would have as the only source of information. Moreover, this also implies that with simultaneous reporting, biased \( A \) and \( B \)’s truth-telling probabilities \( x_A, x_B \) are strategic complements.

This complementarity implies that \( C \) can increase both \( x_A \) and \( x_B \) by increasing either agent’s reputation cost. As both \( A \) and \( B \)’s messages become more accurate, \( C \)’s expected payoff increases.24 In practice, this suggests that policies aiming to improve reporting accuracy from potentially biased sources should not be limited to stricter regulations against one possible source of distortion. It may be more effective to encourage simultaneous reporting, for example, by severing the ties between physicians, researchers and the drug industry.25

V. KEY ASSUMPTIONS AND CONCLUSION

A. Discussion of Key Assumptions

a. Richer Message Space for Agent B. In this model, agent \( B \) sends a binary message \( m_B \in \{0, 1\} \) to \( C \), rather than indicating both what he has heard from \( A \) and his own signal. One reason for this simplifying assumption is that experts often give simple recommendations when the issue at stake is complex, for example,

---

24 One caveat concerning policies encouraging simultaneous reporting is that it may be counterproductive if both agents have sufficiently low reputational concerns that they always distort (\( x_A = 0 \) and \( x_B = 0 \)). The reason is that \( C \) may receive two biased messages instead of one in the main model.

25 The University of Pennsylvania Health System took a step in this direction by requiring that industry can make gifts to departments to support educational programs (but not to individual faculty), while the money is disbursed at the discretion of department chairs.
the effectiveness of a new drug. Another reason is that this simplification helps illustrate how a biased intermediary with reputational concerns aggregates A’s message and his own signal — inefficiently from the decisionmaker’s point of view — to induce the highest action possible from C.

Appendix B briefly considers the case in which intermediary B sends a vector of report to C: $(\hat{m}_A, m_B)$, where B’s report of A’s message is $\hat{m}_A$, and his report of his own signal remains $m_B$. By assumption, objective B reports his information accurately: $\hat{m}_A = m_A, m_B = s_B$, while biased agents still want to induce higher actions and to appear objective. In this new game, as shown in Appendix B, distortion equilibrium qualitatively similar to that in the main model still exists. In equilibrium, the more biased B’s information favors his agenda, the more he distorts. The same information aggregation effect — since $m_A$ and $s_B$ are still private information of biased B — means that biased B ranks his payoffs from potential distortions in a similar fashion.

The richer message space, however, means that biased B can distort either $m_A$, or $s_B$, or both. In equilibrium, biased B is less truthful in reporting $s_B = 0$ than in reporting $m_A = 0$: as his reputational concerns increase, he first reports $\hat{m}_A = 0$ with a positive probability before he reports $m_B = 0$ with any positive probability. When $\beta$ is sufficiently low, biased B always reports $m_B = 1$. Further, if B is very likely objective ($\theta_B$ sufficiently close to 1), then he always reports $\hat{m}_A = 1$ as well because biased B’s distortion benefit is higher than that in the main model. This is because, in comparison with the main model, $(1,1)$ is more favorable news for $\eta = 1$ than $m_B = 1$, and $(0,0)$ is less favorable news for $\eta = 1$ than $m_B = 0$ if $\theta_B$ is sufficiently high. In this case, a complete distortion equilibrium exists and biased B may do so for a larger set of reputational concerns (the cutoff value of $\beta$ is higher than $\beta^C$ in the main model). If $\theta_B$ is not high, or if biased B has moderate reputational concerns, he still reports $m_B = 1$, but he starts reporting $\hat{m}_A = 0$ with positive probabilities. Intuitively, with a richer message space, biased B can attempt to appear objective by reporting $\hat{m}_A = 0$, and still induces a high action by reporting $m_B = 1$. This is because his own signal is more informative, and because C is more likely to believe a more objective intermediary. If his reputational concerns are sufficiently high, then he reports $m_B = 0$ sometimes. And the equilibrium is akin to the strong and weak distortion equilibrium in the main model. Biased A, however, expects his message to reach C with a positive probability due to the presence of objective B, therefore he always distorts to some extent.

The effect of requiring a vector of report on the decisionmaker is ambiguous. With a richer message space for B, on one hand, C may be worse off because both biased agent may lie more which lead to more extreme actions when their reputational concerns are sufficiently low. On the other hand, she may learn A’s message if
she faces an objective $B$ or a biased $B$ with sufficiently high reputational concerns. That is, some information from $A$ may reach $C$ and moderate her action. The net effect depends on parameters of the model.

\textit{b. Relative informativeness of signals.} In this model, the intermediary’s signal is assumed to be more informative than that of $A$’s due to expertise or experience. In some settings, agent $A$ may have better, or exclusive information, for example, military intelligence. If $A$’s signal is far more informative, an objective intermediary follows $m_A$ if it is unlikely that $A$ distorts, but dismisses it otherwise. In particular, if $B$’s signal is sufficiently uninformative, an objective $B$ always follows $m_A$ and becomes a pure intermediary, despite $A$’s potential distortions. Li (2009) shows that if $B$ is a pure intermediary, biased $A$ and $B$’s truth-telling incentives are complements. Moreover, biased $A$ always lies more using an intermediary — and he may prefer a more biased intermediary to a more objective one — because the decrease in his reputation cost due to $B$’s blame sharing outweighs any loss in his message’s credibility due to biased $B$’s possible distortion.

\textit{c. Biased $A$’s ex ante preference.} This paper focuses on $C$’s information loss due to biased agents’ distortions. A related question is whether a biased sender prefers to send his own message to the decisionmaker, for instance, by advertising, or to communicate through an expert. Suppose that agent $A$ commits to a communication method before receiving his signal so that his preference does not reveal information about his signal. Then $A$’s ex ante expected payoff if he communicates through $B$ is simply:

\[
0.5 \left[ \mathbb{E}_{m_B} \left[ \Pr(\eta = 1|m_B) \right] + \alpha \mathbb{E}_\eta \left[ \Pr(A = o|m_B, \eta) \right] \right] \left[ m_B \right].
\]

On one hand, if biased $A$ and $B$ have sufficiently low reputational concerns, biased $A$ may prefer sending a direct message, because in a complete distortion equilibrium, $m_A$ has no influence on $B$, and thus no influence on $C$, as shown in Proposition 2. In that case, if $m_B$ is not credible, $A$’s direct message may have a stronger impact on $C$’s action than $B$’s. Although reporting $m_A = 1$ directly hurts $A$’s reputation, it is negligible due to his low reputational concerns. On the other hand, if $\beta$ is sufficiently high, then recall from Proposition 2 that biased $A$ always reports $m_A = 1$. In this case, biased $B$ reports very truthfully ($y_1$ sufficiently close to 1) that biased $A$ pays a negligible reputation cost, which is proportional to $1 - y_1$; but he still has some influence on $C$ because $B$’s recommendation is highly credible. Thus he can “afford” to lie while he could not do so, at the same level of reputational concerns, if he sends a message of his own. Intuitively, if biased $B$ has sufficiently high reputational concerns, biased $A$ is better off influencing him than risking losing his own reputation by sending $C$ a message directly. This suggests that a drug company highly concerned about its reputation prefers influencing physicians or researchers with high levels of reputational concerns.
d. Role of the objective types. In this model, the objective agent conveys the best information available to him. This assumption, consistent with the stated goals of experts such as stock analysts, critics and physicians, also greatly simplifies the inference problem of decisionmaker \( C \). An objective agent, however, may have reputational concerns as well. Morris (2001) shows that, in a model of direct communication, an objective expert may lie even though he wants the decisionmaker to take the correct action, because he doesn’t want to be grouped with biased experts and lose future influence. Similar incentives may arise here if the objective intermediary is also concerned about his reputation. Recall that \( m_B = 0 \) but \( \eta = 1 \) is a good sign of \( B \)’s objectivity because it suggests that he is not influenced by \( m_A = 1 \). Thus an objective \( B \) may want to report 0 if \( m_A = 0, s_B = 1 \), driving biased \( B \) to do so as well.

B. Conclusion

In the opening example, recommendations from physicians are important for the patients, and independent judgment is a sign of objectivity. However, biased experts may selectively incorporate inferior information from potentially biased sources before making their recommendations. Thus the decisionmaker may want to encourage the experts to give recommendations based on their own information. But policy measures designed to make the source more truthful may reduce the truthfulness of the expert’s recommendation to such an extent that the decisionmaker is strictly worse off. It may be more effective to either increase the expert’s cost of distortion; or to encourage simultaneous reporting from the source and the expert by severing their ties.

Appendix A. Proofs

Proof of Proposition 1: We first consider biased \( B \)’s truth-telling incentives given his signal \( s_B \) and message \( m_A \). Given biased \( A \), \( B \)’s distortion strategy, biased \( B \) chooses \( m_B \) to maximize his expected payoff. If he reports \( m_B = 1 \) instead of \( m_B = 0 \), the difference in biased \( B \)’s expected payoffs, \( \mathbb{E}U_B(m_B = 1|m_A, s_B) - \mathbb{E}U_B(m_B = 0|m_A, s_B) \), is:

\[
\Pr(\eta = 1|m_B = 1) - \Pr(\eta = 1|m_B = 0) - \beta \sum_\eta \Pr(\eta|m_A, s_B)[\Pr(B = o|m_B = 0, \eta) - \Pr(B = o|m_B = 1, \eta)].
\]

Let \( C \)’s belief that agent \( A \) reports \( m_A = 0 \) be \( N_x = \theta_A + (1 - \theta_A)x \); and her belief that agent \( B \) reports \( m_A = 0 \) be \( N_{y_1} = \theta_B + (1 - \theta_B)y_1 \) and \( N_{y_2} = \theta_B + (1 - \theta_B)y_2 \). Also, let \( C \)’s actions given \( B \)’s recommendation be
\[
\hat{\eta}_1^B = \Pr(\eta = 1|m_B = 1) \quad \text{and} \quad \hat{\eta}_0^B = \Pr(\eta = 1|m_B = 0). \]

Then:
\[
\hat{\eta}_1^B = \frac{1 - (1 - p_B)[N_{y_1}(1 - (1 - p_A)N_x) + N_{y_2}(1 - p_A)N_x]}{2 - (1 - p_B)[N_{y_1}(1 - (1 - p_A)N_x) + N_{y_2}(1 - p_A)N_x] - p_B[N_{y_1}(1 - p_A)N_x] + N_{y_2}p_AN_x};
\]
\[
\hat{\eta}_0^B = \frac{(1 - p_B)[N_{y_1}(1 - (1 - p_A)N_x) + N_{y_2}(1 - p_A)N_x] + p_B[N_{y_1}(1 - p_A)N_x] + N_{y_2}p_AN_x]}{(1 - p_B)[N_{y_1}(1 - (1 - p_A)N_x)] + N_{y_2}(1 - p_A)N_x].
\]

Biased \(B\)’s belief the true state is 0 given his information, \(\Pr(\eta = 0|m_A, s_B)\), is:
\[
\Gamma_1 \equiv \Pr(\eta = 0|m_A = 1, s_B = 0) = \frac{(1 - p_A)N_x}{(1 - p_A)N_x + (1 - (1 - p_A)N_x)(1 - p_B)};
\]
\[
\Gamma_2 \equiv \Pr(\eta = 0|m_A = 0, s_B = 0) = \frac{p_A(1 - p_B)}{p_A(1 - p_B) + p_B(1 - p_A)};
\]
\[
\Gamma_3 \equiv \Pr(\eta = 0|m_A = 0, s_B = 1) = \frac{(1 - p_A)N_x}{(1 - p_A)N_x + (1 - (1 - p_A)N_x)p_B}.
\]

Because \(p_B > p_A\), simple calculations can show that \(\Gamma_2 > \Gamma_1 > 0.5 > \Gamma_3 > \Gamma_4\). To simplify notations, denote \(B\)’s posterior objectivity as \(\tau_{m_B,\eta}^B\), and we have:
\[
\tau_{1,0}^B \equiv \Pr(B = o|m_B = 1, \eta = 0) = \frac{(1 - p_B)\theta_B}{1 - p_B + (1 - \theta_B)p_B[(1 - y_1)(1 - p_A)N_x] + (1 - y_2)p_AN_x};
\]
\[
\tau_{0,0}^B \equiv \Pr(B = o|m_B = 0, \eta = 0) = \frac{\theta_B}{\theta_B + (1 - \theta_B)[y_1(1 - p_A)N_x] + y_2p_AN_x};
\]
\[
\tau_{0,1}^B \equiv \Pr(B = o|m_B = 0, \eta = 1) = \frac{\theta_B}{\theta_B + (1 - \theta_B)[y_1(1 - (1 - p_A)N_x)] + y_2(1 - p_A)N_x};
\]
\[
\tau_{1,1}^B \equiv \Pr(B = o|m_B = 1, \eta = 1) = \frac{p_B\theta_B}{p_B + (1 - \theta_B)(1 - p_B)[(1 - y_1)(1 - p_A)N_x] + (1 - y_2)(1 - p_A)N_x}.
\]

For biased \(B\) to report \(s_B\) truthfully, the following four incentive constraints (IC) must be satisfied:
\[
\hat{\eta}_1^B - \hat{\eta}_0^B \leq \Delta_1 \equiv \beta [\Gamma_1(\tau_{0,0}^B - \tau_{1,0}^B) + (1 - \Gamma_1)(\tau_{0,1}^B - \tau_{1,1}^B)];
\]
\[
\hat{\eta}_1^B - \hat{\eta}_0^B \leq \Delta_2 \equiv \beta [\Gamma_2(\tau_{0,0}^B - \tau_{1,0}^B) + (1 - \Gamma_2)(\tau_{0,1}^B - \tau_{1,1}^B)];
\]
\[
\hat{\eta}_1^B - \hat{\eta}_0^B \geq \Delta_3 \equiv \beta [\Gamma_3(\tau_{0,0}^B - \tau_{1,0}^B) + (1 - \Gamma_3)(\tau_{0,1}^B - \tau_{1,1}^B)];
\]
\[
\hat{\eta}_1^B - \hat{\eta}_0^B \geq \Delta_4 \equiv \beta [\Gamma_4(\tau_{0,0}^B - \tau_{1,0}^B) + (1 - \Gamma_4)(\tau_{0,1}^B - \tau_{1,1}^B)].
\]

IC (1) and IC (2) concern the case when \(s_B = 0\); and IC (3) and IC (4) concern the case when \(s_B = 1\). The left hand side (LHS) of all the above ICs is the same, which is biased \(B\)’s distortion benefit from reporting \(m_B = 1\) instead of reporting \(m_B = 0\). The right hand side (RHS) of the above ICs is biased \(B\)’s reputation
cost if he reports \( m_B = 1 \) versus \( m_B = 0 \) given \( m_A, s_B \). Biased \( B \)'s truthful reporting depends on how large his distortion benefit is relative to his reputation cost. Because \( \Gamma_2 > \Gamma_1 > \Gamma_3 > \Gamma_4 \), we can rank the RHS of these ICs. For instance, \( \Delta_2 - \Delta_1 = (\Gamma_2 - \Gamma_1)(\tau_{0,0}^B - \tau_{0,1}^B + \tau_{1,1}^B - \tau_{1,0}^B) \). Other comparisons are similar.

The term \((\tau_{0,0}^B - \tau_{0,1}^B + \tau_{1,1}^B - \tau_{1,0}^B)\) is positive if \( B \)'s expected posterior objectivity of giving correct recommendations \((\tau_{0,0}^B + \tau_{1,1}^B)\) is larger than that of giving wrong recommendations \((\tau_{0,1}^B + \tau_{1,0}^B)\); it is negative otherwise. One sufficient condition for \((\tau_{0,0}^B - \tau_{0,1}^B + \tau_{1,1}^B - \tau_{1,0}^B) > 0\) is if \( p_A \) is sufficiently close to 0.5; or if \( \theta_B \) is sufficiently close to 1. More formally, for every \( p_A \), there exists a cutoff \( \bar{\theta}_B \) such that this condition holds if \( \theta_B > \bar{\theta}_B \). Also, \( \bar{\theta}_B = 0 \) if \( p_A \) is sufficiently close to 0.5; and \( \bar{\theta}_B \) is arbitrarily close to 1 if \( p_A \) is arbitrarily close to \( p_B \).

By assumption, \( p_A \) is sufficiently close to 0.5, and thus \((\tau_{0,0}^B - \tau_{0,1}^B + \tau_{1,1}^B - \tau_{1,0}^B) > 0\). Let the probabilities that biased \( B \) reports his signal \( s_B = 1 \) truthfully be: \( z_1 = \Pr(m_B = 1|m_A = 0, s_B = 1) \) and \( z_2 = \Pr(m_B = 1|m_A = 1, s_B = 1) \). If biased \( B \) reports truthfully, then \( y_1 = y_2 = z_1 = z_2 = 1 \). The following lemma establishes claim (1) and (2) of Proposition 1.

**Lemma 1:** In any (continuation) equilibrium, \( z_1 = 1, z_2 = 1, \) and \( y_1 \leq y_2 \). The inequality is strict if \( y_2 > 0 \). Moreover, \( y_1 \) and \( y_2 \) cannot both be strictly between 0 and 1.

**Proof:** Because \((\tau_{0,0}^B - \tau_{0,1}^B + \tau_{1,1}^B - \tau_{1,0}^B) > 0\), biased \( B \)'s reputation cost can be ranked such that \( \Delta_2 > \Delta_1 > \Delta_3 > \Delta_4 \). Because the LHS of all the ICs is the same, it is not possible for more than one IC to hold with equality, thus at most one of the four probabilities \( y_1, y_2, z_1, z_2 \) can be strictly between zero and one. Also, truth telling cannot be part of the equilibrium. If \( y_1 = 1, y_2 = 1 \), the LHS of IC (1) is strictly positive, while the RHS of IC (1) and (2) are zero, a contradiction.

If any IC holds with equality, there are four possibilities. First, suppose that IC (3) binds, then \( 0 < z_1 < 1 \). This implies that \( z_2 = 1, y_1 = y_2 = 1 \). This possibility corresponds to the case when biased \( B \) reports \( s_B = 0 \) truthfully but lies with some probability if \( s_B = 1, m_A = 0 \). In this case, \( C \) knows that \( s_B = 1 \) if \( m_B = 1 \), which implies that \( \hat{\eta}_1^B = p_B > \hat{\eta}_0^B \). Thus the distortion benefit for biased \( B \) is positive \((\hat{\eta}_1^B - \hat{\eta}_0^B > 0)\). On the reputation side, however, we can show that \( \tau_{0,0}^B < \tau_{1,0}^B \) and \( \tau_{0,1}^B < \tau_{1,1}^B \), and thus biased \( B \)'s reputation cost is negative. Together, biased \( B \) is strictly better off reporting \( m_B = 1 \), which is a contradiction. Intuitively, in this putative equilibrium, \( m_B = 1 \) results from signal \( s_B = 1 \), and it is more likely to come from objective \( B \), therefore it is both credible and a good sign of objectivity. Similarly, IC (4) binds and thus \( z_1 = 0, z_2 > 0 \) cannot be part of an equilibrium. The only remaining possibility is for either IC (1) or IC (2) to bind. In either
case, \( z_1 = 1, z_2 = 1 \). If IC (1) binds, then \( y_1 > 0, y_2 = 1 \); and if IC (2) holds or binds, then \( y_1 = 0, y_2 \leq 1 \), thus \( 0 \leq y_1 < y_2 \). Finally, if none of the ICs binds, then \( z_1 = z_2 = 1, y_1 = y_2 = 0 \), in which case biased \( B \) always reports \( m_B = 1 \) regardless of his information. \( \square \)

For claim (3) of Proposition 1, note that \( \hat{y}_1^B - \hat{y}_0^B \) has the same sign and is proportional to \( (2p_B - 1)Ny_1 + (p_Bp_A - (1 - p_B)(1 - p_A))N_x(N_{y_2} - N_{y_1}) \). It is positive because \( m_B = 1 \) is more likely to come from \( s_B = 1 \). Thus \( C \) always take a higher action after \( m_B = 1 \) than \( m_B = 0 \). Moreover, \( \hat{y}_1^B \) increases in \( y_2 \) and \( \hat{y}_0^B \) decreases in \( y_2 \), thus biased \( B \)'s distortion benefit strictly increases in \( y_2 \): the more truthful he is, the more \( C \) changes her action. In terms of biased \( B \)'s distortion benefit if \( m_A = 1, s_B = 0 \), we can show that both \( \hat{y}_1^B \) and \( \hat{y}_0^B \) increase in \( y_1 \). Here, \( \hat{y}_1^B \) increases in \( y_1 \) because message \( m_B = 1 \) is more likely to come from \( s_B = 1 \); and \( \hat{y}_0^B \) increases in \( y_1 \) because the higher is \( y_1 \), the more likely that biased \( B \) does not use \( m_A \), which may result from \( s_A = 1 \). But when \( p_A \) is sufficiently close to 0.5 (or if \( \theta_B \) is sufficiently large), \( \hat{y}_1^B - \hat{y}_0^B \) increases in \( y_1 \). Also, the reputation cost of distortion for biased \( B \) strictly decreases in \( y_1 \) and \( y_2 \) because both \( \tau_{0,0}^B - \tau_{1,0}^B \) and \( \tau_{0,1}^B - \tau_{1,1}^B \) decrease in \( y_1, y_2 \). Intuitively, the more truthful biased \( B \) is, the less \( C \) changes her estimates about \( B \)'s objectivity. Finally, simple calculations can show that \( \hat{y}_1^B - \hat{y}_0^B \) increases in \( N_x \), and biased \( B \)'s reputation cost decreases in \( N_x \). By the implicit function theorem, if \( y_1 \in (0, 1) \), it strictly decreases in \( N_x \). Similarly, if \( y_2 \in (0, 1) \), it strictly decreases in \( N_x \).

**Proof of Proposition 2:** Given the distortion strategy of biased \( A \) and \( B \), biased \( A \)'s (expected) benefit from reporting \( m_A = 1 \) instead of \( m_A = 0 \) given his signal is:

\[
\begin{align*}
\mathbb{E}_{m_B} & [\Pr(\eta = 1|m_B)|m_A = 1, s_A = 0] - \mathbb{E}_{m_B} [\Pr(\eta = 1|m_B)|m_A = 0, s_A = 0] \\
&= (N_{y_2} - N_{y_1})[p_Ap_B + (1 - p_A)(1 - p_B)](\hat{y}_1^B - \hat{y}_0^B) \\
&= (N_{y_2} - N_{y_1})[p_A(1 - p_B) + p_B(1 - p_A)](\hat{y}_1^B - \hat{y}_0^B).
\end{align*}
\]

Note that \( m_A \) only influences \( C \)'s action if biased \( B \) changes \( m_B \) based on what he hears. Recall that \( \kappa \equiv N_{y_2} - N_{y_1} \), if \( \kappa = 0 \), then biased \( A \) gets no benefit from distortion. Biased \( A \) is also concerned about how objective \( C \) thinks he is at the evaluation stage, given \( B \)'s recommendation and the observed true state. To simplify notations, let \( A \)'s posterior objectivity be \( \tau_{m_B, \eta}^A \). Using Bayes’ rule, we have:

\[
\tau_{0,0}^A \equiv \Pr(A = 0|m_B = 0, \eta = 0) = \frac{p_AN_{y_2} + (1 - p_A)N_{y_1}}{p_AN_{y_2}N_{y_2} + (1 - p_A)N_xN_{y_1}}.
\]
Clearly, if to report and incurs no reputation cost, and thus he is indifferent between his messages and he can report in any way.

\[ \tau_{1,0}^A = \Pr(A = o|m_B = 1, \eta = 0) = \frac{[p_A(1 - p_B N_{y_2}) + (1 - p_A)(1 - p_B N_{y_1})] \theta_A}{p_A N_x (1 - p_B N_{y_2}) + (1 - p_A) N_x (1 - p_B N_{y_1})}; \]

\[ \tau_{0,1}^A = \Pr(A = o|m_B = 0, \eta = 1) = \frac{[1 - (1 - p_A) N_x N_{y_1}] \theta_A}{(1 - p_A) N_x (1 - (1 - p_B) N_{y_1})}; \]

\[ \tau_{1,1}^A = \Pr(A = o|m_B = 1, \eta = 1) = \frac{[1 - (1 - p_A) (1 - p_B) N_{y_1}] + p_A (1 - (1 - p_B) N_{y_1})] \theta_A}{(1 - p_A) N_x (1 - (1 - p_B) N_{y_1})}. \]

Biased A’s reputation cost if he reports \( m_A = 0 \) instead of \( m_A = 1 \) given his signal is:

\[
\mathbb{E}_{\eta, m_B}[\Pr(A = o|m_B, \eta)|m_A = 0, s_A = 0] - \mathbb{E}_{\eta, m_B}[\Pr(A = o|m_B, \eta)|m_A = 1, s_A = 0] \\
= (N_{y_2} - N_{y_1})[p_A p_B \tau_{0,0}^A + (1 - p_A)(1 - p_B) \tau_{0,1}^A] - (N_{y_2} - N_{y_1})[p_A p_B \tau_{1,0}^A + (1 - p_A)(1 - p_B) \tau_{1,1}^A];
\]

\[
\mathbb{E}_{\eta, m_B}[\Pr(A = o|m_B, \eta)|m_A = 0, s_A = 1] - \mathbb{E}_{\eta, m_B}[\Pr(A = o|m_B, \eta)|m_A = 1, s_A = 1] \\
= (N_{y_2} - N_{y_1})[(1 - p_A) p_B \tau_{0,0}^A + p_A (1 - p_B) \tau_{0,1}^A] - (N_{y_2} - N_{y_1})[(1 - p_A) p_B \tau_{1,0}^A + p_A (1 - p_B) \tau_{1,1}^A].
\]

Clearly, if \( \kappa = 0 \), biased A’s reputation cost is zero. If \( \kappa > 0 \), for biased A to report \( s_A = 0 \) and \( s_A = 1 \) truthfully, the following two incentive constraints must hold:

\[
\hat{\eta}_1^B - \hat{\eta}_0^B \leq \alpha \left[ \Pr(\eta = 0|s_A = 0, s_B = 0) (\tau_{0,0}^A - \tau_{1,0}^A) + \Pr(\eta = 1|s_A = 0, s_B = 0) (\tau_{0,1}^A - \tau_{1,1}^A) \right], \tag{5}
\]

\[
\hat{\eta}_1^B - \hat{\eta}_0^B \geq \alpha \left[ \Pr(\eta = 0|s_A = 1, s_B = 0) (\tau_{0,0}^A - \tau_{1,0}^A) + \Pr(\eta = 1|s_A = 1, s_B = 0) (\tau_{0,1}^A - \tau_{1,1}^A) \right]. \tag{6}
\]

The LHS of IC (5) and (6) is the same, and simple calculations can show that the RHS of IC (5) is always larger than the RHS of IC (6). Thus whenever biased A prefers reporting \( m_A = 1 \) if \( s_A = 0 \), he strictly prefers reporting \( m_A = 1 \) if \( s_A = 1 \). Intuitively, biased A has less to lose if \( s_A = 1 \), because agent B is more likely to report \( s_B = 1 \) correctly, which is a better sign of A’s objectivity. Thus for biased A, if neither IC holds, he always reports \( m_A = 1 \). If IC (5) holds with equality and IC (6) holds strictly, he reports \( m_A = 0 \) if \( s_A = 0 \) with some positive probability.

Next, at \( y_1 = y_2 = 0 \), biased B’s highest reputation cost occurs if \( s_B = 0, m_A = 0 \). Thus there exists a cutoff value \( \beta^c \) such that, at \( y_1 = y_2 = 0 \), IC (2) holds with equality:

\[
\frac{2p_B - 1}{2 - \theta_B} = \beta^c (1 - \theta_B) \left[ \frac{\Gamma_2}{1 - p_B \theta_B} + \frac{1 - \Gamma_2}{1 - (1 - p_B) \theta_B} \right].
\]

Clearly, if \( \beta \leq \beta^c \), biased B reports \( m_B = 1 \) with probability one. In this case, biased A has no influence on C and incurs no reputation cost, and thus he is indifferent between his messages and he can report in any way. Hence if \( \beta \leq \beta^c, x \in [0, 1], y_1 = 0, y_2 = 0 \).
If $\beta > \beta^c$, then $y_1 = 0, y_2 = 0$ cannot be part of an equilibrium. Thus biased $B$ must report $m_B = 0$ truthfully with some probability. For biased $B$, IC (1) implicitly defines $\tilde{y}_1(x)$, a function of $y_1$ with respect to $x$ when $y_2 = 1$. Similarly, IC (2) implicitly defines $\tilde{y}_2(x)$, a function of $y_2$ with respect to $x$ when $y_1 = 0$. Biased $A$’s IC (5) also defines two functions of $x$ with respect to $y_1$ and $y_2$: $\tilde{x}(y_1)$ if $y_2 = 1$; and $\tilde{x}(y_2)$ if $y_1 = 0$. Also, because the biased agents’ distortion benefit and reputation cost are continuous in $x, y_1, y_2$, these best responses are continuous. From Proposition 1, both $\tilde{y}_1(x)$ and $\tilde{y}_2(x)$ decreases in $x$, and biased $A$ and $B$’s distortion benefit $\hat{\eta}_1^B - \hat{\eta}_0^B$ strictly increases in $y_1, y_2, x$. For biased $A$, his reputation cost decreases in $x$. Since both $(\tau^A_{0,0} - \tau^A_{1,0})$ and $(\tau^A_{0,1} - \tau^A_{1,1})$ decrease in $y_1$ and increase in $y_2$, his reputation cost decreases in $y_1$ and increases in $y_2$. By the implicit function theorem, $\tilde{x}(y_1)$ decreases in $y_1$.

To begin with, observe that if $\tilde{y}_2(0) \leq 1$, there does not exist a weak distortion equilibrium. In this case, biased $B$’s IC (2) does not hold at $x = 0$ and $y_1 = 0, y_2 = 1$, i.e., biased $B$ prefers reporting $m_B = 1$ if $s_B = 0, m_A = 0$. Because biased $B$’s reputation cost $\Delta_2 > \Delta_1$, and $\tilde{y}_1(x)$ decreases in $x$, IC (1) cannot bind for any $x$, thus $y_1 > 0$ cannot be part of the equilibrium. One sufficient condition for this to occur is if $\beta \leq \beta^s$, where IC (2) binds at $\beta = \beta^s$, and $y_1 = 0, y_2 = 1, x = 0$. In this case, biased $B$’s best response to $x$, $y^{BR}(x)$, is simply $\tilde{y}_2(x)$ and biased $A$’s best response to $y_2$ is $x^{BR}(y_2)$. Also, $x^{BR}(y_2) = 0$ if $y_2$ is sufficiently close to 0, and $x^{BR}(1) < 1$. To see this, note that $y_2$ increases in $\beta$. If $y_2$ is sufficiently close to 0, which occurs if $\beta$ is sufficiently close to $\beta^c$, biased $A$’s reputation cost (the RHS of IC (5)) is sufficiently close to 0, which is strictly smaller than his benefit from distortion (the LHS of IC (5)). Given that $y_1 = 0$, consider $x^{BR}(y_2)$ and $\tilde{y}_2(x)$. Because $x^{BR}(0) \in [0, 1], x^{BR}(1) < 1$; and $\tilde{y}_2(0) > 0, \tilde{y}_2(1) < \tilde{y}_2(0) \leq 1$, by the intermediate value theorem, the two best responses intersect. Thus if $\beta \in (\beta^c, \beta^s]$, there exists a strong distortion equilibrium in which $x \in [0, 1], y_2 \in (0, 1)$.

If $\tilde{y}_1(1) > 0$, however, there does not exist a strong distortion equilibrium. In this case, biased $B$’s IC (1) holds at $x = 1$ and $y_1 = 0, y_2 = 1$, i.e., biased $B$ prefers reporting $m_B = 0$ with some probability if $s_B = 0, m_A = 1$. Thus IC (2) holds strictly for any $y_2, x$. One sufficient condition for this to occur is if $\beta > \beta^w$, where the cutoff is defined such that at $\beta = \beta^w$, IC (1) binds at $y_1 = 0, y_2 = 1, x = 1$. Because IC (1) binds at $\beta > \beta^w$ and $x = 1, y_1 = 0, y_2 = 1$, IC (2) strictly holds at these values. Since the LHS of IC (2) increases in $x$ and the RHS decreases in it, $\beta^s < \beta^w$. In this case, biased $B$’s best response $y^{BR}(x)$ is simply $\tilde{y}_1(x)$ and biased $A$’s best response $x^{BR}(y_1)$ is $\tilde{x}(y_1)$. Also, $x^{BR}(y_1) = 0$ if $y_1$ is sufficiently close to 1, which occurs if $\beta$ is sufficiently large. To see this, note that $y_1$ increases in $\beta$. If $y_1$ is sufficiently close to 1,
then biased $A$’s reputation cost (the RHS of IC (5)) is sufficiently close to 0, which is strictly smaller than his distortion benefit (the LHS of IC (5)). Given that $y_2 = 1$, consider $x^{BR}(y_1)$ and $\tilde{y}_1(x)$. Because $x^{BR}(0) \geq 0$, $x^{BR}(1) \in [0, 1]$; and $\tilde{y}_1(0) > 0$, $\tilde{y}_1(1) < 1$, by the intermediate value theorem, the two best responses intersect. Thus if $\beta \geq \beta^w$, there exists a weak distortion equilibrium in which $x \in [0, 1], y_1 \in (0, 1)$.

By continuity, and by the intermediate value theorem, the two best responses intersect and there exists a weak distortion equilibrium cannot be ruled out. Thus multiple equilibria may exist if $\beta \in (\beta^s, \beta^w]$. Finally, if $\beta \in (\beta^s, \beta^w]$, then $\tilde{y}_2(0) > 1$ and $\tilde{y}_1(1) = 0$, then there exists a cutoff value $\hat{x}$ such that $y^{BR}(x) = \tilde{y}_1(x)$ if $x \leq \hat{x}$; and $y^{BR}(x) = \tilde{y}_2(x)$ if $x > \hat{x}$. As for biased $A$, at $y_1 = 0, y_2 = 1$, either IC (5) fails to hold, or it binds at some $x$. Let $x' = 0$ in the former case and $x' = x$ in the latter. First, if $x' \leq \hat{x}$, then $0 < \tilde{y}_1(x') < 1$, $0 < \tilde{y}_1(\hat{x}) < \tilde{y}_1(x')$, $x^{BR}(y_1) = 0$ for $y_1$ sufficiently close to 1, $x^{BR}(0) = x'$. By continuity, and by the intermediate value theorem, the two best responses intersect and there exists a weak distortion equilibrium. However, because $x^{BR}(y_2)$ depends on parameter values, a strong distortion equilibrium cannot be ruled out. Second, if $x' > \hat{x}$, then because $\tilde{y}_2(\hat{x}) = 1$, $0 < \tilde{y}_1(x') < \tilde{y}_1(\hat{x}) < 1$, and $x^{BR}(y_2) = 0$ for $y_2$ sufficiently close to 0, $x^{BR}(1) = x'$, the two best responses intersect and there exists a strong distortion equilibrium. However, because the curvatures of $x^{BR}(y_1)$ and $\tilde{y}_1(x)$ depend on parameter values, a weak distortion equilibrium cannot be ruled out. Thus multiple equilibria may exist if $\beta \in (\beta^s, \beta^w]$. 

**Proof of Proposition 3:** Recall from the text that $\mathbb{E}U_C = -0.5 \left[ \tilde{\eta}_1^B (1 - \tilde{\eta}_1^B) + \tilde{\eta}_0^B (1 - \tilde{\eta}_0^B) \right]$. Differentiate with respect to $x$, we have:

$$-0.5 \left[ (1 - 2\tilde{\eta}_1^B) \frac{\partial}{\partial x} \tilde{\eta}_1^B + (1 - 2\tilde{\eta}_0^B) \frac{\partial}{\partial x} \tilde{\eta}_0^B \right].$$

Recall from the proof of Proposition 1 that $B$’s message is informative, which implies that $\tilde{\eta}_1^B > 0.5$ while $\tilde{\eta}_0^B < 0.5$. Moreover, $\tilde{\eta}_1^B$ increases in $x, y_1$ while $\tilde{\eta}_0^B$ decreases in $x$. Thus $\mathbb{E}U_C$ increases in $x$. Even though $\tilde{\eta}_0^B$ increases in $y_1$, it is proportional to $2p_A - 1$, which is close to zero by assumption. Thus the $\mathbb{E}U_C$ increases in $x$. Similarly, $\mathbb{E}U_C$ increases in $y_1$.

We now examine how the truth-telling incentives of biased $B$ change with $p_A$ for a given $x$ and its effect on $C$. Simple calculations show that the derivative of biased agents’ distortion benefit, $\tilde{\eta}_1^B - \tilde{\eta}_0^B$, with respect to $p_A$, is positive and proportional to $N_{y_2} - N_{y_1}$. Differentiate biased $B$’s reputation cost with respect to $p_A$:

$$\frac{\partial \Delta_1}{\partial p_A} = \frac{\partial \Gamma_1}{\partial p_A} \left[ \tau_{0,0} - \tau_{1,0} - (\tau_{0,1} - \tau_{1,1}) \right] + \Gamma_1 \frac{\partial}{\partial p_A} \left( \tau_{0,0} - \tau_{1,0} \right) + (1 - \Gamma_1) \frac{\partial}{\partial p_A} \left( \tau_{0,1} - \tau_{1,1} \right).$$

Note that the first half is the information aggregation effect: it measures the change in biased $B$’s estimates of the true state due to $m_A$. Also, $\frac{\partial \Delta_1}{\partial p_A} < 0$ since as $p_A$ increases, biased $B$ believes more strongly in $\eta = 1$ if $m_A = 1$, and less so if $m_A = 0$. The second part is the reputation sensitivity effect: it measures how $B$’s
posterior objectivity responds to the possibility that biased $B$ may have used $m_A$. Specifically, the derivative of $\tau_{0,0}^B - \tau_{1,0}^B$ with respect to $p_A$ is negative and proportional to $N_{y_2} - N_{y_1}$; and that of $\tau_{0,1}^B - \tau_{1,1}^B$ with respect to $p_A$ is positive and proportional to $N_{y_2} - N_{y_1}$. By Proposition 2, if $\beta$ is sufficiently large, $N_{y_2} - N_{y_1}$ is sufficiently close to 0. In these cases, the information aggregation effect dominates. Using the implicit function theorem and biased $B$’s IC (1), $y_1$ decreases in $p_A$.

In a weak distortion equilibrium, $C$’s expected payoff changes both with $p_A$ and $N_{y_1}$:

$$\frac{d}{dp_A} EU_C = -0.5 \left[ (1 - 2\bar{\eta}_1^B) \frac{\partial \bar{\eta}_1^B}{\partial p_A} + (1 - 2\bar{\eta}_0^B) \frac{\partial \bar{\eta}_0^B}{\partial p_A} \right] - 0.5 \left[ (1 - 2\bar{\eta}_1^B) \frac{\partial \bar{\eta}_1^B}{\partial N_{y_1}} + (1 - 2\bar{\eta}_0^B) \frac{\partial \bar{\eta}_0^B}{\partial N_{y_1}} \right] \frac{dN_{y_1}}{dp_A}. \tag{7}$$

The first half of expression 7 is how $C$’s expected payoff changes directly with $p_A$, which is positive and proportional to $N_{y_2} - N_{y_1}$. The second half is the indirect effect through change in biased $B$’s truth telling. If $\beta$ is sufficiently high, the direct effect approaches 0. Also, from the result above, $y_1$ decreases in $p_A$, and thus the indirect effect dominates, which is negative. In this case, biased $B$ distorts more against his better signal and $C$’s expected payoff falls.

**Proof of Proposition 4:** For claim (1), by Proposition 1 and 2, biased $B$ and $A$’s possible binding incentive constraints are IC (1), (2) and (5). If the respective IC holds with equality, by the implicit function theorem, biased $A$’s truth-telling probability $x$ increases in $\alpha$ while biased $B$’s truth-telling probability $y_1$ increases in $\beta$ in a weak distortion equilibrium, and $y_2$ increases in $\beta$ in a strong distortion one.

For claim (2), suppose that $x > 0, y_2 > 0$ in a strong distortion equilibrium. Recall that biased $A$ and $B$’s incentive constraints IC (5) and IC (2) hold with equality in this case. These two ICs can be rewritten as: $\xi^*(x, y_2, \alpha) = 0$ and $\psi^*(x, y_2; \beta) = 0$. Denote $\xi_1^*, \xi_2^*, \xi_3^*$ as the partial derivative of $\xi^*$ with respect to $x, y_2$ and $\alpha$ respectively; and $\psi_1^*, \psi_2^*$ are similarly defined. Differentiate with respect to $\alpha$, we have:

$$\frac{dx}{d\alpha} = \frac{\xi_3^*\psi_2^*}{\xi_2^*\psi_1^* - \xi_1^*\psi_2^*}; \text{ and } \frac{dy_2}{d\alpha} = -\frac{\xi_3^*\psi_1^*}{\xi_2^*\psi_1^* - \xi_1^*\psi_2^*}.$$

Signs of some of the above partial derivatives are straightforward, namely, $\xi_1^* > 0, \psi_2^* > 0, \psi_1^* > 0, \xi_3^* < 0$. The key is the sign of $\xi_2^*$. Differentiate the LHS and the RHS of IC (5), then use biased $A$’s indifference condition $\xi^*(x, y_2, \alpha) = 0$ itself, we can show that if $y_2$ is sufficiently close to 0, which occurs if $\beta$ is sufficiently close to $\beta^*$, or if $\theta_B$ is sufficiently large, $\xi_2^* < 0$. Thus $y_2$ decreases in $\alpha$ while $x$ increases in $\alpha$.

For claim (3), suppose that $x > 0, y_1 > 0$ in a weak distortion equilibrium. Recall that biased $A$ and $B$’s incentive constraints IC (5) and IC (1) hold with equality in this case. Also, from the proof of Proposition 2,
we know that biased $A$’s best response decreases in $y_1$ and biased $B$’s best response decreases in $x$. Thus $x$ and $y_1$ are substitutes: $x$ increases in $\alpha$ while $y_1$ decreases in it; also $x$ decreases in $\beta$ while $y_1$ increases in it. If $\beta$ is sufficiently large, then $x = 0$. In this case, $y_1$ and $C$’s expected payoff doesn’t change with $\alpha$.

**Proof of Proposition 5**: Let the weights on biased $i, j$’s reputation be $\alpha_i$ and $\alpha_j$ and their signal qualities be $p_i, p_j$ respectively. Recall from the text that biased $i, j$ maximizes:

$$E_{\eta_j}\left[\Pr(\eta = 1|m_i, m_j)|s_i]\right] + E_{\eta_i}\left[\Pr(i = o|m_i, \eta)|s_i]\right].$$

To simplify notations, let $C$’s action given the messages be $\Pr(1|m_i, m_j)$. Then for biased $i$ to report $s_i = 0$ and $s_i = 1$ truthfully, the following two incentive constraints must hold:

$$\Pr(m_j = 1|s_i = 0)\left[\Pr(1|1, 1) - \Pr(1|0, 1)\right] + \Pr(m_j = 0|s_i = 0)\left[\Pr(1|1, 0) - \Pr(1|0, 0)\right]$$

$$\geq \alpha_i \sum_\eta \Pr(\eta|s_i = 0)\left[\Pr(i = o|m_i = 0, \eta) - \Pr(i = o|m_i = 1, \eta)\right];$$

$$\Pr(m_j = 1|s_i = 1)\left[\Pr(1|1, 1) - \Pr(1|0, 1)\right] + \Pr(m_j = 0|s_i = 1)\left[\Pr(1|1, 0) - \Pr(1|0, 0)\right]$$

$$\geq \alpha_i \sum_\eta \Pr(\eta|s_i = 1)\left[\Pr(i = o|m_i = 0, \eta) - \Pr(i = o|m_i = 1, \eta)\right].$$

Let $N_i \equiv \theta_i + (1 - \theta_i)x_i$, $N_j \equiv \theta_j + (1 - \theta_j)x_j$, then $C$’s action after receiving both messages is:

$$\Pr(1|0, 0) = \frac{(1 - p_i)(1 - p_j)}{p_ip_j + (1 - p_i)(1 - p_j)};$$

$$\Pr(1|0, 1) = \frac{(1 - p_i)(1 - (1 - p_j)N_j)}{(1 - p_i)(1 - (1 - p_j)N_j) + p_i[1 - p_jN_j]};$$

$$\Pr(1|1, 0) = \frac{(1 - p_j)(1 - (1 - p_i)N_i)}{(1 - p_j)(1 - (1 - p_i)N_i) + p_j[1 - p_iN_i]};$$

$$\Pr(1|1, 1) = \frac{[1 - (1 - p_j)N_j][1 - (1 - p_i)N_i] + [1 - p_jN_j][1 - p_iN_i]}{[1 - (1 - p_j)N_j][1 - (1 - p_i)N_i] + [1 - p_jN_j][1 - p_iN_i]}. $$

Moreover, because of the presence of objective agent, it is simple to show that $\Pr(1|1, 1) > \Pr(1|0, 1)$ and $\Pr(1|1, 0) > \Pr(1|0, 0)$. Next, it can also be shown that the difference $\Pr(1|1, 1) - \Pr(1|0, 1) - [\Pr(1|1, 0) - \Pr(1|0, 0)] > 0$. Intuitively, $m_i = 1$ has a higher marginal impact on $C$ if $m_j$ supports, rather than contradicts $m_i$, because $C$ believes that $\eta = 0$ is more likely if $m_j = 0$ than $m_j = 1$. This also implies that if IC (8) holds with equality or fails to hold, IC (9) holds strictly. Because if $s_i = 1$, it is more likely that $s_j = 1$ as well, thus the benefit from distortion is higher for biased $i$ to report $m_j = 1$ as opposed to $m_i = 0$. Moreover, the reputation cost is smaller for $i$ if $s_i = 1$ because $m_i = 1$ is more likely to be correct. Thus biased $i$ always
From the discussion above, the second line is negative. Also, 

There are twelve such ICs for biased 

\( \hat{\text{strategies be:}} \)

\( \text{distortion benefit is decreasing in} \)

\( \text{truthfully:} \)

\( m(1) \text{ There does not exist an equilibrium in which biased} \)

\( LHS \text{ of IC (8) with respect to} \)

\( s \) \( \text{with probability} \)

\( \text{defined as the probability} \)

\( \text{with probability one such that} \)

\( \text{be strictly between} \)

\( \text{max} \)

\( \text{hold, thus biased} \)

\( \text{reports} \)

\( \text{m} \text{B} \text{reports} \)

\( \text{with probability} \)

\( \text{B} \text{0} \text{B} \text{0} \)

\( \text{A} \text{A} = 0 \text{A} \text{A} = 0 \text{A} \text{A} = 0 \text{A} \text{A} = 0 \text{A} \text{A} = 0 \text{A} \text{A} = 0 \text{A} \text{A} = 0 \text{A} \text{A} = 0 \)

\[ \left[ 1 - (p_ip_j + (1-p_i)(1-p_j))N_j \right] \left[ \frac{\partial}{\partial N_j} \Pr(1|1,1) - \frac{\partial}{\partial N_j} \Pr(1|0,1) \right] \]

\[ - [p_ip_j + (1-p_i)(1-p_j)][\Pr(1|1,1) - \Pr(1|0,1) - [\Pr(1|1,0) - \Pr(1|0,0)]]. \]

From the discussion above, the second line is negative. Also, \( \frac{\partial}{\partial N_j} \Pr(1|1,1) < \frac{\partial}{\partial N_j} \Pr(1|0,1) \). Thus biased \( i \)'s distortion benefit is decreasing in \( x_j \). Together, \( x_i \) increases in \( x_j \), and vice versa.

**Appendix B. A Richer Message Space for Agent B**

In this section, \( C \) receives a vector of reports \((\hat{m}_A, m_B)\) from \( B \). By assumption, objective \( A \) and \( B \) report truthfully: \( m_A = s_A \) and \( \hat{m}_A = m_A, m_B = s_B \). Consider the following distortion strategies for biased \( A \) and \( B \), the first part of which is similar to those in the main model: biased \( A \) reports \( m_A = 1 \) if \( s_A = 1 \); and \( m_A = 0 \) with probability \( x \) if \( s_A = 0 \). Biased \( B \) reports \( m_B = 1 \) if \( s_B = 1 \). When \( s_B = 0 \), he reports \( m_B = 0 \) with probability \( y_2 \) if \( m_A = 0 \); and with probability \( y_1 \) if \( m_A = 1 \). The second part concerns how biased \( B \) distorts \( m_A \): he reports \( \hat{m}_A = 1 \) if \( m_A = 1 \). When \( m_A = 0 \), biased \( B \) reports \( \hat{m}_A = 0 \) with probability \( q_{00} \) if \( s_B = 0 \); and \( \hat{m}_A = 0 \) with probability \( q_{01} \) if \( s_B = 1 \). Given these distortion strategies, we have:

**Lemma 2:** (1) There does not exist an equilibrium in which biased \( B \) reports either \( m_A \) or \( s_B \) truthfully with probability one such that \( q_{00} = q_{01} = 1 \) or \( y_1 = y_2 = 1 \). (2) In any equilibrium, \( q_{00} \) and \( q_{01} \) cannot both be strictly between 0 and 1. Further, \( q_{00} \geq q_{01} \) and the inequality is strict if \( \max \{q_{00}, q_{01}\} > 0 \). (3) In any equilibrium, \( y_1 \) and \( y_2 \) cannot both be strictly between 0 and 1. Further, \( y_2 \geq y_1 \) and the inequality is strict if \( \max \{y_2, y_1\} > 0 \). (4) In any equilibrium, if \( q_{00} = 0 \), \( y_2 = 0 \).

**Proof:** For each \( m_A, s_B \), biased \( B \)'s truth telling requires that: for all \( m_A', s_B' \),

\[
\Pr(\eta = 1|m_A, s_B) + \beta \mathbb{E}_\eta[\Pr(B = o|m_A, s_B)|m_A, s_B] \geq \Pr(\eta = 1|m_A', s_B') + \beta \mathbb{E}_\eta[\Pr(B = o|m_A', s_B')|m_A, s_B].
\]

There are twelve such ICs for biased \( B \), three for each possible history. Recall that \( \Gamma_1, \Gamma_2, \Gamma_3 \) and \( \Gamma_4 \) were defined as the probability \( B \) believes that \( \eta = 0 \) given his private information. Let \( C \)'s actions given these strategies be: \( \hat{\eta}_{m_A,m_B}^B \equiv \Pr(\eta = 1|\hat{m}_A, m_B) \), then we have:

\[
\frac{1}{\hat{\eta}_{0,0}^B} = 1 + \frac{p_A p_B}{(1-p_A)(1-p_B)}; \quad \frac{1}{\hat{\eta}_{1,0}^B} = 1 + \frac{p_B [1 - p_A N_x + (1 - \theta_B) p_A N_x (1 - q_{00})]}{(1 - p_B) [1 - (1 - p_A) N_x + (1 - \theta_B) (1 - p_A) N_x (1 - q_{00})].}
\]

28
\[
\frac{1}{s_{0,1}} = 1 + \frac{p_A[(1 - \theta_B)(1 - \theta_B)q_00y_1 + p_B(1 - \theta_B)q_00(1 - y_2)]}{(1 - \theta_B)[(1 - \theta_B)q_00y_1 + p_B(1 - \theta_B)q_00(1 - y_2)]},
\]
\[
\frac{1}{s_{1,1}} = 1 + \frac{(1 - p_A)N_y + p_A(1 - \theta_B)[(1 - p_B)(1 - q_01) + p_B(1 - q_00)(1 - y_2)]}{(1 - (1 - p_A)N_y + p_A(1 - \theta_B)[(1 - p_B)(1 - q_01) + p_B(1 - q_00)(1 - y_2)]}
\]

Given the distortion strategies, \(s_{0,0}^B\) leads to the lowest action because it results from \(s_A = 0, s_B = 0\). Next, \(s_{1,0}^B\) increases in \(q_00\) because the more truthful biased \(B\) is in reporting \(m_A = 0\), the more likely \(\tilde{m}_A = 1\) comes from \(m_A = 1\). Also, \(s_{0,1}^B\) decreases in \(q_00\) and increases in \(q_01\): the higher is \(q_01\), the more likely that \(m_A = 0\) and \(s_B = 1\), which induces a higher action because \(s_B\) is more informative. Finally, \(s_{1,1}^B\) increases in \(q_00\), but may increase or decrease in \(q_01\). Because as \(q_01\) increases, \((1, 1)\) is less likely to come from \(m_A = 0, s_B = 1\).

If biased \(B\) distorts \(s_B\) a lot such that \(m_B = 1\) is not credible, then \((1, 1)\) may lead to a lower action than \((0, 1)\). Together, biased \(B\)'s distortion benefit, \(s_{1,0}^B - s_{0,0}^B\) and \(s_{1,1}^B - s_{0,0}^B\), increase in \(q_00\).

Let \(B\)'s posterior objectivity be: \(\tau_{m_A, m_B; \eta}^B \equiv \Pr(B = o|m_A, m_B, \eta)\), then we have:

\[
\tau_{0,0;0}^B = \frac{\theta_B}{\theta_B + (1 - \theta_B)q_00y_1}; \quad \tau_{0,0;1}^B = \frac{\theta_B}{\theta_B + (1 - \theta_B)q_00y_2};
\]
\[
\tau_{0,1;0}^B = \frac{\theta_B(1 - p_B)}{\theta_B(1 - p_B) + (1 - \theta_B)[(1 - p_B)q_01 + p_Bq_00(1 - y_2)]};
\]
\[
\tau_{0,1;1}^B = \frac{\theta_Bp_B + (1 - \theta_B)[p_Bq_01 + (1 - p_B)q_00(1 - y_2)]}{\theta_B(1 - p_A)N_y};
\]
\[
\tau_{1,0;0}^B = \frac{\theta_B(1 - p_A)N_y + (1 - \theta_B)[(1 - p_A)N_yq_1 + p_A(1 - q_01)y_2]}{\theta_B(1 - p_B)(1 - p_A)N_y};
\]
\[
\tau_{1,0;1}^B = \frac{\theta_B(1 - p_A)N_y + (1 - \theta_B)[(1 - p_A)N_yq_1 + p_A(1 - q_01)y_2]}{\theta_B(1 - p_B)(1 - p_A)N_y};
\]
\[
\tau_{1,1;0}^B = \frac{\theta_B(1 - p_A)N_y + (1 - \theta_B)[(1 - p_A)N_yq_1 + p_A(1 - q_01)y_2]}{\theta_B(1 - p_B)(1 - p_A)N_y};
\]
\[
\tau_{1,1;1}^B = \frac{\theta_B(1 - p_A)N_y + (1 - \theta_B)[(1 - p_A)N_yq_1 + p_A(1 - q_01)y_2]}{\theta_B(1 - p_B)(1 - p_A)N_y};
\]

Note that the difference in biased \(B\)'s posterior objectivity \(\tau_{m_A, m_B; \eta}^B - \tau_{m_A, m_B; \eta}^B\) decreases in \(q_00, q_01\); and the difference \(\tau_{m_A, m_B; \eta}^B - \tau_{m_A, m_B; \eta}^B\) decreases in \(y_1, y_2\). Intuitively, the more truthful biased \(B\) in reporting \(m_A\) (\(s_B\)), the less his reputation depends on \(\tilde{m}_A\) (\(m_B\)).

If \(m_A = 0, s_B = 0\), the following truth-telling ICs need to hold for biased \(B\):

\[
\tau_{0,0;0}^B - \tau_{0,0;1}^B \leq \beta \left[\tau_{0,0;0}^B - \tau_{1,0;0}^B\right] \quad \tau_{1,0;0}^B - \tau_{1,0;1}^B \leq \beta \left[\tau_{1,0;0}^B - \tau_{1,0;1}^B\right]
\]
\[
\tau_{0,1;0}^B - \tau_{0,1;1}^B \leq \beta \left[\tau_{0,1;0}^B - \tau_{1,1;0}^B\right] \quad \tau_{1,1;0}^B - \tau_{1,1;1}^B \leq \beta \left[\tau_{1,1;0}^B - \tau_{1,1;1}^B\right]
\]
Other incentive constraints are similar and thus left out.

For Claim (1), suppose that $q_{00} = 1, q_{01} = 1$, then biased $B$’s reputation cost depends only on $y_1$ and $y_2$. Moreover, for any message $m_B$, it can be shown that $\Pr(\eta = 1|1, m_B) > \Pr(\eta = 1|0, m_B)$: biased $B$ induces a higher action by reporting $m_A = 1$ with no reputation cost, a contradiction. Thus if $m_A = 0$, biased $B$ never reports $m_A = 0$ truthfully with probability 1. Similarly, he never reports $s_B = 0$ truthfully with probability 1.

For Claim (2), suppose that $q_{00} \in (0, 1)$, there are three cases. If $y_2 = 0$, then biased $B$ is indifferent between reporting $\tilde{m}_A = 0, m_B = 1$ and $\tilde{m}_A = 1, m_B = 1$: $EU_B(0, 1) = EU_B(1, 1)$ if $m_A = 0, s_B = 0$. Because $\eta = 0$ is more likely when $m_A = 0, s_B = 0$ than $m_A = 0, s_B = 1$ ($\Gamma_2 > \Gamma_3$), $EU_B(1, 1) > EU_B(0, 1)$ if $m_A = 0, s_B = 1$, and thus $q_{01} = 0$. If $y_2 \in (0, 1)$, then it must follow that all three ICs of biased $B$ hold with equality, which implies that $EU_B(1, 1) > EU_B(0, 1)$ if $m_A = 0, s_B = 1$, and thus $q_{01} = 0$. If $y_2 = 1$, then biased $B$ is indifferent between reporting $\tilde{m}_A = 0, m_B = 0$ and $\tilde{m}_A = 1, m_B = 0$ if $m_A = 0, s_B = 0$. However, because $y_1 < 1$ in this case, $EU_B(0, 1) < EU_B(1, 0) \leq EU_B(1, 1)$ if $m_A = 1, s_B = 0$. Because $\eta = 1$ is more likely when $m_A = 0, s_B = 1$ than $m_A = 1, s_B = 0$ ($\Gamma_3 > \Gamma_1$), $q_{01} = 0$. Thus $q_{00} = 0$ if $q_{00} \in (0, 1)$. If $q_{01} \in (0, 1)$, then $EU_B(0, 1) = EU_B(1, 1)$ if $m_A = 0, s_B = 1$. In that case, however, $EU_B(0, 1) > EU_B(1, 1)$ if $m_A = 0, s_B = 0$, and thus $q_{00} = 1$. Thus $q_{00} > q_{01}$ if $\max\{q_{00}, q_{01}\} > 0$.

For Claim (3), suppose first that $y_2 = 0$, then biased $B$’s expected payoff $EU_B(1, 0) < EU_B(0, 1) \leq EU_B(1, 1)$. Because $\eta = 1$ is more likely when $m_A = 1, s_B = 0$ than when $m_A = 0, s_B = 0$ ($\Gamma_2 > \Gamma_1$), it implies that $EU_B(1, 1) > EU_B(1, 0)$ as well, and thus $y_1 = 0$. If $y_2 \in (0, 1)$, then $q_{00} = 1$. Suppose $q_{00} < 1$ instead, then it follows that all three ICs of biased $B$ hold with equality. But this is impossible because in this case $y_1 = 0, q_{01} = 0$, and thus there are only two mixing probabilities to be determined in equilibrium and three equalities to hold. If $q_{00} = 0$ instead, then from the proof of Claim (2), $q_{01} = 0$. Then $EU_B(1, 0) \leq EU_B(1, 1)$ if $m_A = 0, s_B = 0$, thus $y_1 = 0$. If $y_2 = 1$, from the proof of Claim (1), $y_1 < 1$ because truthful revelation of $s_B = 0$ is impossible.

For Claim (4), suppose that $q_{00} = 0$ and $y_2 > 0$, then from Claim (2), $q_{01} = 0$. Thus if $m_A = 0, s_B = 0$, $EU_B(0, 1), EU_B(1, 1) < EU_B(1, 0)$. However, if $q_{00} = 0, q_{01} = 0$, biased $B$ is believed to be objective for sure if he reports $\tilde{m}_A = 0$, and thus message $(0, 1)$ induces a higher action, and a better reputation, than $(1, 0)$, a contradiction. Thus $y_2 = 0$ if $q_{00} = 0$, which also implies that $y_1 = 0$ from Claim (3). $\square$

Lemma 2 shows that biased $B$ still uses both pieces of information before choosing what to report. Also, he is more truthful if the state is more likely to be 0 given his information. The following result shows that a
distortion equilibrium similar to that in the main model exists.

**PROPOSITION 6:** A distortion equilibrium exists. (1) In equilibrium, biased $A$ reports $m_A = 0$ with probability $x \in [0, 1]$ if $s_A = 0$. (2) When $\beta$ is sufficiently low, biased $B$ always reports $m_B = 1$. Further, if $\theta_B$ is sufficiently close to 1, biased $B$ always reports $\tilde{m}_A = 1, m_B = 1$: $q_{00} = 0, q_{01} = 0$; otherwise $q_{00} > 0, q_{01} = 0$. (3) When $\beta$ is moderate, then $q_{00} = 1, q_{01} \in [0, 1)$ and $y_1 = 0, y_2 \in (0, 1)$. (4) When $\beta$ is sufficiently high, then $q_{00} = 1, q_{01} \in [0, 1)$ and $y_1 \geq 0, y_2 = 1$.

**Proof:** Biased $A$’s incentives to distort depend on the pivotal event when his message changes $C$’s action. In contrast with the main model, biased $A$’s message reaches $C$ truthfully with positive probability in expectation.

To see this, suppose that $s_A = 0$, then biased $A$’s influence on $C$’s action if he reports $m_A = 1$ versus $m_A = 0$ is at least: 

$$\theta_B(p_A p_B + (1 - p_A)(1 - p_B))(\hat{\eta}_{1,0}^B - \hat{\eta}_{0,0}^B) + \theta_B(p_A(1 - p_B) + (1 - p_A)p_B)(\hat{\eta}_{1,1}^B - \hat{\eta}_{0,1}^B).$$

For the same reason, biased $A$ also pays a reputation cost. Let $A$’s posterior objectivity be: $\tau_{m_A, m_B, \eta}^A \equiv \Pr(A = o|m_A, m_B, \eta)$. Then we have:

$$\tau_{0, m_B, \eta}^A = \frac{\theta_A}{\theta_A + (1 - \theta_A)x};$$

$$\tau_{1, m_B, \eta}^A = \frac{\theta_A[1 - p_A + p_A(1 - \theta_B)(1 - q_{00})]}{\theta_A[1 - p_A + p_A(1 - \theta_B)(1 - q_{00})] + (1 - \theta_A)[1 - p_A x + p_A x (1 - \theta_B)(1 - q_{00})]};$$

$$\frac{1}{\tau_{1, m_B, \eta}^A} = 1 + \frac{\theta_A[(1 - p_A) + p_A(1 - \theta_B)(1 - q_{00})]}{\theta_A[(1 - p_A) + p_A(1 - \theta_B)] + (1 - p_A x + p_A x (1 - \theta_B) (1 - q_{00}))};$$

$$\frac{1}{\tau_{1, m_B, \eta}^A} = 1 + \frac{\theta_A(p_A + (1 - p_A)(1 - x) + p_A x (1 - \theta_B) (1 - y_2) + p_B(1 - q_{00}))}{\theta_A[p_A x + (1 - p_A)(1 - \theta_B) (1 - y_2) + p_B(1 - q_{00})]}.$$ 

Biased $A$’s reputation cost decreases in $x$. Argument similar to the main model can show that biased $A$ always distorts $s_A = 0$ to some extent: $x \in [0, 1)$. In particular, note that biased $A$ now has influence even in the case when biased $B$ completely distorts and thus he has no influence in the main model.

For biased $B$, first consider whether a complete distortion equilibrium exists in which he always reports $\tilde{m}_A = 1, m_B = 1$, or $q_{00} = 0, q_{01} = 0$ and $y_1 = 0, y_2 = 0$. If $\beta$ is sufficiently close to 0, biased $B$ is primarily concerned about $C$’s action. From Lemma 2, $q_{00} = 0$ is sufficient to guarantee such an equilibrium to exist. To begin with, simple algebra can show that at $q_{00} = 0, q_{01} = 0, \hat{\eta}_{0,1}^B > \hat{\eta}_{1,0}^B$. This is an important new constraint: biased $B$ can improve his posterior objectivity by reporting $\tilde{m}_A = 0, m_B = 1$ without inducing a low action, because $s_A$ is very uninformative. Next, we need to compare $\mathbb{E}U_B(1, 1)$ with $\mathbb{E}U_B(0, 1)$. There are two cases. First, $\hat{\eta}_{1,1}^B > \hat{\eta}_{1,0}^B$ if $(2p_A - 1)p_B(1 - p_B)\theta_B - (p_B - p_A)(1 - \theta_B) > 0$, which holds if $\theta_A$ is sufficiently close
to 1, or if \( p_B \) is sufficiently close to \( p_A \). In this case, the cutoff \( \beta \) for a complete distortion equilibrium is:

\[
\beta^{ac} (1 - \theta_B) \left[ \frac{\Gamma_2}{\theta_B (1 - p_B) (1 - p_A N_x) + 1 - \theta_B} + \frac{1 - \Gamma_2}{\theta_B p_B (1 - (1 - p_A) N_x) + 1 - \theta_B} \right].
\]

Observe that here, biased \( B \)'s report \((0, 0)\) induces a lower action than \( m_B = 0 \); while \((1, 1)\) may induce a higher action than \( m_B = 1 \). However, his expected reputation cost also increases, thus the cutoff \( \beta^{ac} \) may be higher or lower than the cutoff \( \beta^c \).

Second, if \( \hat{\eta}^{B}_{1,1} \leq \hat{\eta}^{B}_{1,0} \), a complete distortion equilibrium does not exist because the incentive to deviate to \( \tilde{m}_A = 0, m_B = 1 \) is too strong. This occurs if \( \theta_A \) is sufficiently bounded away from 1, or if \( p_B \) is sufficiently close to 1. In this case, because \( \hat{\eta}^{B}_{1,1} \) increases in \( q_{00} \) and \( \hat{\eta}^{B}_{1,0} \) decreases in \( q_{00} \); also because the reputation cost decreases in \( q_{00} \), there exists a \( q_{00} \in (0, 1) \) such that \( \hat{\eta}^{B}_{1,1} = \hat{\eta}^{B}_{1,0} \). If at this \( q_{00} \), \( \hat{\eta}^{B}_{1,0} < \hat{\eta}^{B}_{1,1} \), then there exists a cutoff value of \( \beta^{nc} > \beta^{ac} \) such that for \( \beta \in (\beta^{nc}, \beta^{ac}) \), \( q_{00} > 0, q_{01} = 0, y_1 = 0, y_2 = 0 \) is an equilibrium.

If \( \beta > \beta^{nc} \), then in any equilibrium, \( y_2 > 0 \). From Lemma 2, \( q_{00} = 1 \). This implies that if \( m_A = 0, s_B = 0, \) \( \mathbb{E}U_B(0, 0) \geq \mathbb{E}U_B(0, 1) \). If \( y_2 \in (0, 1) \), then \( \mathbb{E}U_B(1, 0) < \mathbb{E}U_B(0, 1) \) at \( m_A = 1, s_B = 0 \), thus \( y_1 = 0 \). There exists a cutoff value \( \beta^{mcs} \) such that \( \mathbb{E}U_B(0, 0) = \mathbb{E}U_B(0, 1) \) at \( y_2 = 1, y_1 = 0 \). If \( \beta \in (\beta^{mc'}, \beta^{mcs}) \), \( q_{00} = 1, q_{01} \in [0, 1), y_2 \in (0, 1), y_1 = 0 \) is an equilibrium. This is the counterpart of the strong distortion equilibrium in the main model. If \( \beta \) is sufficiently large, then if \( m_A = 0, s_B = 0, \) \( \mathbb{E}U_B(0, 0) > \mathbb{E}U_B(0, 1) \) at \( y_2 = 1, y_1 = 0 \). In this case, \( q_{00} = 1, q_{01} \in [0, 1), y_2 = 1 \) and \( y_1 \in (0, 1) \) is an equilibrium, which is the counterpart of the weak distortion equilibrium in the main model. □

References


