American Economic Journal: Microeconomics 2009, 1:2, 220–242
http://www.aeaweb.org/articles.php?doi=10.1257/mic.1.2.220

Corruptible Advice†

By Erik Durbin and Ganesh Iyer*

We study information transmission to a decision maker from an advisor who values a reputation for incorruptibility in the presence of a third party who offers unobservable payments/bribes. While it is common to ascribe negative effects to such bribes, we show that given reputational concerns, bribes can play a positive role by restoring truthful communication that would otherwise not occur. Thus, while bribes can influence self-interested bad advisors to lie about the unfavorable state, they can also be used to motivate good advisors who care more about the decision maker's utility to truthfully report the favorable state. (*JEL* D82, D83)

Many decisions in market settings depend upon the information received from informed advisors. We consider a problem in which a decision maker consults an informed advisor about taking an action that affects a third party. The advisor has some information that is relevant for the optimal decision, and, all else equal, has preferences that are aligned with those of the decision maker. However, the third-party's preferences may differ from those of the decision maker, and the third party might, therefore, attempt to influence the advisor's report by offering a payment that is contingent on that decision. The possibility of pecuniary side payments or bribes to the advisor influences the credibility of the advisor's report or the corruptibility of the advice provided to the decision maker.

For example, expert advice plays an important role in medical decisions. Patients may consult specialists about the best course of treatment for a particular condition. Pharmaceutical companies benefit from the extent to which their drugs are recommended by specialist doctors. Therefore, they may offer gifts, grants, travel, or other benefits to doctors to encourage them to prescribe their drugs, and in some cases these benefits may be implicitly or explicitly tied to the level of prescription of a particular drug. Many industry observers have argued that the proliferation of such promotional efforts may create conflicts of interests and may sometimes lead to inappropriate prescription behavior.[1] There are many other examples of situations

* Durbin (e-mail: edurbin@ftc.gov); Iyer: Haas School of Business, University of California, Berkeley, Berkeley, CA 94720-1900 (email: giyer@haas.berkeley.edu). We thank Liang Guo, Dmitri Kuksov, John Morgan, Marko Tervio, V. Srinivasan, and various seminar participants for comments.

† To comment on this article in the online discussion forum, or to view additional materials, visit the articles page at: http://www.aeaweb.org/articles.php?doi=10.1257/mic.1.2.220.

[1] See http://www.nofreelunch.org/ for examples of such promotional efforts. As another example, Italian prosecutors recently accused GlaxoSmithKline of offering incentives to doctors to prescribe certain drugs, including cash payments that were tied to the number of patients treated with the drug (John Hooper and Heather Stewart, "Over 4,000 Doctors Face Charges in Italian Drugs Scandal," *The Guardian*, April 27, 2004, 2).

in which an outside party would like to influence the communication between an advisor and a decision maker:

- Policymakers often rely on consultants, academic experts, or government employees to recommend a course of action that may create benefits for third parties, such as with procurement contracts. Third parties with a stake in the decision may try to influence the advice provided to policy makers.[2]
- Stock analysts who advise investors may be influenced by the companies whose securities they analyze.[3]
- Product reviewers or movie critics may receive undisclosed incentives from the companies whose products they review. For example, the technology editor for NBC's "Today" show admitted that he had accepted payments from Apple, Sony, and other electronics companies to promote their products on news shows (see Howard Kurtz, "Firms paid tech gurus to promote their products," *Washington Post,* April 20, 2005, C1).

Common across these situations is the idea that bribes or side payments made by third parties would have negative effects because they may influence the advisor to misreport and prevent the truthful transmission of information. It is this well-known information-corrupting effect of third-party side payments that is subject to criticism by commentators in the examples described above. Contrary to this criticism, a goal of this paper is to suggest that when advisors have reputational concerns, payments offered by third parties may play a positive role by promoting truthful information transmission.

Specifically, we look at a model with the following characteristics:

- A decision maker consults an advisor about a 0-1 decision (where "1" is the preferred decision of the third party).
- A third party stands to benefit if the decision maker takes a positive decision.
- The third party can offer the advisor a payment that is conditional on the decision maker's action.
- The advisor cares about his reputation for not being corruptible.[4]
- The payment to the advisor is not (perfectly) observable to the decision maker.

In the example of medical advice, the third party represents a pharmaceutical company that can offer unobservable incentives that depend on whether the doctor's

[2] Such influence might range from financial support for academic research to outright bribery of bureaucrats. As an example of the latter, in 2004, a senior Air Force official was sentenced to jail for having steered defense contracts toward Boeing before accepting a high-paid position as a Boeing executive (Leslie Wayne, "A growing military contract scandal," *New York Times*, October 8, 2004, sec. C1, 2).

[3] For example, in 2003 ten major investment banks paid $1.4 billion to settle Securities and Exchange Commission (SEC) charges that their analysts' investment advice was "fraudulent," and, specifically, that they had encouraged analysts to offer positive analysis of companies in order to win those companies' investment banking business (Jeffrey Krasner, "$1.4B Wall St. Settlement Finalized," *Boston Globe*, April 29, 2003, D1).

[4] It must be noted that this is not necessarily the same as a reputation for accuracy. Corruptibility in this paper pertains to the advisor being perceived by the decision maker as being susceptible to influence by the payments/bribes that the third party might offer. A reputation for accuracy implies that the advisor cares about whether his message was consistent with the ex post realization.

patients use the company's drug—a dependence that may be implicit, but has, in at least some cases, been made explicit. The possibility of such unobservable payments creates a conflict of interest situation. The doctor's professional credibility can be reduced if patients come to believe that his advice is influenced by the pharmaceutical company, so he wishes to be seen as making recommendations that reflect only his own judgment about the best interests of the patient. However, patients cannot observe payments from the pharmaceutical company, so the doctor cannot necessarily maintain a reputation for incorruptibility by simply refusing to accept payments from the pharmaceutical company.

Cheap-talk models of advice in the tradition of Vincent P. Crawford and Joel Sobel (1982) typically involve an informed advisor and a decision maker who shares the advisor's objectives to a certain degree. Typically in such models, the advisor has a bias relative to the decision maker's preferred outcome, and this bias prevents the advisor from fully revealing the information he has. We present a model in which the preferences of the advisor are not inherently biased relative to those of the decision maker. Rather, the bias arises endogenously from the possibility of side payments offered by the third party. Thus, in the absence of the third-party payments, there is no bias in the model of this paper, and the advisor would prefer the same choice as the decision maker. Nevertheless, the advisor may not be able to fully communicate what he knows because the decision maker will be concerned that the advisor's report has been corrupted by a payment from the third party.

Reputation effects play a role in making an advisor desire not to appear "corruptible" in the message he conveys. Suppose that advisors differ in the weight they place on the decision maker's utility with good advisors placing a greater weight than bad advisors. Then, if the advisor values a future relationship with the decision maker, the advisor would like the buyer to believe that he attaches a greater weight to the buyer's utility, and is therefore not easily influenced by third party side payments. This reputational motive means that the advisor will hesitate to recommend the action that the third party prefers, for fear that the decision maker will think the recommendation is motivated by a side payment. This means that the advisor's concern for reputation affects communication. Depending on the parameters, the case where the advisor cares about reputation can lead to more information transmission or less information transmission compared to the case without reputational concerns. Specifically, the case with reputational concerns can enhance (reduce) information transmission, compared to the case without reputational concerns, if the cost of lost reputation is higher (lower) for the good advisor than it is for the bad advisor.

An interesting implication is that the information loss that arises from the advisor's desire to be perceived as incorruptible may motivate the third party to offer a payment to the advisor even when the third party's interests are aligned with those of the decision maker. For example, a pharmaceutical firm may offer doctors benefits even when it knows that its treatment is of high quality and in the patient's best interest. In other words, it is not only the low-type firms that will attempt to offer a side payment to the advisor as a "bribe" for lying about quality. But, even high quality firms can offer payments in response to the reputational concerns of advisors created by the endogenous actions of the firm. Thus third-party payments might not only have the function of influencing the bad advisor to lie about low quality, but, interestingly, they may also motivate the good advisor to correctly report high quality.

An associated implication pertains to the motivations of the advisors. It is obvious that "bad" advisors who care more about their pecuniary self-interest would have the incentive to misreport low quality in order to collect the payment. However, the notable point is that even "good" advisors might actually misreport a high quality product as one of low quality given the possibility of side payments. While bad advisors would be more susceptible to the bribes, and may lie and misreport a bad state of the world as good, good advisors, who care relatively more about the decision maker's utility, might lie about the good state of the world by misreporting the good state as bad. This bias arises because reporting the state as bad enhances the advisor's reputation. This is similar to the effects described in the papers on bad reputation (Jeffrey C. Ely and Juuso Valimaki 2003; Ely, Drew Fudenberg, and David K. Levine 2008) and the political correctness effect described in Stephen Morris (2001). However, in this paper, the bias arises from the endogenous incentives of the third party who strategically chooses side payments. Consequently, our paper can be seen as providing a theory of when reputation may be bad based on the equilibrium incentives of third-party market participants.

I. Related Research

Our paper is related to the cheap talk literature initiated by Crawford and Sobel (1982) involving strategic communication between an advisor and a decision maker when the advisor's preferences are inherently different from those of the decision maker. Sobel (1985) introduces reputation effects in a cheap talk model. The decision maker is uncertain about the advisor's preferences, so that the advisor's past reports determine his future credibility. It is assumed that good advisors have interests that are aligned with those of the decision maker, while bad advisors have opposing interests. This leads bad advisors to sometimes invest in reputation by telling the truth so that the reputation may later be exploited. Roland Benabou and Guy Laroque (1992) extend Sobel's analysis to the case where advisors have noisy private signals.[5] Unlike in the cheap talk models, advisor preferences are such that the advisor prefers the same choice as the decision maker. The bias in communication arises endogenously from the motives of the third party to influence the message resulting in third-party payments. Consequently, this paper characterizes the nature and effects of these endogenous payments made by the third party.

The work is also related to Ely and Valimaki (2003) and Ely, Fudenberg, and Levine (2008) who show that reputational concerns for a long-lived player interacting with a sequence of short-term players can be unambiguously bad, leading to loss of surplus. This arises if the long-term player's reputational incentive to separate from the bad type leads to actions that also hurt the short-term players, creating

[5] Vijay Krishna and John Morgan (2001) extend this literature to the case of multiple advisors and show that eliciting advice from multiple advisors sequentially is beneficial only when they are biased in opposite directions. Joseph Farrell and Robert Gibbons (1989) model cheap talk when there are multiple decision-making audiences, and the possibility of private or public communication, to show how the presence of one audience may discipline the information transmission to the other.

surplus loss. Indeed, when the reputational incentive to separate from the bad type is substantial, the potential surplus loss can be significant enough to induce market failure by inducing the short-lived players to not participate. Our paper uncovers a reputational incentive in a communication/advice game that may induce good and bad advisors to misreport to decision makers. This arises because of the presence of a third-party market participant who strategically chooses side payments to influence the advice.

In the advice literature, a paper by Morris (2001) shows how reputational concerns can generate perverse incentives for a good advisor who wishes to separate from a bad advisor. If a bad advisor is biased toward a certain message, then a good advisor may avoid sending that message, even when it is accurate, to avoid damaging his reputation.[6] Another strand of the literature focuses on the case where the advisor has no interest in how his advice affects decision making, but only cares about its impact on his reputation for accuracy. David S. Scharfstein and Jeremy C. Stein (1990) show that this can lead managers to have an incentive to say the expected thing and indulge in herd behavior, and this leads to information loss. Marco Ottaviani and Peter Norman Sorensen (2006a, b) analyze the reporting of private information by an expert who has exogenous reputational concerns for being perceived as having accurate information, and investigate the nature of the information loss in this setting. In contrast to this literature, our advisor does not care about a reputation for accuracy, but rather a reputation for incorruptibility. While an absolutely incorruptible advisor would always report accurately, in general, these two types of reputational incentives do not coincide. Specifically, we show that the advisor may well send an inaccurate message in order to bolster his reputation for incorruptibility.

II. The Model

We start by describing a basic one-shot model of advice involving three players: the decision maker, indexed by D; the advisor (A); and the third-party (T). Define the payoffs of these three as U^D, U^A, and U^T, respectively. There are two possible states of the world: "high" and "low" ($s \in \{l, h\}$). There are two possible types of advisors: "good" and "bad" ($t \in \{b, g\}$). We assume that D does not observe either s or t, whereas A observes both.[7] For the main analysis, we assume that the third party observes s, and is uninformed of the advisor type t, but we also comment on the effect that alternative assumptions on T's information about s and t has on communication in Section IVB.

The decision maker makes a "yes" or "no" decision $d \in \{0,1\}$, where $d = 1$ is interpreted as a "yes." If the decision maker chooses $d = 1$, then he receives a payoff in a manner that will be made precise below. However, if he chooses $d = 0$, he gets

[6] Morgan and Phillip C. Stocken (2003) examine the problem of stock analysts whose payoffs may be driven by a benefit that comes from inflating the stock prices as well as by a cost due to bad performance. The analyst's interests may either be aligned with the investor or be misaligned because the analyst likes to have a higher stock price than what is warranted by the information. The paper shows that an analyst with fully aligned interests may be able to credibly convey bad news but not good news.

[7] In other words, we assume that the advisor observes something (the state) that is relevant to the decision maker's decision but is unobservable to the decision maker.

a reservation value of 0. If the decision maker goes forward and chooses $d = 1$, the decision will lead to either "success" or "failure," and the probability of success depends on the state of the world which is unobserved by D. The value of a success for the decision maker is $G > 0$, and a failure is $-L < 0$. The probability of success depends on the state of the world, and is given by θ_s, with $0 < \theta_l < \theta_h < 1$. Note that the state cannot be perfectly inferred by the decision maker after the action is taken because $\theta_h < 1$ and $\theta_l > 0$. Given s, the expected utility for D from choosing $d = 1$ is $\pi_s = \theta_s G - (1 - \theta_s)L$ and the expected utility for D from choosing $d = 0$ is the reservation value 0. We will assume that $\pi_h > 0$ and $\pi_l < 0$ and, as argued later, this is necessary for the advisor's communication to be decisive and nontrivial. The decision maker's expected utility, given s, can then be denoted by $U^D = \pi_s d$. The ex ante probability that $s = h$ is assumed to be $1/2$, and this is common knowledge. If the decision maker chooses $d = 1$, then the third party's utility increases by w. Depending upon the state, the third party can choose to make a payment $q_s \geq 0$ to the advisor, which is contingent on $d = 1$. This payment q_s is unobservable to the decision maker. The third-party's payoffs are given by $U^T = [w - q_s]d$.

Nature draws the state s and reveals it to the advisor. The advisor's utility is a function of his own wealth as well as D's utility. The advisor's type is the weight he attaches to wealth relative to D's utility. Specifically, the utility function of an advisor of type t's is given by

$$U_t^A = \alpha_t U^D + q_s d. \tag{1}$$

The parameter α_t represents the weight that the advisor attaches to the decision maker's utility. Let $\alpha_g > \alpha_b$. That is, a good advisor is less easily influenced by a monetary payment than a bad advisor. This can also be interpreted as the bad advisor being less altruistic than a good advisor. The prior probability that the advisor is good is common knowledge and is given by $\lambda \in (0,1)$. We first consider this one-shot game that involves no reputational effects, where communication is only affected by α_t. Subsequently, in Section IV, we analyze the model in which the advisor is motivated by reputational effects and may receive second-period continuation payoffs based on reputation.

The timing of the game in period one is shown in Figure 1. Nature determines the state of the world $s \in \{l,h\}$ and advisor type $t \in \{b,g\}$. The state is revealed to the advisor but not to the decision maker. In the next stage, the third party decides the level of the contingent payment q_s to the advisor (which can possibly be zero). Note that the payment q_s is dependent on the state because the third party observes s. Next, the advisor of type t sends the message to the decision maker denoted by $m_t \in \{l,h\}$. In the fourth stage, D makes the decision $d \in \{0,1\}$. Finally, D updates beliefs about the advisor's type, and the payoffs of all the players in the game are realized.

We will look for a perfect Bayesian equilibrium (PBE) of the game just described. A PBE of this game is given by $\{q_s, m_t(s,q_s), \mu(m), d(\mu)\}$, where $\mu(m) = \Pr(s = h \mid m)$ is D's belief about the state conditional on m.[8] In a PBE, all players' strategies are

[8] As we will show, the equilibrium of this one-shot game of communication is in pure strategies. However, for the game with reputational effects that we analyze in Section IV, it is possible for the advisor's equilibrium reporting to be in mixed strategies.

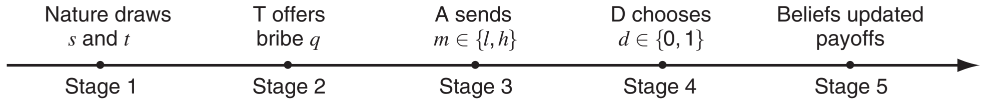

Figure 1. Timing of the Model

optimal given $\mu(m)$, and $\mu(m)$ is determined by Bayes rule wherever possible. Note that this is a cheap talk game, meaning that the message sent by the advisor does not directly affect the advisor's payoff (nor that of any other player). Therefore, there will always be "babbling" equilibria, in which the decision maker ignores the advisor's report, and because of this, the advisor makes a report uncorrelated with his signal. We look for equilibria in which the advisor's message is *informative* and *decisive*, where informative means that the message conveys (at least some) information about the state over and above the prior, and decisive means that D's action depends on the message. The advisor's message is informative if the conditional probability that the state is h, given that the advisor's message is h, is greater than the unconditional probability of the same event. It is possible for the message to be informative, but not decisive, if D's optimal action never depends on what he learns from the advisor. The approach to identifying the equilibrium will be to describe D's beliefs given the reporting strategy of A, then check whether this reporting strategy is consistent with T's strategy and with D's actions that would be induced.

III. Preliminaries: The Game Without Reputation Effects

We start with a brief analysis of the model above without reputation effects, as it serves as a baseline for the analysis with reputation effects. Given that $0 < \theta_l < \theta_h < 1$, it is possible that there will be a failure even if $s = h$ and a success even if $s = l$. Because we are interested in communication of the advisor's information, we are interested in the case in which the advisor's message is decisive in that it has the potential to influence D's choice.

DEFINITION 1: *A's message is decisive if D's optimal strategy is to choose $d = 1$ if and only if $s = h$.*

A's message can be decisive only if D's optimal choice under full information depends on the state. That means we will restrict attention to the case where $\pi_l = \theta_l G - (1 - \theta_l)L < 0 < \theta_h G - (1 - \theta_h)L = \pi_h$.[9] Given this, D's optimal choice is to choose $d = 1$ if and only if $s = h$, so she will choose $d = 1$ whenever the assessed probability of the state being h is close enough to one. The analysis begins with the following lemma which notes that in any informative and decisive equilibrium, there are no mixed strategy equilibria (all proofs are in the Appendix).

[9] Note that if $\pi_l > 0$ ($\pi_h < 0$), then D should always (never) choose $d = 1$ regardless of A's information, and the advisor's message becomes irrelevant.

LEMMA 1: *In any informative and decisive equilibrium, the type t advisor chooses* $m_t(h) = h$ *with probability one and* $m_t(l) = l$ *with probability one (except possibly in the knife-edge case of* $w = -\alpha_t \pi_l$*).*

Thus, the PBE will be one in which the advisor chooses a pure strategy upon observing the state. The PBE with honest communication is characterized in the proposition below.

PROPOSITION 1: *There is an informative and decisive PBE in which A is honest irrespective of his type, and D believes A's report, if and only if* $w \leq -\alpha_b \pi_l$*. In this equilibrium, there are no bribes (i.e.,* $q_h = q_l = 0$*).*

If the payment is zero and D believes A's report, then A will be honest. In the absence of bribes, the advisor's preferences are such that the advisor will prefer the same choice as the decision maker. So, there is nothing to prevent the advisor from communicating his information. But this will break down if T is willing to offer a large enough bribe to change A's report irrespective of the advisor type. Suppose that $s = l$, and that A's report is honest and decisive. The advisor would be willing to change a negative report to a positive one if $\alpha_t \pi_l + q_l > 0$, or if $q_l > -\alpha_t \pi_l$.

If A's message is decisive, T would be willing to offer a bribe as high as w to change the advisor's report. The third party could either offer a bribe that convinces a bad advisor to change his report or a larger bribe that convinces a good advisor to change his report. For honest communication to be an equilibrium, neither must be appealing, and so the relevant condition is that T does not want to offer a bribe that the bad advisor would accept. Thus, honest communication cannot be part of a PBE if $w > -\alpha_b \pi_l$. Therefore, a PBE in which both types of A are honest, and D believes A's report, exists if and only if $w \leq -\alpha_b \pi_l$.[10] In other words, honest communication is an equilibrium when the decision maker's loss from a "failure" is large, relative to the gain from a "success." For example, if a doctor is advising on a new drug treatment that carries relatively significant side effects compared to its potential therapeutic benefits, it will be difficult for a pharmaceutical company to influence the doctor's recommendation. Conversely, honest communication is unlikely when the decision maker does not have much at stake (so that the expected loss is small when $s = l$ and $d = 1$). In this case, a smaller payment is likely to change the advisor's report and therefore, in equilibrium, D would believe the advisor's report only when G is relatively low.

Note, also, that zero bribes are paid on the equilibrium path in this full communication equilibrium. However, the possibility of bribes/side payments reduces information transmission, in that it limits the parameter values for which informative communication is possible. It is intuitive that meaningful communication is impossible when w is large and when α_b is small. That is, when the third party has a strong interest in the decision and the bad advisor's altruistic incentives are weak.

[10] Note that this condition is independent of λ, the prior probability that the advisor is good. If T offers a payment that only a bad advisor would accept, it is only paid if the advisor is, indeed, bad. The fact that q_s is conditional on d means that T does not risk paying a bribe that does not accomplish its goal.

Communication is also impossible when π_l is close enough to zero, for example, if G or θ_l is relatively large. In such cases, the stakes are lower for the decision maker, so the altruistic concerns are weaker for the advisor.

Partially Informative Equilibrium.—With two types of advisors there is also a partially informative equilibrium in which one type of advisor is honest, but the other misreports the state. In such an equilibrium, the bad advisor always accepts the payment and reports the state as high, and the good advisor is honest.[11] Knowing this, the decision maker must consider the possibility of advisor dishonesty in interpreting the advisor's message. If a message of h is received, D knows that either $t = g$ and $s = h$, or $t = b$ and the message is independent of the state. Based on this, D can calculate

$$\Pr[s = h \,|\, m = h] = \frac{1/2}{(1/2\,\lambda) + (1-\lambda)} = \frac{1}{2-\lambda}.$$

The expected payoff to D of setting $d = 1$, given $m = h$, is $[1/(2-\lambda)]\,\pi_h + [(1-\lambda)/(2-\lambda)]\,\pi_l$, so D will act on a positive recommendation if this payoff is positive, or if

$$\pi_h + (1-\lambda)\pi_l > 0.$$

If a message of l was received then D knows that the advisor is honest, so the strategy after a negative report is still to set $d = 0$ given our assumption that $\pi_l < 0$. So the report is decisive for $-\pi_l \in (0, \pi_h/(1-\lambda))$. For a partially informative PBE, it must be that T is willing to offer a bribe large enough to convince a type b advisor to lie, but not large enough to convince a type g advisor to lie. The following proposition characterizes this equilibrium.

PROPOSITION 2: *There exists a PBE in which A's message is decisive, and only the good advisor is honest, if and only if* $0 < -\pi_l < \pi_h/(1-\lambda)$ *and,*

$$-\alpha_b\pi_l < w < \frac{-\pi_l[\alpha_g - (1-\lambda)\alpha_b]}{\lambda}. \tag{2}$$

In equilibrium, positive payments are made by T to the advisor.[12]

[11] If either type always reported l and the message was decisive, then that advisor could strictly improve his payoffs by switching to $m(h) = h$. If the bad advisor were honest and the good advisor always reported h, then we must have $\alpha_g\pi_l + q_l > 0$ and $\alpha_b\pi_l + q_l < 0$, a contradiction given $\alpha_g > \alpha_b$.

[12] As before, there are no equilibria in mixed strategies. If one type is mixing between $m_t(l) = l$ and $m_t(l) = h$, then we must have $\alpha_t\pi_l + q_l = 0$. T could increase q_l slightly and obtain a discrete increase in the probability of $d = 1$, improving his payoff except in the knife-edge case $-\alpha_t\pi_l = w$. However, as we shall see in the model with reputation effects presented in Section IV, mixed strategy equilibria are possible.

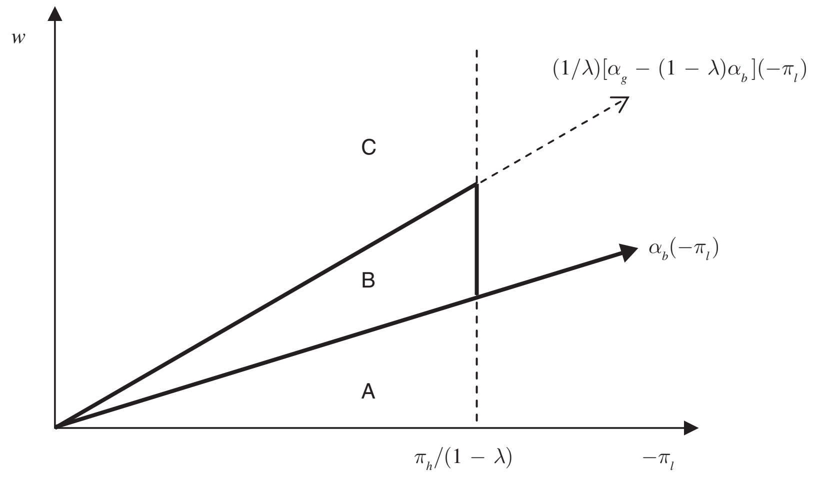

Figure 2. Equilibrium without Reputation Effects

Figure 2 illustrates this proposition. For $w < -\alpha_b \pi_l$ (region A), there is a fully informative PBE, similar to the case with one type of advisor. The threshold w is increasing in $-\pi_l$, reflecting the fact that as the wrong decision becomes more costly for D, it requires a higher bribe for T to overcome A's altruism. Region B represents the partially informative PBE. Again, the threshold w is increasing in $-\pi_l$, but for $-\pi_l$ large enough, a partially informative message is no longer decisive. In region C, no informative communication is possible.

The interesting point of the proposition is that in this partially informative equilibrium, positive bribes are paid by the third party in region B, where only good advisors are honest. If both types are honest (region A) or if no message is believed (region C), then there is no reason to pay bribes. Given that the advisor types are unobservable to T, the bribes, while collected by either type, are only successful in changing the reporting strategy of the bad advisor. The equilibrium size of the bribe will be $-\alpha_b \pi_l$, and it decreases if θ_l, the probability with which there is a success in the low state (for example, the quality offered by a low quality seller), increases. An increase in θ_l decreases the loss in utility that the advisor must overcome to convince the advisor to change his report, allowing the third party to reduce the bribe.

IV. Reputational Effects

We now introduce the consideration of reputational effects to derive the main results of the paper. If the decision maker and advisor interact repeatedly, the advisor must consider the reputational impact of current advice. The influence that the advisor's message will have on D's decision depends on the decision maker's beliefs about whether the advisor is good or bad, which depends, in turn, on past advice and outcomes. This means that the advisor's payoffs are affected by his reputation. A higher reputation gives him a greater chance to sway the decision maker, providing benefits both through A's altruistic motives and through a greater opportunity for bribes.

To capture reputation effects, consider an extension of the game in which the advisor and decision maker interact for two periods. The structure of the first period game is identical to the game in the previous section. The second period models the value of reputation for the advisor and represents the idea that the advisor may get payoffs from future interactions which are affected by the advisor's updated reputation at the end of the first period. Define the advisor's reputation, $\hat{\lambda}$ as the decision maker's belief that $t = g$ at the end of the first period, and let the advisor's second-period payoff be a function of this updated reputation. We denote by $V_t(\hat{\lambda})$, the second-period value of the updated reputation $\hat{\lambda}$ for the type t advisor. So the advisor's payoff, including reputation effects, is given by:

(3) $$W_t = \alpha_t U^D + q_s d + V_t(\hat{\lambda}).$$

For the analysis and results that follow, we simply require that $V_t(\hat{\lambda})$ is an increasing and continuous function of the updated reputation. Note that this assumption does not necessarily require a twice-repeated version of the game in the previous section. All that is necessary is that the advisor has future interactions (with the same or different decision makers) whose payoffs are increasing in his reputation. In the Appendix, we present a characterization of the second-period interaction between the advisor and the decision maker which implies that $V_t(\hat{\lambda})$ is an increasing function.

For the analysis with reputation effects to matter, it cannot be the case that there is a fully informative equilibrium in the first-period game. This is because if both types of advisors are honest in the first period, then the advisor reputation will not be affected, because both types behave identically. With identical advisor strategies, the decision maker can infer nothing about an advisor's type from his behavior or the realization of the state. Consequently, the second period will not affect play in the first period.[13]

A. *Reputational Updating*

Consider, therefore, the equilibrium in which only the good advisor is honest in the first period. In this case, the advisor's report does indeed affect his reputation, since a bad advisor sets $m = h$ more often than a good advisor. We start by describing how the advisor's reputation depends on his report in the first period and the outcome of D's decision. Define the potential outcomes by $x \in \{S, F, 0\}$, where S represents $d = 1$ followed by success, F represents $d = 1$ followed by failure, and 0 represents $d = 0$. Further define $\hat{\lambda}(m, x)$ as D's updated belief about the advisor's type, as a function of the advisor's report, and the actual outcome. In the following

[13] Even with reputation effects, there exists a fully informative equilibrium in which both types of A are honest, if and only if $w < -\alpha_b \pi_l$. Suppose that both types are honest in the first period, and that the threshold is higher than the one with no reputation effects. Given that the beliefs about A's type are the same in period 2 as in period 1, lying would have no reputational consequences. But, this means that a bad advisor would have accepted a bribe that would have worked in changing the report in the case without reputation effects. Thus, the threshold is the same as in the case without reputation effects.

proposition, we identify the condition under which the only possible informative equilibrium is one in which the good advisor is honest in the first period.[14]

PROPOSITION 3: *Define* $W(\hat{\lambda}) = V_g(\hat{\lambda}) - V_b(\hat{\lambda})$. *Then if* $W'(\hat{\lambda}) < min[-\pi_l(\alpha_g - \alpha_b); (\alpha_g - \alpha_b)\pi_h)$, *the only pure-strategy partially informative equilibrium that can exist is one in which the good advisor is honest and the bad advisor always reports* $m = h$.

Proposition 3 establishes a condition sufficient to rule out all partially informative equilibria in the first period, except the one in which the good advisor is honest while the bad advisor is not. Note that $W(\hat{\lambda})$ is the additional value that a good advisor places on reputation compared to a bad advisor. The left-hand side of the condition $W'(\hat{\lambda})$ is the change in this additional value with respect to reputation. The condition is that this change in reputational value is not too large. This ensures the expected outcome that the good advisor is more honest than the bad advisor in the first period.[15] The right-hand side of the inequality represents the maximum possible difference between the value of reputation to the good advisor and to the bad advisor. Thus, the idea in Proposition 3 is that reputational incentives would be not so much more important to the good advisor that the good advisor will mislead the decision maker in the first period (in order to improve his reputation), while the bad advisor tells the truth in the first period.

Consider then the case in which the good advisor is honest while the bad advisor always reports the state as high. If a bad advisor always chooses the message that the state is high, then a message that the state is low implies with certainty that the advisor is good. That is, $\hat{\lambda}(l, 0) = 1$. If the advisor recommends $m = h$, then his reputation will depend on the outcome x. If D experiences a success, then we have

$$\hat{\lambda}(h, S) = \frac{\lambda \frac{1}{2}\theta_h}{\lambda \frac{1}{2}\theta_h + (1 - \lambda)[\frac{1}{2}\theta_h + \frac{1}{2}\theta_l]} = \frac{\lambda\theta_h}{\theta_h + (1 - \lambda)\theta_l}$$

and if D experiences a failure,

$$\hat{\lambda}(h, F) = \frac{\lambda \frac{1}{2}(1 - \theta_h)}{\lambda \frac{1}{2}(1 - \theta_h) + (1 - \lambda)[\frac{1}{2}(1 - \theta_h) + \frac{1}{2}(1 - \theta_l)]}$$
$$= \frac{\lambda(1 - \theta_h)}{1 - \theta_h + (1 - \lambda)(1 - \theta_l)}.$$

[14] Reputation effects may also introduce the possibility of other types of partially informative communication. As we show in the Appendix, if second-period payoffs are large relative to first-period payoffs, then advisors might follow perverse strategies in the first period. For example, bad advisors might be honest while good advisors might always lie.

[15] For example, in the equilibrium, if a bad advisor were honest in the first period, but a good advisor always reported $m_g(l) = h$, then a message that the state is high would improve the advisor's reputation even when the message is deceptive.

The point to now note is that $\hat{\lambda}(h,F) < \hat{\lambda}(h,S) < \lambda$. Naturally, a recommendation of h that is followed by a failure is more harmful to the advisor's reputation than a recommendation followed by a success. But the more interesting point is that the advisor's reputation suffers whenever $m = h$, even if the result is a success. This feature of reputational updating can lead to the advisor choosing *to lie about the high state* in order to establish himself as a good type. Not only does honestly reporting $s = h$ harm the advisor's reputation relative to the prior, but in an informative equilibrium, lying about $s = h$ establishes the advisor as certainly good. This creates a reputational incentive to lie about state h on the part of both types of advisors. This feature of the reputational updating, which can cause the advisor to lie even when his incentives are aligned with those of the decision maker, is akin to the political correctness effect of Morris (2001). However, in this model, this effect arises because of endogenous incentives (i.e., the presence of the third party who can offer a bribe to maximize payoff). Consequently, the main point of our analysis is the characterization of the endogenous bribe offers that we proceed to do below.

In the partially informative equilibrium, each type of advisor must find the specified report optimal given the bribe offered, and T must not want to change the bribe. Define $\overline{V}_t(l)$ as the expected reputation value of the type t advisor resulting from $m = h$ when $s = l$. Thus, $\overline{V}_t(l) = \theta_l V_t(\hat{\lambda}(h,S)) + (1 - \theta_l)V_t(\hat{\lambda}(h,F))$, and likewise for $s = h$ is $\overline{V}_t(h) = \theta_h V_t(\hat{\lambda}(h,S)) + (1 - \theta_h)V_t(\hat{\lambda}(h,F))$. In equilibrium, T's choice of q_s must be consistent with the proposed advisor strategy $m_b(h) = m_b(l) = m_g(h) = h$; $m_h(l) = l$, which imply the following:[16]

$$m_b(l) = h \;\Rightarrow\; q_l \geq q_1 = V_b(1) - \overline{V}_b(l) - \alpha_b \pi_l \tag{4}$$

$$m_b(h) = h \;\Rightarrow\; q_h \geq q_2 = V_b(1) - \overline{V}_b(h) - \alpha_b \pi_h \tag{5}$$

$$m_g(l) = l \;\Rightarrow\; q_l \leq q_3 = V_g(1) - \overline{V}_g(l) - \alpha_g \pi_l \tag{6}$$

$$m_g(h) = h \;\Rightarrow\; q_h \geq q_4 = V_g(1) - \overline{V}_g(h) - \alpha_g \pi_h. \tag{7}$$

Let us define the variable $X_t(s) = V_t(1) - \overline{V}_t(s)$. Note that $V_t(1)$ is the value placed by the type t advisor for being perceived as the good type with certainty or, in other words, the maximum reputation value for the advisor of type t. Thus, $X_t(s)$ can be seen as the loss in reputation value (relative to the maximum value) for the advisor from sending the report of $m = h$ when the state is s. Consider, now, T's decision when $s = l$. As in the case with no reputation effects, T chooses from among three options: (i) no bribe, (ii) the minimum bribe necessary to obtain $m_b(l) = h$ (i.e., $q_1 = X_b(l) - \alpha_b \pi_l$), (iii) and the minimum bribe necessary to obtain $m_g(l) = h$ (i.e., $q_3 = X_g(l) - \alpha_g \pi_l$).

For (ii) to be preferred to (i), we must have $(1 - \lambda)(w - q_1) > 0$, or $w > -\alpha_b \pi_l + X_b(l)$ and for (ii) to be preferred to (iii), we must have $(1 - \lambda)(w - q_1) > (w - q_3)$,

[16] Note, as in the case without reputation effects, we need, when the advisor reports $m = h$, the decision maker will choose $d = 1$, which is the condition $0 < -\pi_l < \pi_h/(1 - \lambda)$.

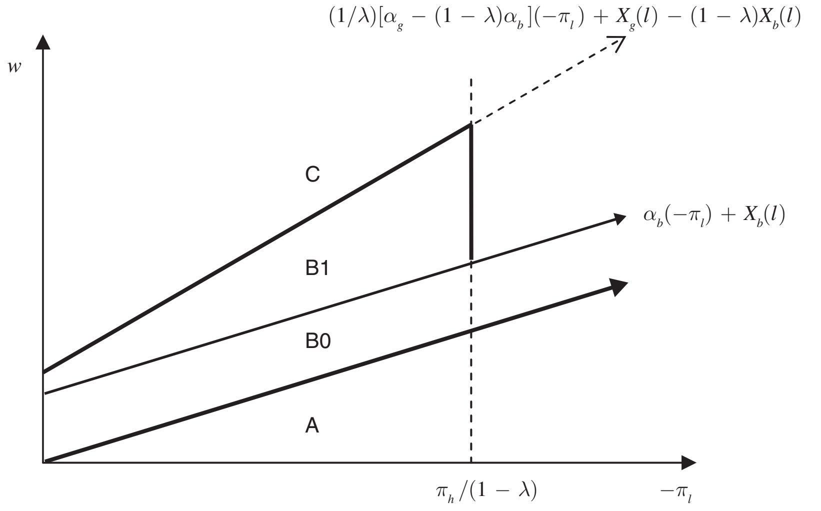

FIGURE 3. EQUILIBRIUM WITH REPUTATION EFFECTS

or $w < (1/\lambda)\{[(1-\lambda)\alpha_b - \alpha_g]\pi_l - (1-\lambda)X_b(l) + X_g(l)\}$. This leads to the condition stated in the following proposition:

PROPOSITION 4: *The partially informative equilibrium in which the good advisor is honest, while the bad advisor always reports $m = h$, exists when*

$$(8)\qquad -\alpha_b\pi_l + X_b(l) < w < \frac{1}{\lambda}\{[(1-\lambda)\alpha_b - \alpha_g]\pi_l - (1-\lambda)X_b(l) + X_g(l)\}.$$

In this equilibrium, T offers a payment $q_l^* \geq \max\{0, q_1\}$ when $s = l$.[17] *Further, T offers a payment $q_h^* \geq \max\{0, q_2, q_4\}$ to both types of advisors to induce a report of $m = h$ even when $s = h$.*

Figure 3 illustrates these conditions. First, note that region A is the same as with no reputation effects, because when both advisor types are honest, the advisor's reputation does not change in the second period. On the other hand, by comparing the width of the intervals in (8) and (2), we can show that the partially informative equilibrium in the case with reputational effects is feasible over a larger region relative to that with no reputation effects as long as $X_g(l) > X_b(l)$. In other words, a partially informative equilibrium becomes more likely if the loss in the reputation value for the good advisor, who lies about the bad state, is larger than the corresponding

[17] Note that we assume that $q_3 > q_1$, otherwise the bribe that induces $m_b(l) = h$ will also induce $m_g(l) = h$, which then implies that the partially informative equilibrium does not exist. The requirement $q_3 > q_1$ simply means that the bribe needed to get a good advisor to lie about l is larger than the corresponding bribe needed for the bad advisor.

loss for the bad advisor. If $X_g(l)$ is large enough, a larger bribe will be necessary to induce the good advisor to misreport a low state as high, while if $X_b(l)$ is not very large, then a small enough bribe will make the bad advisor lie. Under this condition, reputational concerns expand the parameter space over which the partially informative equilibrium exists, because inducing A to lie requires T to overcome not only A's altruistic motives, but also the harm to A's reputation. Conversely, the above also implies that reputational effects can reduce the parameter range for the partially informative equilibrium if the loss in the reputation value for the good advisor from sending a report of high is smaller than it is for the bad advisor. The region B1 represents the area in which a type b advisor always chooses $m(l) = h$. In region B0, a type b advisor will mix between $m(l) = h$ and $m(l) = l$ (see the Appendix for details).[18] In this region, the fact that even a type b advisor sometimes reports $m(l) = l$ lowers the reputational benefit of an honest report.

Turning to the equilibrium bribes, when $s = l$, then T will offer a bribe in order to induce the bad advisor to misreport the state exactly as in the case without reputation effects. But the situation is different when $s = h$. Recall that in the case without reputation, either type of advisor would truthfully report the high state. In contrast, in the presence of reputation effects, we must also consider the incentives of T when $s = h$. This is specified at the end of Proposition 4 which also highlights one of the main points of the paper. Any positive recommendation is costly to the advisor's reputation, and so we cannot assume that honest communication is optimal for A whenever $s = h$. To induce $m = h$ from both types of advisors, T must set q_h^* as derived in Proposition 4. If reputation effects are strong enough, it will be necessary for T to offer a positive bribe in order to induce honest communication about the good state.

An immediate implication is that it is possible in an informative equilibrium for T to offer a strictly positive bribe even if $s = h$ and $d = 1$ is optimal for the decision maker. It is intuitively straightforward that a positive monetary incentive (q_l^*) would be necessary to get the bad advisor to misreport the low state as high. But the more interesting implication is that a positive bribe may be necessary to get even the good advisor to truthfully report the high state. Indeed, it is possible that it is this concern that determines the equilibrium size of the payment that the third party offers. This would be the case when $q_4 > q_2$, or when $X_g(h) - X_b(h) > (\alpha_g - \alpha_b)\pi_h$. In other words, if the loss in reputation value from truly reporting the high state for the good advisor is sufficiently high compared to that for the bad advisor, then not only would T offer a bribe when the state was truly high, but also the magnitude of this bribe will be determined by the incentive to motivate the good rather than the bad advisor to correctly report the high state.

Figure 4 illustrates how this effect determines communication in equilibrium. In area D, $m(h) = h$ cannot be part of a partially informative equilibrium strategy with probability one because the third party is unwilling to pay a bribe large enough to overcome the lost reputation associated with such a recommendation. As in region B0, a mixed strategy on the part of the advisor can reduce the reputational incentives

[18] Thus, unlike in the case without reputational effects, an equilibrium in mixed strategies is possible.

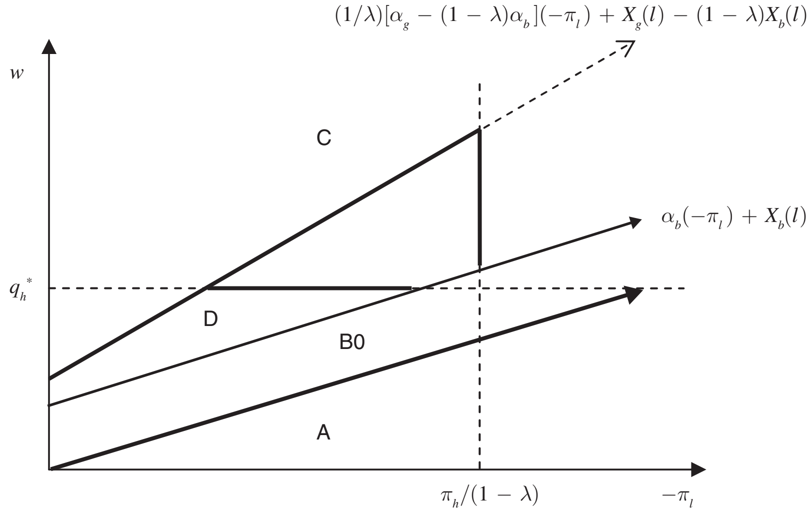

FIGURE 4. BRIBE TO INDUCE GOOD ADVISOR TO REPORT $m(h) = h$

to make communication possible. Here, a bad advisor will report $m(h) = l$ with positive probability. This lowers the reputational cost associated with $m = h$, so that T is able to induce $m = h$ by choosing a payment below q_h^*.

In summary, the possibility of a bribe reduces the scope for informative communication. But communication is not eliminated. While the potential for a bribe can have the negative effect of interfering with communication, in the presence of reputational concerns, the actual bribe can also have the positive effect of restoring communication that would not have otherwise taken place.

B. *Extensions*

***T* knows the Advisor Type:** We have assumed that the decision maker and the third party are symmetrically uninformed about the advisor's type. Here, we consider the case where the third party knows the advisor type while the decision maker does not. This may represent the case of movie critics or stock analysts, where the third party has a greater stake and ability in knowing about the advisor than any individual decision maker. If the third party knows the advisor's type, he can offer a more efficient bribe, tailoring the size of the payment to the type of the advisor. One would expect this to make information transmission more difficult. T is more willing to offer a bribe that changes a type g advisor's message, since he knows he will not be overpaying for a type b advisor's message.[19]

For example, in an equilibrium in which only the good advisor is honest, the bribe necessary to get the advisor to change his report is the same as before, and the necessary conditions on q (q_1 to q_4) are the same as before as in (4)–(7). But

[19] Bribes are also more efficient in the event of $s = h$. The third party does not waste bribes on someone who would have made a positive recommendation anyway.

the difference will be in T's payoffs, which now includes four possible bribes. The necessary conditions for T's strategy are $w \geq \max(q_1, q_2, q_4)$ and $w \leq q_3$. The first condition is also necessary in the base model, where the third party also knows the advisor type. The second condition is weaker. When T does not know t, the condition is $(1 - \lambda)(w - q_1) \geq w - q_3$, or $w \leq (q_3 - (1 - \lambda)q_1)/\lambda$. So, as long as $q_3 > q_1$ (that is, as long as the bribe needed to get a good advisor to lie is larger than the bribe needed to get a bad advisor to lie), the conditions necessary for communication are stricter when the third party knows t. Or, in other words, no communication in the base model implies no communication when T knows t.

***T* is Uninformed of *s*:** What would happen if the third party knows less about the state than the advisor? Often the advisor may know more than the third party about the optimal decision. An example is when there is uncertainty about the quality of the match between a specific patient and a new drug. In such a case, the accuracy of the prescription may depend not only upon the characteristics of the drug, but also upon the specific characteristics of the individual patient. Thus, an advisor (doctor) who knows about the characteristics of a drug as well as about the patient history may be in a better position to describe how well it suits an individual patient's needs than the pharmaceutical firm. Accordingly, assume that the third party does not know the state of the world. However, it still has an incentive to influence the advisor's message, as long as the advisor's message is decisive.

The effect of the third party lack of information about s is easily seen in the full communication equilibrium. When T does not know the state, honest communication will lead to a "yes" decision by the decision maker with a probability $1/2$, implying an expected payoff of $w/2$. From this we can show that the necessary condition for an equilibrium (i.e., T does not want to offer a payment that A would accept) will be $w \leq -2\alpha_b \pi_l$, which is weaker than the necessary condition for the case where T is informed about the state ($w \leq -\alpha_b \pi_l$). While an uninformed T also has an incentive to offer payments to influence the report of the advisor, this incentive is weaker than in the case where T knows s because the third party does not know whether a bribe is necessary or not. If the state is in fact h, T may end up paying the advisor for the message that would have been sent even without the payment. This means that an informative equilibrium is more likely when the third party does not know the state. Lack of information for the third party facilitates honest communication. An analysis which parallels that of the previous section can show that this result and intuition is equally true for the partially informative equilibrium with reputation effects presented in Section IV.

V. Conclusion

A decision maker seeking advice must always be concerned with the independence of the advisor he consults. In many circumstances, there is the opportunity for those with a stake in the decision maker's action to try to influence the advisor through direct economic incentives that are unobservable to the decision maker. We generally think of reputation as reinforcing the independence of the advisor, since the appearance of bias will undermine the advisor's influence in the future.

This desire to appear independent can cut both ways. It can prevent an advisor from acceding to the influence of interested third parties, but it can also prevent the advisor from recommending an action that he knows would be beneficial for the decision maker, out of a desire to appear incorruptible.

It is common to ascribe negative effects to bribes or side payments made by third parties because they influence the advisor and prevent the transmission of truthful information. This well-known information corrupting effect of third-party side payments has meant that third-party side payments have often been subjected to criticism in the context that motivates this paper. Contrary to this criticism, our analysis shows that when advisors have reputational concerns, bribes offered by third parties may indeed play a positive role of restoring truthful communication that would otherwise have not taken place. Thus, bribes from interested third parties may have two distinct roles. The obvious role is to influence the bad advisor to lie about the bad state (example, a poor quality product). But bribes may have a second more interesting role. They may be offered to compensate even a good advisor to truthfully report a good state. In this case, these bribes are a compensation for the loss of reputational capital of the good advisor. An implication of our results is that, when we observe a seller attempting to influence an advisor, we should not necessarily infer that the seller is trying to encourage bad advice. Even when he is offering the buyer good advice, the advisor may require a payment to compensate for his loss of reputational capital.

The possibility of bribes interferes with communication. But, the presence of bribes does not necessarily imply that the advisor's message is counter to the interests of the decision maker. In the presence of advisor reputational concerns, bribes may actually serve to restore communication which might otherwise have failed. We also show that information transmission from the advisor is decreased with better third-party knowledge about the state. Similarly, if the third party knows the advisor's type better than the decision maker, then it becomes more difficult for the advisor to credibly convey useful information to the decision maker.

Appendix

PROOF OF LEMMA 1:

Suppose, for the type t advisor, $\Pr(m_t(h) = h) = y \in (0,1)$, and that $m_t(s)$ is decisive. Then A must be indifferent between $d = 0$ and $d = 1$ when $s = h$. Thus, $\alpha_t \pi_h + q_h = 0$. Since $q_h \geq 0$, this is impossible. Suppose $\Pr(m_t(l) = l) = y \in (0,1)$, and that $m_t(s)$ is decisive. Then A is indifferent between $d = 0$ and $d = 1$ when $s = l$. Thus, $\alpha_t \pi_l + q_l = 0$. T's payoff is given by $(1 - y)(w - q_l)$. If $w > q_l$, then T can do strictly better by choosing $q_l + \varepsilon$, and obtaining payoff $w - q_l - \varepsilon$. If $w < q_l$, then T can do strictly better by choosing $q_l = 0$. So there cannot exist a mixed strategy equilibrium except possibly for the knife-edge case $w = -\alpha_t \pi_l$. In the knife-edge case, T can offer a maximum bribe of $-\alpha_t \pi_l$ which can make A indifferent between $d = 0$ and $d = 1$ when $s = l$.[20]

[20] Even in this knife-edge case of $w = -\alpha_t \pi_{l,}$ there cannot be a mixed strategy equilibrium if we assume the tie-breaking rule that T will not choose to offer a bribe if its payoffs from bribing and not bribing are identical.

PROOF OF PROPOSITION 1:

(Necessity) In an informative equilibrium, D believes that A is honest. Given these beliefs, the optimal strategy for the advisor $m_t(s, q_s)$ is given by

$$m_t(h, q_h) = \begin{cases} h \text{ if } \alpha_t \pi_h + q_h \geq 0 \\ l \text{ if } \alpha_t \pi_h + q_h < 0 \end{cases}$$

$$m_t(l, q_l) = \begin{cases} h \text{ if } \alpha_t \pi_l + q_l \geq 0 \\ l \text{ if } \alpha_t \pi_l + q_l < 0 \end{cases}.$$

Since $q_h \geq 0$, $m_t(h, q_h) = h$ always. Also, $m_t(l, q_l) = l$ if and only if $q_l < -\alpha_t \pi_l$. Further, because $\alpha_b < \alpha_g$, the advisor will report a message of l when $s = l$ irrespective of his type when $q_l < -\alpha_b \pi_l$. If $m_t(l) = l$, irrespective of the type of the advisor, then T's payoff is 0, while if T is able to afford a bribe q_l to at least induce the bad advisor to misreport $m_b(l) = h$, T's payoff is $(1 - \lambda)(w - q_l)$. Therefore, T will choose $q_l > -\alpha_b \pi_l$ whenever $w > -\alpha_b \pi_l$. Therefore, $w \leq -\alpha_b \pi_l$ is a necessary condition.

(Sufficiency) If $w \leq -\alpha_b \pi_l$, then $q_h = q_l = 0$ is an equilibrium strategy for T and is consistent with an honest, decisive message.

PROOF OF PROPOSITION 2:

As discussed in the text, $\pi_h + (1 - \lambda)\pi_l > 0$ implies that a partially informative message from the advisor will be decisive. Given a decisive message, when $s = l$, T's choice of q_l will fall in one of three ranges: (i) it will be too small to affect the message of either type of advisor, (ii) it will be large enough to cause a bad advisor to choose $m_b(l) = h$, but not large enough to cause a good advisor to do so, or (iii) it will be large enough to cause either type of advisor to choose $m(l) = h$. Clearly, T will prefer the minimum q_l within each range, meaning T chooses from (i) $q_l = 0$, in which case $U^T = 0$, (ii) $q_l = -\alpha_b \pi_l$, in which case $U^T = (1 - \lambda)(w + \alpha_b \pi_l)$, and (iii) $q_l = -\alpha_g \pi_l$, in which case $U^T = w + \alpha_g \pi_l$. Therefore, for a partially informative equilibrium, T must prefer (ii) to (i), or $w > -\alpha_b \pi_l$, and must prefer (ii) to (iii), or $w < -(1/\lambda)[\alpha_g - (1 - \lambda)\alpha_b]\pi_l$.

A. *A Characterization of the Second-Period Game that Generates $V_t(\hat{\lambda})$*

Consider a second-period continuation game denoted by the subscript 2 which is as follows: Assume that in the second period there is no third-party action, but that the advisor receives a known benefit $r > 0$ if the second-period decision $d_2 = 1$. The state in the second period, $s_2 \in \{l, h\}$, is independent of the state in the first period, and each state occurs with probability 1/2. After the state is observed, the advisor chooses m_2, and then the decision maker chooses d_2. Whether the advisor will enjoy the benefit r in the second period will depend upon whether the decision maker's action is influenced by the advisor's message. This, in turn, is affected by the decision maker's updated belief of the advisor's type at the end of period 2.

Assume that if $d_2 = 0$, then $U_2^D = 0$. If $d_2 = 1$, then $U_2^D = \pi_{2s}$, with $\pi_{2l} < 0 < \pi_{2h}$. Advisor second-period payoffs, as before, are $U_2^A = \alpha_t U_2^D + d_2 r$. Next, assume that π_{2l} is a known parameter, but that π_{2h} is drawn from a distribution $F(\pi_{2h})$, which is everywhere continuous with support on $(0, \infty)$, at the beginning of the second period, at which point it is observed by both parties. As with s_1, the second-period state s_2 is also observed by the advisor and not by the decision maker.

In order for reputation effects to matter in the second period, we assume that $-\alpha_g \pi_{2l} > r > -\alpha_b \pi_{2l}$, which implies that in the second period a good advisor is always honest and a bad advisor is never honest.[21]

Define $V_t(\hat{\lambda})$ as the second-period payoff to an advisor of type t with an updated reputation $\hat{\lambda}$ at the end of the first period. Given $-\alpha_g \pi_{2l} > r > -\alpha_b \pi_{2l}$, in the second period a good advisor is always honest and a bad advisor is never honest. When $m = l$, D will know that $t = g$ and that $s = l$. When $m = h$, the expected utility from $d_2 = 1$ is $\hat{\lambda}\pi_{2h} + (1 - \hat{\lambda})[1/2\,\pi_{2h} + 1/2\,\pi_{2l}]$, so the decision maker will choose $d_2 = 1$ if

$$\pi_{2h} > -\frac{(1 - \hat{\lambda})}{(1 + \hat{\lambda})}\pi_{2l}.$$

Define $a(\hat{\lambda}) = -((1 - \hat{\lambda})/(1 + \hat{\lambda}))\pi_{2l}$. Then the payoffs to a type g advisor, as a function of π_{2h}, are

$$\begin{array}{ll} 0 & \text{if} \quad \pi_{2h} < a(\hat{\lambda}) \\ \frac{1}{2}[\alpha_g \pi_{2h} + r] & \text{if} \quad \pi_{2h} > a(\hat{\lambda}). \end{array}$$

And, so, the expected payoffs at the start of the second period are

$$V_g(\hat{\lambda}) = \frac{1}{2}\int_{a(\hat{\lambda})}^{\infty} [\alpha_g x + r]\, dF(x). \tag{9}$$

Payoffs to a type b advisor are

$$\begin{array}{ll} 0 & \text{if} \quad \pi_{2h} < a(\hat{\lambda}) \\ r + \dfrac{\alpha_b(\pi_{2h} + \pi_{2l})}{2} & \text{if} \quad \pi_{2h} > a(\hat{\lambda}), \end{array}$$

and payoffs at the end of the first period are

$$V_b(\hat{\lambda}) = \int_{a(\hat{\lambda})}^{\infty} [\frac{\alpha_b(x + \pi_{2l})}{2} + r]\, dF(x). \tag{10}$$

[21] If $-\alpha_g \pi_{2l} > -\alpha_b \pi_{2l} > r$, then both types will be honest in the second period, and if $r > -\alpha_g \pi_{2l} > -\alpha_b \pi_{2l}$, then neither type will be honest in the second period. Thus, our assumption is necessary for reputation effects to be relevant in the model.

From the above expressions for $V_t(\hat{\lambda})$, and given that $a(\hat{\lambda})$ is continuous, we can show that $V_t(\hat{\lambda})$ are increasing, continuous functions of $\hat{\lambda}$. For both types, an increase in $\hat{\lambda}$ increases payoffs at the end of the first period, by increasing the range of π_{2h} over which the message is decisive. A better reputation is valuable to the advisor because it makes his advice more influential; it increases the range of parameters over which the decision maker's choice is determined by the advisor's report. This has value both for a good and a bad advisor. For a bad advisor, the benefit is an increased opportunity to obtain the full private benefit r when $d_2 = 1$ (while getting only the mean altruism value $\alpha_b(\pi_{2h} + \pi_{2l})/2$). A good advisor who is honest receives the benefit r only half the time when the state is h, but also gets the full payoff from his altruistic motives, since greater influence will imply that the decision maker makes better choices.

PROOF OF PROPOSITION 3:

There are three alternative classes of informative equilibria that need to be ruled out. The procedure for the proof will be as follows. First, we derive the necessary conditions for equilibria in each of these classes. Next, we show, under the assumption about $W'(\hat{\lambda})$ shown in Proposition 3, these necessary conditions cannot be met.

(i) The good advisor always reports l and only the bad advisor is honest, i.e., $m_g(h) = m_g(l) = l$, $m_b(h) = h$, $m_b(l) = l$. Note that in such an equilibrium a report of $m = h$ will establish the advisor as surely bad. However, for $m = l$, D will have to update his belief about λ based on the outcome. Let us define $\dot{V}_t(s)$ as the expected reputation value for the type t advisor resulting from a report of $m = l$ when the state was s. This can be defined as $\dot{V}_t(s) = \theta_s V_t(\hat{\lambda}(l, S)) + (1 - \theta_s)V_t(\hat{\lambda}(l, F))$, where S and F, respectively, denote success or failure in outcomes for D. Consider, now, the reporting strategies of both types of advisors when h is realized. $m_b(h) = h$ implies $q_h + \alpha_b\pi_h + V_b(0) \geq \dot{V}_b(h)$, and $m_g(h) = l$ implies $\dot{V}_g(h) \geq V_g(0) + \alpha_g\pi_h + q_h$, so these strategies imply that a necessary condition for this type of partially informative equilibrium is

$$[\dot{V}_g(h) - V_g(0)] - [\dot{V}_b(h) - V_b(0)] > (\alpha_g - \alpha_b)\pi_h. \tag{11}$$

(ii) The good advisor always reports h and only the bad advisor is honest: $m_g(h) = m_g(l) = h$, $m_b(h) = h$, $m_b(l) = l$. Now, let $\ddot{V}_t(s)$ represent the expected reputation value from a report of $m = h$ when the state is s. Considering the reporting strategies when $s = l$, we have that $m_b(l) = l$ implies $q_l + \ddot{V}_b(l) + \alpha_b\pi_l < V_b(0)$, and $m_g(l) = h$ implies $q_l + \ddot{V}_g(l) + \alpha_g\pi_l > V_g(0)$. So, these strategies imply that a necessary condition for this type of partially informative equilibrium is

$$[\ddot{V}_g(l) - V_g(0)] - [\ddot{V}_b(l) - V_b(0)] > -(\alpha_g - \alpha_b)\pi_l. \tag{12}$$

(iii) The good advisor is always honest, and the bad advisor always reports l: $m_g(h) = h$, $m_g(l) = l$, $m_b(h) = m_b(l) = l$. Letting the expected reputation value of reporting l to be $\tilde{V}_t(s)$, we have: $m_b(l) = l$ implies $q_h + V_b(1) < \tilde{V}_b(l)$, a contradiction because $V_b(1) > \tilde{V}_b(l)$. So, this type of partially informative equilibrium will never exist.

Note that $W(\hat{\lambda}) = V_g(\hat{\lambda}) - V_b(\hat{\lambda})$ is continuous and assume that it is also differentiable in $\hat{\lambda}$.[22] Now, suppose that $W'(\hat{\lambda}) < (\alpha_g - \alpha_b)\pi_h \;\forall\; \hat{\lambda}$. Because $W(\hat{\lambda})$ is a continuous function, we have, from the intermediate value theorem that there exist $\hat{\lambda}_1 > \hat{\lambda} > \hat{\lambda}_0$ such that $(W(\hat{\lambda}_1) - W(\hat{\lambda}_0))/(\hat{\lambda}_1 - \hat{\lambda}_0) < (\alpha_g - \alpha_b)\pi_h \;\forall\; \hat{\lambda}_1 > \hat{\lambda}_0$. Because $\hat{\lambda}$ is a probability measure, it follows that $W(\hat{\lambda}_1) - W(\hat{\lambda}_0) < (\alpha_g - \alpha_b)\pi_h$ $\forall\; \hat{\lambda}_1 > \hat{\lambda}_0$. Now, note that $[\dot{V}_g(l) - V_g(0)] - [\dot{V}_b(l) - V_b(0)] = \theta_l W(\hat{\lambda}(l,S)) + (1-\theta_l)W(\hat{\lambda}(l,F)) - W(0) < W(\hat{\lambda}(l,S)) - W(0)$. So, $\dot{V}_g(l) - V_g(0) - [\dot{V}_b(l) - V_b(0)] < (\alpha_g - \alpha_b)\pi_h$, and this contradicts the necessary condition in (11).

In a similar manner, we can show that if $W'(\hat{\lambda}) < -\pi_l(\alpha_g - \alpha_b) \;\forall\; \hat{\lambda}$, then this leads to a contradiction of the necessary condition in (12). Thus, if $W'(\hat{\lambda}) < \min[-\pi_l(\alpha_g - \alpha_b); (\alpha_g - \alpha_b)\pi_h]$, we will have that the only partially informative equilibrium that may exist is one in which the good advisor is honest while the bad advisor always reports $m = h$.

PROOF OF PROPOSITION 4:

From the discussion in the text preceding the proposition, we can note that when the condition in (8) is satisfied, T is willing to offer a bribe of at least q_1, which is high enough to induce $m_b(l) = h$, but not a bribe q_3, which would be high enough to induce $m_g(l) = h$.[23] So, the equilibrium conditions (4) and (6) will be satisfied and equilibrium payment made by T when $s = l$ is q_l^* as shown in Proposition 4. Next, when $s = h$, if T offers $q_h^* \geq \max\{0, q_2, q_4\}$, then the conditions (5) and (7) are also satisfied. Finally, if the advisor's reporting strategies are as given in (4)–(7), this is consistent with D's choice of $d = 0$ when the message is l and $d = 1$ when the message is h (given $0 < -\pi_l < \pi_h/(1-\lambda)$, which ensures that when $m = h$, D will choose $d = 1$), which, in turn, is consistent with T's payment strategy.

Mixed strategies with reputation effects: Suppose that a type b advisor chooses $m(l) = l$ with probability z. We will show that for any $w \in [-\alpha_b\pi_l, -\alpha_b\pi_l + V_b(1) - \overline{V}_b(l)]$ (region B1 in Figure 3), there exists a z such that the maximum bribe T is willing to pay leaves a type b advisor indifferent between $m(l) = l$ and $m(l) = h$. Given z, updated reputations are given by

$$\hat{\lambda}(l,0,z) = \frac{\lambda}{\lambda + (1-\lambda)z}$$

$$\hat{\lambda}(h,S,z) = \frac{\lambda\theta_h}{\theta_h + (1-\lambda)(1-z)\theta_l}$$

[22] For example, for the second-period continuation game described in Section A of the Appendix, $W(\hat{\lambda})$ can be shown to be a continuous and differentiable function. Subtracting (10) from (9) and simplifying, we have that $W(\hat{\lambda}) = ((\alpha_g - \alpha_b)/2)[\overline{\pi}_{2h} - F(a(\hat{\lambda}))E(\pi_{2h} \mid \pi_{2h} < a(\hat{\lambda}))] - [1 - F(a(\hat{\lambda}))][(\alpha_b\pi_{2l} + r)/2]$, where $\overline{\pi}_{2h}$ is the mean of the random variable π_{2h}, and $E(\cdot)$ is the conditional expectation operator. Given that $F(\pi_{2h})$ is continuous and that $a(\hat{\lambda})$ is continuous and differentiable in $\hat{\lambda}$, it follows that $W(\hat{\lambda})$ is continuous and differentiable in $\hat{\lambda}$.

[23] Assume that $q_3 > q_1$, otherwise the bribe that induces $m_b(l) = h$ will also induce $m_g(l) = h$, which then implies that the partially informative equilibrium does not exist.

$$\hat{\lambda}(h, F, z) = \frac{\lambda(1 - \theta_h)}{1 - \theta_h + (1 - \lambda)(1 - z)(1 - \theta_l)}.$$

Define $\hat{V}_b(m, z)$ by

$$\hat{V}_b(m, z) = \begin{cases} V_b(\hat{\lambda}(l, 0, z)) & \text{if} \quad m = l \\ \theta_l V_b(\hat{\lambda}(h, S, z)) + (1 - \theta_l) V_b(\hat{\lambda}(h, F, z)) & \text{if} \quad m = h. \end{cases}$$

A type b advisor will be indifferent between $m(l) = l$ and $m(l) = h$ if $q + \alpha_b \pi_l + \hat{V}_b(h, z) = \hat{V}_b(l, z)$, or

$$q = -\alpha_b \pi_l + \hat{V}_b(l, z) - \hat{V}_b(h, z).$$

When $z = 1$, $\hat{\lambda} = \lambda$ regardless of m, so $\hat{V}_b(l, z) - \hat{V}_b(h, z) = 0$. When $z = 0$, $\hat{V}_b(l, z) - \hat{V}_b(h, z) = V_b(1) - \overline{V}_b(l)$. Since $\hat{\lambda}$ is continuous on $z \in [0, 1]$, and V_b is assumed to be continuous in $\hat{\lambda}$, for any $w \in [-\alpha_b \pi l, -\alpha_b \pi_l + V_b(1) - \overline{V}_b(l)]$, there exists a z such that $w = -\alpha_b \pi_l + \hat{V}_b(l, z) - \hat{V}_b(h, z)$.

REFERENCES

Benabou, Roland, and Guy Laroque. 1992. "Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility." *Quarterly Journal of Economics*, 107(3): 921–58.

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

Ely, Jeffrey C., and Juuso Valimaki. 2003. "Bad Reputation." *Quarterly Journal of Economics*, 118(3): 785–814.

Ely, Jeffrey, Drew Fudenberg, and David K. Levine. 2008. "When Is Reputation Bad?" *Games and Economic Behavior*, 63(2): 498–526.

Farrell, Joseph, and Robert Gibbons. 1989. "Cheap Talk with Two Audiences." *American Economic Review*, 79(5): 1214–23.

Krishna, Vijay, and John Morgan. 2001. "A Model of Expertise." *Quarterly Journal of Economics*, 116(2): 747–75.

Morgan, John, and Phillip C. Stocken. 2003. "An Analysis of Stock Recommendations." *RAND Journal of Economics*, 34(1): 183–203.

Morris, Stephen. 2001. "Political Correctness." *Journal of Political Economy*, 109(2): 231–65.

Ottaviani, Marco, and Peter Norman Sorensen. 2006a. "Professional Advice." *Journal of Economic Theory*, 126(1): 120–42.

Ottaviani, Marco, and Peter Norman Sorensen. 2006b. "Reputational Cheap Talk." *RAND Journal of Economics*, 37(1): 155–75.

Scharfstein, David S., and Jeremy C. Stein. 1990. "Herd Behavior and Investment." *American Economic Review*, 80(3): 465–79.

Sobel, Joel. 1985. "A Theory of Credibility." *Review of Economic Studies*, 52(4): 557–73.