An offer you *can* refuse: the effect of transparency with endogenous conflict of interest*

Melis Kartal[†]
Vienna University of Economics and Business & VCEE

James Tremewan[‡]
University of Vienna & VCEE

September 26, 2016

Abstract

We study the effects of transparency on information transmission and decision making theoretically and experimentally. We develop a model in which a decision maker seeks advice from a better-informed adviser whose advice might be swayed by a side payment from an "interested" third party. Transparency enables the decision maker to learn the decision of the adviser regarding the side payment. Prior theoretical and experimental research has found that transparency is ineffective or harmful to decision makers. Our model predicts that transparency is not harmful and, depending on equilibrium selection, may strictly improve the accuracy of decision makers. Our experiment shows that transparency does indeed improve accuracy.

Keywords: Experimental game theory, strategic information transmission, transparency, conflict of interest.

*We would like to thank Thomas de Haan, Tim Johnson and Dimitri Landa, as well as seminar audiences at GATE-University of Lyon, Heidelberg University, University of Vienna, WZB Berlin Social Science Center, NYU CESS Annual Experimental Political Science Conference, and the 2015 ESA European Meeting.

[†]Department of Economics, University of Vienna, Email: *melis.kartal@univie.ac.at*.

[‡]Department of Economics, University of Vienna, Email: *james.tremewan@univie.ac.at*.

1 Introduction

In 2013, the Occupational Safety and Health Administration (OSHA) in the U.S. began a public consultation on setting new limits for working with silica dust, which is a major health hazard for construction workers that causes serious lung disease. The OSHA created substantial controversy in the Senate when it requested for the first time that those submitting scientific evidence should disclose their funding sources. A number of senators protested against the request arguing that revealing this type of information would bias the judgement of the agency. In turn, the head of the OSHA defended the request vigorously, claiming that transparency is indispensable so that the information that the agency bases its decision upon meets the highest standard of integrity.[1] How transparency affects advice and whether it improves the accuracy of decision making in settings such as this, where the expert might be influenced by a third party, is the topic of this paper. Transparency enables the decision maker to learn whether or not the expert accepted a side payment from the third party—without transparency, the decision maker learns only the advice.

Advice is prevalent in a variety of settings, ranging from regulatory agencies, legislatures and judiciaries to medical services and financial markets. In such settings, decision makers often face complex decisions with uncertain outcomes, and therefore seek the advice of an expert in order to increase the likelihood of a successful decision. However, information transmission from the expert to the decision maker may be impaired: Even if the expert and the decision maker do not have a conflict of interest *in the absence of a third party*, a third party (e.g. a special interest group or an industry) may sway the expert's advice in its favor by offering him a financial reward. Such concern regarding the impartiality of experts funded by third parties was raised in a recent study by the New York Times which identified dozens of examples of think tank researchers who helped shape the U.S. government policies while being paid by corporations who had stakes in those policies.[2]

On the one hand, transparency is assumed to remedy this sort of situation: It protects the decision maker by revealing whether the expert has a conflict of interest that might lead him to give biased advice. On the other hand, one counter-argument against transparency is that disclosing this type of information results in a bias itself: Even if the expert's advice is truthful, the decision maker may dismiss the advice if the expert has accepted funding from an industry or a special interest group.[3] According to the proponents of this idea, the

[1] Published in Nature on March 4, 2014. In the same issue, an editorial piece argued that regulatory agencies must demand conflict of interest statements for the research that they use.

[2] http://www.nytimes.com/2016/08/09/us/politics/think-tank-scholars-corporate-consultants.html

[3] See, for example, Stossel (2005), Weber (2009) and Stossel and Stell (2011).

bias against experts funded by third parties is harmful, so "the conflict of interest mania" must be cured.[4] An intimately related debate is whether transparency should be *voluntary* or *mandatory*. For example, organizations such as Transparency International advocate mandatory registering of lobbying activity. Most politicians and lobbyists agree that this should be the case; however, in countries where registers exist they often remain voluntary.[5,6]

Although it is a widely-held belief that transparency should ameliorate concerns related to conflicts of interest, as suggested above it is not immediately clear how it should do so. Moreover, prior research has produced bleak results regarding the effects of transparency on decision making. We provide the first formal model which illustrates a precise mechanism through which transparency can lead to better decision making. However, transparency does not guarantee better outcomes due to the existence of multiple equilibria. Therefore, we run an experiment to establish whether or not transparency improves information transmission in practice.

In the model, there are two states of the world (labeled L and R) and two possible policies (labeled l and r). The adviser is an expert who is perfectly informed about the state of the world whereas the decision maker knows only the prior probability of each state. The adviser recommends a policy to the decision maker, after which the decision maker makes a policy choice. The payoff of the decision maker is maximized if the chosen policy matches the state of the world. All else being equal, the adviser and the decision maker have no conflict of interest; i.e., their payoffs are aligned. However, prior to the policy recommendation stage—but after the adviser learns the true state of the world—a third party offers a side payment to the adviser. The third party strictly prefers policy r regardless of the state of the world. If the adviser accepts the offer, then he must recommend policy r.

We consider the following scenarios. In the *transparency* condition, the decision maker is informed whether or not the adviser accepted the third-party payment. In the *non-transparency* condition, the decision of the adviser regarding the payment is not disclosed. We also study a *voluntary-transparency* condition in which transparency is not enforced and the adviser chooses whether or not to disclose his decision regarding the third-party payment.

To highlight the basic mechanism through which transparency may improve decision making, we focus on a simple model with stark assumptions, but we show that this mechanism is robust to rich extensions in which these assumptions are relaxed. In particular, it is robust to assuming that the adviser is (ex-ante) imperfectly informed and obtains higher quality information if he accepts the third-party payment, and that the adviser is free to

[4] Rago, Joseph. "A Cure for 'Conflict of Interest' Mania." Wall Street Journal, 26 June 2015.
[5] http://www.transparencyinternational.eu/wp-content/uploads/2015/04/Lobbying_web.pdf
[6] http://www.oecd.org/gov/ethics/Lobbying-Brochure.pdf

choose his recommendation (with some probability) even if he accepts the payment.[7]

Previous experimental research has shown that transparency is either ineffective or has *adverse* effects on decision makers (Cain *et al.*, 2005, 2011; Koch and Schmidt, 2010; Rode, 2010; Loewenstein *et al.*, 2011; Loewenstein *et al.*, 2012; and Ismayilov and Potters, 2014). Our model differs from the previous studies in that the conflict of interest between the adviser and the decision maker is endogenous. On the one hand, the adviser and the decision maker have no conflict of interest if the adviser rejects the third-party payment. Hence, the transparency condition allows the adviser to send a costly signal to the decision maker that his advice is honest by rejecting the payment. Note, however, that accepting the third-party payment per se is not inherently dishonest. Whether the adviser does accept the third-party payment and how the decision maker interprets the advice of an adviser who accepted the third-party payment are determined in equilibrium.

Our predictions regarding the effect of transparency are as follows.[8] In the non-transparency condition, the adviser always accepts the payment from the third party and recommends r. As a result, the adviser's recommendation is uninformative. We denote this equilibrium the "corrupt equilibrium." In the transparency condition, there are two equilibria of interest. The first one is the corrupt equilibrium, in which behavior is the same as in the equilibrium of the non-transparency condition. The second equilibrium involves honest adviser behavior: The adviser always rejects the third-party payment and makes a truthful recommendation, which the decision maker follows. We denote this the "honest equilibrium." Note that sustaining the honest equilibrium requires the type of bias which opponents of transparency argue will be the result of disclosure: If the adviser accepts the third-party payment and recommends policy r, this prompts the decision maker to believe that the adviser is dishonest and choose policy l.[9] Both honest and corrupt equilibria also exist with voluntary transparency.

Transparency may strictly improve information transmission and decision making in our model, but our theoretical predictions are not sharp due to equilibrium multiplicity. In order to evaluate whether the honest equilibrium can arise in practice and gain further insights regarding the effect of transparency on decision making, we designed and ran an

[7] We also note that mounting evidence suggests that third parties who fund scientific and academic research do expect the conclusions to be consistent with their interests. See Footnote 11 and Section 4.4 for a discussion of highly publicized, large-impact examples.

[8] While we provide a theoretical analysis with all possible parameter values, here we focus on the parameter values such that equilibrium outcomes under transparency and nontransparency conditions are not identical.

[9] This decision maker bias is necessary to improve decision making with transparency because there is a significant agency problem. However, the bias emerges only off-the-equilibrium path as the adviser rejects the payment in the honest equilibrium.

experiment on the basis of our model, implementing each of the three conditions discussed above. Overall, we find that transparency clearly improves decision making relative to non-transparency condition. However, the evidence regarding the effect of voluntary transparency is much weaker. We find that transparency improves the accuracy of decisions made in state L, the state in which the adviser has a monetary incentive to give a "dishonest" recommendation. While transparency improves the accuracy of decisions in state L, it has no impact on the accuracy in state R. Thus, we conclude that transparency improves decision making.

The mechanism through which decision making is improved in the transparency condition is consistent with our theory. Many decision makers and advisers view rejecting the payment from the third party as a way to enhance the adviser's credibility. When the state is R, more advisers reject the payment and recommend r in the transparency treatment than in the non-transparency treatment. When the state is L, many advisers reject the payment and recommend the correct policy even in the non-transparency treatment—this can be explained by lying-aversion—however, even more advisers do so with transparency. Thus, advisers' willingness to reject the payment in the transparency treatment stems not only from lying-aversion but also from strategic signalling of honesty. Signalling honesty by refusing the third-party payment is potentially beneficial because if an adviser accepts the payment and recommends policy r in the transparency treatment, a sizeable proportion of decision makers find it suspicious and choose policy l. One caveat is that although the fraction of decision makers who mistrust advisers that accept the payment is nonnegligible, it is also far from being a majority. As a result, positive effects of transparency weaken over time: Many advisers learn that the negative bias among decision makers against advisers who accept the third-party payment is not too prevalent and adjust their behavior accordingly.

Our study sheds light on the effects of transparency on the adviser and the decision maker behavior in an environment where the adviser might be swayed by a third-party who favors the implementation of a specific policy. In particular, ours is the first study which shows both theoretically and experimentally that transparency may help decision makers—especially if transparency is mandatory. Transparency is becoming more and more important because the share of private enterprise in the funding of research has been rising steeply. According to the National Science Foundation (NSF) in the US, the share of industry and government roughly tracked each other until late 80's. However, industry has since considerably outpaced government in terms of research funding.[10] This has bestowed corporations with an immense influence and ability to shape policy-making as well as public opinion

[10] http://www.nsf.gov/statistics/seind14/index.cfm/chapter-4/c4s1.htm

via the research institutions and experts that they fund.[11] In response to this, regulatory agencies, academic journals, NGOs and government agencies have started demanding more transparency. However, this demand for transparency is not without its backlash, as we discussed. Therefore, transparency is a not only very important but also sensitive issue that calls for a thorough theoretical and empirical evaluation.

The paper proceeds as follows. Section 2 discusses the related literature. Section 3 presents the theoretical model and Section 4 the theoretical predictions. Sections 5 and 6 describe the experimental setup and the experimental results, respectively. Presenting all three conditions in full generality requires a considerable increase in length and complexity of exposition with little corresponding increase in insight; therefore, we focus on only the transparency and non-transparency conditions throughout Sections 3-6. The main results for the voluntary-transparency condition are presented in Section 7. Section 8 concludes.

2 Related Literature

2.1 Experimental Literature

Prior experimental research on transparency pointed out that disclosing conflict of interests is either ineffective or harmful to decision makers (Cain *et al.*, 2005, 2011; Koch and Schmidt, 2010; Rode, 2010; Loewenstein *et al.*, 2011; Loewenstein *et al.*, 2012; and Ismayilov and Potters, 2014). Loewenstein et al. (2011) put forth two explanations for the adverse effects of transparency that they observe. Firstly, transparency may result in "moral licensing." That is, information disclosure can undermine the willingness of the adviser to engage in moral behavior, which harms the decision maker. Secondly, decision makers fail to discount recommendations by advisers whose biases are disclosed. Via these two channels, decision makers obtain worse outcomes with transparency.

As we already mentioned, the most important distinction of our model from the previous experiments is that the adviser can choose to avoid a conflict of interest. In prior research, advisers did not have a choice regarding their incentives and were always conflicted.

[11] For example, the energy industry, the tobacco industry, and the sugar industry have used experts extensively in their endeavour to supress information regarding the hazards of their products. One highly publicised case involves ExxonMobil, the world's biggest oil company. ExxonMobil made the headlines recently due to its funding of climate change denial. It was reported that the company knew as early as 1981 of climate change, and despite this, the company spent a huge amount of money over the next three decades to promote its denial (The Guardian, July 8, 2015). On a related note, 47 percent of US voters believe that climate change is caused by human activities, compared with 97 percent of climate scientists (Yale Project on Climate Change Communication 2013).

The only experiment other than ours that documents positive effects of transparency is the independent study by Sah and Loewenstein (2014), which we became aware of while completing our work. Sah and Loewenstein (2014) does not provide a formal model of advice. In their experimental design, the adviser can accept or reject a conflict of interest. If the adviser accepts the conflict of interest, then the adviser and the decision maker have misaligned payoffs; otherwise, their payoffs are perfectly aligned. In the transparency treatment, the decision maker is informed whether or not the adviser accepted the conflict of interest. Our model differs from their design because accepting the third-party funding is not always equivalent to having a conflict of interest as an adviser who accepts the funding in state R will give honest advice. More generally, our design is consistent with the idea that a relationship between the adviser and the third party is not inherently wrong.[12]

Our model and experimental results also relate to studies on lying aversion. Previous experiments have shown that people may find lying morally costly and avoid it (see, for example, Gneezy, 2005; Sanchez-Pages and Vorsatz, 2007; Hurkens and Kartik, 2009; Gibson *et al.*, 2013). Our results are consistent with previous findings. Many advisers choose to tell the truth even if there is a material incentive to lie.

2.2 Theoretical Literature

There is an extensive literature on strategic information transmission from a better-informed sender to a receiver, dating back to the seminal works by Crawford and Sobel (1982) and Sobel (1985). In this literature, the receiver and the sender have payoffs that are misaligned to a certain degree. This misalignment results in a bias in the sender's communication; the sender's equilibrium message to the receiver is noisy and can even be uninformative, depending on the precise structure of payoffs.

In most of the literature on strategic information transmission, the extent to which sender and receiver payoffs are (mis)aligned is exogenous. Our model endogenizes it since the adviser may choose to decline the third-party payment, and transparency affects the incentives of the adviser to do so. To be more precise, our transparency condition allows the adviser to engage in *costly signaling*; that is, the adviser can reject the payment and signal to the decision maker that their payoffs are aligned.

The extent to which the adviser and the decision maker payoffs are aligned is also

[12] On the empirical side, research by Djankov et al. (2010) investigated the rules and practices of financial and conflict disclosure by members of Parliament in 175 countries, and documented that a public form of disclosure (but not internal disclosure to parliament) is positively correlated with government quality. The authors do not provide a causal link but they suggest that disclosure might be a significant ingredient of political accountability.

endogenous in Durbin and Iyer (2009) and Inderst and Ottaviani (2012) but this is due to strategic third parties. There is a single third party in Durbin and Iyer (2009) whereas there are multiple third parties in Inderst and Ottaviani (2012). In these models, third parties set side payments for the adviser (sales commissions, bribes, etc.) in order to maximize their return. Inderst and Ottaviani (2012) also analyzes the effects of transparency in their setting and show that transparency can have adverse effects on the decision maker, unlike in our model. The focus of Inderst and Ottaviani is on a setting different than ours: Third parties produce horizontally-differentiated products and compete for consumers using advisers who are incentivized by sales commissions. We abstract from the competition of third parties and focus on the effects of transparency with a single—or disproportionately powerful—third party (e.g., oil and energy industry, tobacco industry, sugar industry, gun rights lobby, *etc.*) that can affect policy and sway public opinion through political lobbying and funding research institutes and experts.

Also related is the model by Potters and van Winden (1992, 2000) in which there are two players, a policy-maker and a better-informed interest group (i.e., a lobby) that can send the policy-maker a costly message. Their model differs from the cheap talk literature since sending a message entails a cost to the lobbyist—this relates to lobbying costs. Our model can be thought of as a lobbying model (absent lobbying costs) in which there is a third party that can influence the message of the lobbyist to the policy-maker.

3 The Model

We develop a model that involves two active players, a decision maker (D) and an adviser (A), and an inactive third-party, namely the special interest group (SIG).[13] There are two possible states of the world. Nature draws the state $S \in \{L, R\}$ such that

$$S = \begin{cases} L, & \text{with probability } p \\ R, & \text{otherwise.} \end{cases}$$

The prior probability p is common knowledge. The adviser (A) learns the true state of the world whereas the decision maker (D) knows only the prior p. After learning the state, A recommends a policy $s \in \{l, r\}$ to D, who then chooses a policy. D prefers the policy to match the state of the world. After learning the state and prior to making a recommendation,

[13] In our experiment, we want to focus on the decision of the adviser regarding the side payment with and without transparency and how the decision-maker interprets the advice from an adviser who accepted the payment. Therefore, we abstract from the strategic behaviour of the SIG and assume that it always offers a side payment to the adviser and that the side payment is an exogeneously determined amount.

A decides whether or not to accept a payment from the special interest group (SIG).[14] SIG strictly prefers policy r regardless of the state of the world, and if A accepts the payment, then he must recommend r. If A rejects the payment, then he decides whether to recommend l or r.

In Section 4.4., we discuss our modeling assumptions and their relevance in light of the evidence regarding industry-funded research. In that section, we also show that our main results are robust to relaxing these assumptions. In particular, the main results are robust to assuming that the adviser is (ex-ante) imperfectly informed and obtains higher quality information if he accepts the third-party payment, and that the adviser is free to choose his recommendation (with some probability) even if he accepts the payment.

We consider the following scenarios in our analysis.

(i) In the "transparency condition", D learns whether or not A accepted the payment before choosing the policy.
(ii) In the "non-transparency condition", D has no information about the decision of A regarding the payment.
(iii) In the "voluntary-transparency condition", transparency is not enforced and A chooses whether or not to disclose his decision regarding the third-party payment.

As stated before, presenting all three conditions in full generality requires a considerable increase in length and complexity of exposition with little corresponding increase in insight; therefore, we focus on a comparative analysis of transparency and non-transparency conditions throughout Sections 3-6; Section 7 presents the main results for the voluntary-transparency condition.

3.1 Payoffs

D obtains a payoff of $\bar{\pi}$ if the chosen policy matches the state of the world, and a payoff of $\underline{\pi}$ otherwise, where $\underline{\pi} < \bar{\pi}$. A receives

(i) $\alpha > 0$ if D chooses the policy that A recommends, and
(ii) $\gamma > 0$ if A recommends the "better" policy for D, and
(iii) $\beta(S) \geq 0$ if A accepts the payment in state $S \in \{L, R\}$,

We assume that $\alpha > 0$ because being followed is good for the adviser's reputation and enables future business prospects. To be more concrete, if (for example) the adviser is a lobbyist, then α will materialize as better access to policy makers and future lobbying

[14] Therefore, accepting the payment is not inherently dishonest: The timing of the model allows the adviser to accept the side payment only in state R and remain honest.

prospects. If the adviser is a scientist, then α represents greater influence in the adviser's field, scientific publications, research grants, and professorial positions.

Importantly, in our modelling choice, we have in mind settings where there may be a significant lag between the advice and the resolution of uncertainty regarding the true state, as in our example of silica dust regulation. The resolution of uncertainty will take (or has taken) decades in numerous, important circumstances—consider the relationship between tobacco use and cancer, the relationship between a high-sugar diet and coronary heart disease, the climate change debate, the current debate on asylum and immigration policies, the debate on genetically modified organisms, the debate on affordable healthcare policies, and so on. This means that if an expert chooses to provide advice that goes against his own information, he cannot easily be held accountable.[15] Hence, α represents the financial benefit from being consulted and followed, irrespective of the consequences, because it may take a long time to evaluate the correctness of the advice—in some cases, even longer than the expert's lifetime.

Unlike α, the parameter γ captures the intangible consequences of A's recommendation. We assume that $\gamma > 0$ due to reasons such as lying aversion or having a concern for the outcome. Note that our modeling approach aligns the payoffs and the interests of the adviser and the decision maker in the absence of third-party payments—see Section 4.1 for details. It is also consistent with the experimental literature which has shown that people may find lying morally costly and avoid it. An alternative way of modeling A's concern for truthtelling or for the outcome would be to assume that A obtains a payoff $\gamma > 0$ if D chooses the "better" policy. However, this approach gives rise to pathological equilibria in which A tells the truth by lying; such an equilibrium cannot be eliminated by reasonable equilibrium selection criteria—see Section 4.1 for details. Therefore, we impose $\gamma > 0$ as a payoff for "explicit" truthtelling, to have cleaner predictions for the experiment.

We allow for state-dependent side payments; for example, it is intuitive to allow for $\beta(L) > \beta(R)$ because L is the state in which an adviser who accepts the side-payment has to give a dishonest recommendation and lose $\gamma > 0$—this is not the case in state R.[16]

[15] On a related note, a meta-study of (anonymous) scientist surveys by Fanelli (2009) found that on average, over 14% of respondents reported to have observed fabrication, falsification and modification of data by colleagues, and up to 72% have observed other questionable practices. However, Fanelli (2009) notes that fabrication and falsification are very hard to detect in the data, and are rarely reported by whistleblowers.

[16] Note that $\beta(L)$ and $\beta(R)$ can differ if, for example, SIG has its in-house researchers; this is indeed common practice (see the case of ExxonMobil in Footnote 10).

4 Equilibrium Analysis

Let $a \in \{r_A, r_R, l\}$ denote A's action, where $a = r_A$ if A accepts the payment and recommends r; $a = r_R$ if A rejects the payment and recommends r; and $a = l$ if A recommends l (recall that A *must* reject the payment in order to recommend l). Next, let m denote A's "message" to D. In the transparency condition, D observes A's action completely. So, A's message is equivalent to his action; i.e., $m = a$. In the non-transparency condition, A's message is equivalent to his policy recommendation. If $a \in \{r_A, r_R\}$, then $m = r$. If $a = l$, then $m = l$. Hence, the message space is $\{r_A, r_R, l\}$ in the transparency condition and $\{r, l\}$ in the non-transparency condition.

We look for the perfect Bayesian Equilibria (PBE) of the games described above. Let $\Pr_A(m|S)$ denote the probability that A chooses message m in state S. Also, let $\mu(S|m)$ denote the belief D attaches to state S conditional on m, and let $\Pr_D(s|m)$ denote the probability that D chooses policy s after observing message m. In a PBE, (i) D's strategy maximizes his payoff given m and $\mu(S|m)$; (ii) $\mu(S|m)$ is formed using A's strategy $a(S)$ by applying Bayes' rule whenever possible; and (iii) given $\mu(S|m)$ and D's strategy, A's strategy maximizes his payoff.

In the main text, we impose the inequality $\beta(L) > \gamma$ because it rules out the case in which transparency does not make difference relative to the non-transparency condition.[17] For example, if $\gamma > \beta(L)$, then it is easy to see that being honest is very important to the adviser, and the adviser will never give dishonest advice "in a reasonable equilibrium" with or without transparency.

Assumption 1 $\beta(L) > \gamma$.

We also maintain the assumption that $p < 1/2$ and implement it in our experimental design.

Assumption 2 $p < 1/2$.

We show in Section 4.3 that transparency is *never* harmful to the ex-ante accuracy of decision making and sometimes strictly improves it. Importantly, this is also true under the assumption that $p \geq 1/2$.[18] More generally, transparency never harms decision making regardless of whether or not Assumptions 1 and 2 hold. However, the assumption that $p < 1/2$ gives rise to a clean and stark contrast between the equilibrium predictions for the transparency and non-transparency conditions, which is convenient for our experimental design (Footnote 22 provides more details).

[17]We prove this claim in the Online Appendix at the end of the proof of Proposition 2.

[18]The proof of this claim is omitted due to space considerations and is available upon request.

4.1 Baseline Model: The Game Without SIG

We first study the model without SIG, as a benchmark for subsequent analysis. This analysis illustrates in a simple way the need for reasonable equilibrium selection criteria. Since there is no SIG, $a \in \{l, r\}$ and $a = m$. In this case, there are two possible equilibria. The first equilibrium is fully informative: A always recommends the correct policy and D follows the recommendation. The second one is uninformative: A always recommends r regardless of the realized state, and D always follows.[19] The second equilibrium is not only pareto-dominated but also dependent on the following "unreasonable" out-of-equilibrium belief. If A recommends l, then D interprets this as "strong evidence" for state R, which deters A from recommending l in state L.[20] But why should D believe that A is more likely to recommend l in state R than in state L? This is not *intuitive* given that A obtains an additional payoff $\gamma > 0$ if he recommends l in state L, not in state R. In order to eliminate equilibria (with and without SIG) that rely on unreasonable out-of-equilibrium beliefs, we impose a modification of the Intuitive Criterion by Cho and Kreps (1987).

Definition 1 *Message m is* ***equilibrium-dominated in state*** *$S \in \{L, R\}$ if the equilibrium payoff of A in S is greater than the highest possible payoff of A from m in S.*

Given this, we present a version of the Intuitive Criterion suited to our two-state setting.

Definition 2 An equilibrium **fails to satisfy the Intuitive Criterion (IC)** if there exists an out-of-equilibrium message m such that:
(i) m is equilibrium-dominated in state S; and
(ii) m is a profitable deviation for A in state $S' \neq S$ if D best-responds to m according to the belief $\mu(S|m) = 0$.

The uninformative equilibrium does not satisfy the Intuitive Criterion (IC) because recommending l is equilibrium dominated in state R, and recommending l in state L is a profitable deviation for A if D best-responds according to the belief $\mu(R|l) = 0$.[21] As a result, the only PBE that satisfies the IC in this game is fully informative: A always provides truthful advice and D follows the advice. As mentioned before, if we assumed that A obtains $\gamma > 0$ conditional on D choosing the "better" policy (instead of conditional on A recommending the better policy), we would have a pathological equilibrium in which A tells the truth by lying. To be more precise, in this case there exists an equilibrium such that A always recommends

[19] Note that this equilibrium exists if and only if $\alpha \geq \gamma$.

[20] More formally, either $\mu(R|l) > 0.5$ and $\Pr_D(l|l) = 0$ or $\mu(R|l) = 0.5$ and $\Pr_D(l|l) < 5/6$ holds in the uninformative equilibrium. Both deter A from recommending l in state L.

[21] The equilibrium payoff of A in state R is $\alpha + \gamma$ whereas the highest possible payoff from choosing l in state R is α, which is smaller than $\alpha + \gamma$ as $\gamma > 0$.

the worse policy—thus, A's behavior is fully informative, and D chooses the better policy. This equilibrium cannot be eliminated by IC. To avoid pathological cases, we impose $\gamma > 0$ as a payoff for "explicit" truthtelling.

4.2 The Game With SIG and Non-transparency

Next, we analyze the game with SIG assuming that D cannot observe the decision of A regarding the side payment. We show that there exists a generically unique equilibrium, which we denote as the "corrupt equilibrium." In the corrupt equilibrium, A always chooses r_A (i.e., accepts the payment and recommends r) and D always follows A's advice and chooses r.[22]

Proposition 1 *There exists a (generically) unique equilibrium in the non-transparency condition. On the equilibrium path, A always accepts the payment and recommends r, and D follows the advice with probability one.*

Proof. In the Online Appendix. ■

It is easy to see how the corrupt equilibrium arises.[23] Since A's message is always r in the corrupt equilibrium, it is uninformative and the decision maker must rely on his prior to choose a policy. Then, it is in the best interest of D to choose r—state R is more likely than state L by Assumption 1. Given this, A attains the highest possible payoff in both states if he chooses r_A and has no incentive to deviate. Thus, the belief $\mu(L|r) = p < 0.5$ is consistent with A's strategy.[24]

4.3 The Game With SIG and Transparency

We now analyze the game assuming that D observes whether or not A accepted the payment from the SIG. We focus on equilibria that satisfy the IC. For brevity in the main text, we assume that $\alpha + \gamma > \beta(L)$—this is the case which we implement in the experiment. However, in the Appendix we characterize fully the set of equilibria that satisfy the IC without this restriction and show that, regardless of whether or not $\alpha + \gamma > \beta(L)$ holds, transparency

[22] What we mean by generic uniqueness is that the equilibrium path play and payoffs are unique. However, there are various belief systems that support the corrupt equilibrium because A never chooses l in equilibrium.

[23] This is where the assumption that $p < 1/2$ matters; a corrupt equilibrium exists and is the unique equilibrium only if $p < 1/2$. As a result, the comparison between the transparency and non-transparency conditions is much more clean with regard to our experimental purposes.

[24] It may seem strange in that D takes advice from A even though this has no informational advantage. However, taking advice has no cost and allowing for a tiny probability that A never lies (indeed some advisers do not lie in our experiment) makes taking advice strictly optimal.

(i) never harms the ex-ante accuracy of decision making and (ii) strictly improves it in an equilibrium if $\alpha \geq \beta(R)$. Under the assumption that $\alpha + \gamma > \beta(L)$, there exist at most three equilibria. The first is the corrupt equilibrium in which A always chooses r_A and D always chooses r. The second is what we denote the "honest equilibrium." The honest equilibrium is fully informative: A always rejects the payment and provides honest advice, and D follows A's advice. The honest equilibrium exists if $\alpha \geq \beta(R)$ holds. Finally, there exists a mixed-strategy equilibrium if $\beta(L) \geq \beta(R) + \gamma$. In this equilibrium, A always chooses r_A in state L (i.e., accepts the payment and gives false advice), and randomizes between r_R and r_A in state R; and D randomizes between l and r if A chooses r_A and chooses r if A chooses r_R.

The mechanism of the corrupt equilibrium is identical to that in the non-transparency condition. As for the honest equilibrium, first note that A must reject the payment in both states in order to sustain truthful information revelation because the side-payment $\beta(L)$ exceeds γ by Assumption 1 and thus, there is a significant "agency problem." If A always gave honest advice and yet accepted the payment only in state R, then D would be better off following A's recommendation. But this would give A an incentive to deviate and choose r_A in state L.

Proposition 2 *Assume that Assumptions 1-2 and $\alpha + \gamma > \beta(L)$ hold. There are (at most) three equilibria that satisfy IC:*

1) *Corrupt Equilibrium:* $\Pr_A(r_A|L) = \Pr_A(r_A|R) = 1$; $\Pr_D(r|r_A) = 1$; $\mu(R|r_A) = 1 - p$.[25]

2) *Honest Equilibrium (if $\alpha \geq \beta(R)$):* $\Pr_A(l|L) = 1$; $\Pr_A(r_R|R) = 1$; $\Pr_D(r|r_R) = 1$; $\Pr_D(l|l) = 1$; $\mu(R|r_R) = 1$, $\mu(R|l) = 0$ and $\mu(R|r_A) \leq 0.5$; $\Pr_D(r|r_A) = 0$.

3) *Mixed-strategy Equilibrium (if $\beta(L) \geq \beta(R) + \gamma$):* $\Pr_A(r_A|L) = 1$; $\Pr_A(r_A|R) = \frac{p}{1-p}$ and $\Pr_A(r_R|R) = \frac{1-2p}{1-p}$; $\Pr_D(r|r_R) = 1$; $\mu(R|r_R) = 1$; $\Pr_D(r|r_A) = 1 - \beta(R)/\alpha$; $\mu(R|r_A) = 0.5$.[26]

Proof. In the Online Appendix. ∎

Our "normative" measure of interest is the ex-ante accuracy of decision making rather than the efficiency of the aggregate expected payoff because the former has wider social welfare implications. In that sense, the honest equilibrium is superior to both the corrupt equilibrium and the mixed-strategy equilibrium whereas the corrupt equilibrium and the

[25] Just as in the corrupt equilibrium with nontransparency, there are various out-of-equilibrium beliefs that support the equilibrium.

[26] In the knife-edge case with $\beta(R) + \gamma = \beta(L)$, there is also a continuum of mixed-strategy equilibria in which A randomizes between r_A and l in state L and between r_A and r_R in state R. See the details in the Appendix.

mixed-strategy equilibrium are ex-ante identical.[27] Thus, transparency increases the (ex-ante) expected accuracy relative to non-transparency. To reiterate, transparency is *never* harmful to the ex-ante accuracy of decision making in our model and can strictly improve it regardless of whether or not Assumptions 1-2 and $\alpha + \gamma > \beta(L)$ hold.

4.4 Extensions

We close the theory section with a discussion of richer variations of our model. It may be argued that in certain fields of expertise, the adviser may use third-party payment in order to acquire more accurate information. However, allowing for this cannot change our main results unless we relax the assumption that the adviser who accepts the third party funding must recommend r. Our results are robust to jointly assuming that accepting the payment increases the accuracy of the adviser's information and an adviser who accepts the third party payment can choose between recommending l or r with a positive—but not too high—probability.

Assume, as an example, that the adviser who accepts the third party payment is perfectly informed about the state whereas an adviser who does not accept it gets a signal $\sigma \in \{L, R\}$ about the state such that $\Pr(\sigma|S) = 0.8$ if $\sigma = S$. Further, assume that $p = 0.4$, $\alpha = 6$, $\beta(L) = 5$, $\beta(R) = 2$, $\gamma = 1$ (these are the parameter values that we use in our experimental design). If the adviser who accepts the side payment is free to choose the policy recommendation with a probability that is weakly lower than 0.38, then our main findings go through. For example, an honest equilibrium still exists and is the most informative equilibrium—even though the adviser is perfectly informed about the state only if he accepts the side payment. If, however, the aforementioned probability is higher than 0.38, then an honest equilibrium that satisfies the Intuitive Criterion does not exist in the transparency condition.

How plausible is it to assume that an adviser who accepts the third party payment will have to recommend the policy that the third party favors (with a sufficiently high probability)? We believe that it is a reasonable assumption in many circumstances given the body of evidence accumulated so far. In a very recent study, Kearns *et al.* (2016) documented that sugar industry executives funded prominent nutrition scientists in 1960s in return for review articles that would falsify studies which pointed out that a high-sugar diet posed a major risk for coronary heart disease (CHD). Moreover, the sugar industry successfully cast doubt about the hazards of sugar for decades while promoting fat as the main

[27] The ex-ante expected accuracy is 100% in the honest equilibrium whereas it equals $100(1-p)\%$ in both the corrupt equilibrium and the mixed-strategy equilibrium.

dietary risk for CHD. These findings led Kearns *et al.* (2016) to argue that "policymaking committees should consider giving less weight to food industry–funded studies".[28] Another highly-publicized case involves tobacco companies, which have been repeatedly sued both for fraud—hiding from the public what they knew about their product—and in order to recover health costs associated with smoking, employed dozens of experts in order to testify on their behalf. Kenneth Ludmerer, currently a distinguished professor of history and medicine at Washington University in St. Louis testified as an expert on medical history on behalf of the tobacco industry over a period of 15 years. From the testimony of Ludmerer in 2002:

Question: Doctor, is it your opinion that cigarette smoking contributes to the development of lung cancer in human beings?
Answer: I have no opinion on that.

Ludmerer's testimony implies that tobacco companies cannot be held liable for any wrongdoing since he, the expert, has no opinion on whether cigarette smoking contributes to the development of lung cancer—as recently as in 2002. Ludmerer was reportedly paid more than $550,000 by the tobacco industry (Delafontaine, 2015). It seems unlikely that the industry would pay this amount to an expert who could give an affirmative answer to the question above. On the contrary, Robert Proctor and Louis Kyriakoudes, two (of only three) experts who testified against tobacco companies were subject to harassment by the industry (Delafontaine, 2015).[29]

5 Experimental Design and Hypotheses

As Propositions 1 and 2 indicate, our theoretical predictions regarding the comparison of the transparency and the nontransparency conditions are ambigious due to equilibrium multiplicity. To gain further insights regarding the behavioral effects of transparency, we designed and ran an experiment which implemented our model. Our aim is to empirically answer the following research questions:

1) Does transparency help decision makers?

2) Do advisers behave differently if decision makers observe whether or not they accepted the side payment? In particular, are advisers less likely to accept the payment in order to

[28] Following the media interest regarding this discovery, Dr. Walter Willett, chairman of the nutrition department at the Harvard University stated that while academic conflict-of-interest rules had changed since the 1960s, the discovery was a reminder of "why research should be supported by public funding rather than depending on industry funding" (The New York Times, September 13, 2016).

[29] http://www.thenation.com/article/big-tobacco-and-historians/

"signal" their honesty?

3) Do decision makers take into account or ignore the decision of advisers regarding the side payment when they choose the policy?

4) Does the accuracy of decisions depend on whether transparency is mandatory or voluntary?

We ran three treatments implementing each of the transparency, non-transparency, and voluntary-transparency conditions. Subjects participated in only one of the three treatments. The experiment consisted of 40 rounds. We used stranger matching; i.e., subjects were randomly rematched with a new counterpart each period. At the beginning of the experiment, each subject was randomly assigned to be a receiver (i.e., decision maker) or a sender (i.e., an adviser). Subjects remained in the same role for the first 20 rounds. After 20 rounds were over, the roles were switched and subjects remained in their new role until the end of the experiment. Subjects were not informed at the beginning of the experiment that roles would be switched after the first 20 rounds. We applied role-switching because it may facilitate learning.

Before the start of the actual experiment and the assignment of roles, subjects went through a tutorial and answered control questions in order to enhance their understanding of the game. Once the tutorial was over, the actual experiment began. Decision making in each round of the experiment was as described in the theory section. We used the strategy method with decision makers, giving us more observations at each information set and analyze better whether subjects use equilibrium strategies. We used the strategy method only with decision makers because the game would be more difficult to explain if we used the strategy method with both advisers and decision makers.

Parameter values that we used in the experiment are as follows: $p = 0.4$, $\alpha = 6$, $\beta(L) = 5$, $\beta(R) = 2$, $\gamma = 1$, $\bar{\pi} = 10$, $\underline{\pi} = 5$. Given these values, there exist three equilibria that satisfy the IC, and they are as described in Proposition 2. As discussed before, having $\beta(L) > \beta(R)$ is intuitive because L is the state in which an adviser who accepts the payment has to give a dishonest recommendation and loses $\gamma > 0$. Moreover, $\beta(R) > 0$ because, otherwise, accepting a third-party payment is unequivocal evidence that the state is L. For the transparency and the voluntary-transparency treatments to have a good shot at improving experimental decision making, we chose parameter values such that the honest eequilibrium satisfies not only IC but also D1 (Cho and Kreps, 1987), which is a more stringent condition than IC. Proposition 3 in the Appendix provides the definition of D1 as well as the set of equilibria that satisfy D1.[30]

[30] We have also run sessions in which all the parameter values were the same except that $\beta(R) = \beta(L) = 4$.

We used neutral language in the treatments, and there was no mention of a special interest group. Instructions only stated that the sender could choose to accept an extra payment, in which case the sender had to recommend r.[31] Other than the adviser's recommending l or r, there was no communication between the adviser and the decision maker. The experiment was conducted at the experimental laboratory of Vienna Center for Experimental Economics (VCEE) at the University of Vienna. Subjects were recruited from the general student population in Vienna via e-mail solicitations. We ran three sessions for each treatment. In each session, there were 24 subjects. At the beginning of each session, each subject was randomly assigned to a matching group of 12 subjects. There was no interaction between subjects assigned to different matching groups. Therefore, we have six independent observations for each treatment.

All sessions were conducted using a computer program written in Z-Tree (Fischbacher, 2007). Earnings were denoted in "points" which were exchanged at the rate of 1 euro for 3 points. Each subject was paid for two randomly drawn rounds in each role. Sessions lasted around 75 minutes and the average payoff per subject was approximately 16 euros (including a fee for completing the questionnaire at the end of the experiment).

Table 1: Decision Maker Payoffs by State and Equilibrium type

Equilibrium Type	**State L**	**State R**
Corrupt Equilibrium	5	10
Honest Equilibrium	10	10
Mixed-Strategy Equilibrium	6.67	8.89

Finally, we present the two experimental hypotheses that we derive from our theoretical results given the parameters used in the design. The first hypothesis is the accuracy of decision making in each state. Note that even though the hypothesis refers to decision maker payoffs, there is a one-to-one mapping from decision maker payoff to accuracy.

Hypothesis 1 (a) *Transparency makes decision makers better off in state L (via the honest equilibrium and the mixed-strategy equilibrium).* **(b)** *Transparency makes decision makers worse off in state R (via the mixed-strategy equilibrium).*

In this case, the honest equilibrium is not robust to D1, and in the experiment, transparency did not have a statistically significant impact on decision-making, and it had either zero or little impact on other dimensions. Therefore, we decided to move forward with treatments where the honest equilibrium is robust to both IC and D1.

[31] Instructions can be found in the supplementary appendix.

We derived Hypothesis 1 taking into account all possible equilibria given the realized state. In other words, Hypothesis 1 concerns interim payoffs. Part a is due to the following. L is the state in which A has a financial incentive to give a "dishonest" recommendation—A accepts the third-party payment and recommends r in state L in the corrupt equilibrium of either treatment. As a result, D makes the wrong decision in state L and obtains a payoff of 5 in the corrupt equilibrium. The corrupt equilibrium is the unique equilibrium of the non-transparency treatment. However, the transparency treatment has other equilibria (namely, the honest equilibrium and the mixed-strategy equilibrium) which make D better off in state L than the corrupt equilibrium (see Table 1 for decision maker payoffs by state and equilibrium type). Part b follows only because of the mixed-strategy equilibrium in the transparency treatment: If the state is R, D always makes the correct decision in the corrupt equilibrium and the honest equilibrium, but there are mistakes in the mixed-strategy equilibrium.

Finally, the existence of the honest equilibrium and the mixed-strategy equilibrium in the transparency treatment generates the following hypotheses about adviser behavior.

Hypothesis 2 *Honest equilibrium behavior is more prevalent under transparency in both L and R—i.e., transparency increases the fraction of advisers who reject the payment and recommend the correct policy in both states.*

6 Results

6.1 Does Transparency Help Decision makers?

In this section, we analyze the accuracy of decisions by treatment. We start with state L. Hypothesis 1 states that transparency makes the decision maker better off in state L. As explained before, this is because the transparency treatment has the honest equilibrium and the mixed-strategy equilibrium both of which are better than the corrupt equilibrium in state L for the decision maker.

Table 2: Proportion of Correct Decisions by State and Treatment

	Transparency	**Non-transparency**	**Mann-Whitney p–value**
state L	0.461	0.297	0.008
state R	0.756	0.793	0.169

Note: One-sided Mann-Whitney tests using matching-group averages as observations (n=12).

Table 2 displays the proportion of correct choices by state and treatment: 46.1% of decisions in state L are correct in the transparency treatment, whereas only 29.7% are correct in the non-transparency treatment. The difference across the two treatments is statistically significant according to a one-sided Mann-Whitney test ($p = 0.008$).[32] We conclude that transparency improves the accuracy of decision makers in state L, in line with Hypothesis 1 (a).

The decision maker accuracy in state L in the non-transparency treatment is much larger than the theoretical prediction of zero mainly due to two reasons. As the next section shows, there are many advisers who recommend the correct policy in state L rejecting the payment even in the non–transparency treatment, likely due to lying-aversion. Decision makers who follow such advisers make the correct decision in state L.[33]

Next, we study the accuracy of decisions in state R. Recall that Hypothesis 1 predicts that the decision maker is worse-off in state R with transparency but this is only because of the mixed strategy equilibrium (see Table 1). Table 2 shows that 75.6% of decision makers choose policy r if the state is R in the transparency treatment. The percentage in the non-transparency treatment is 79.3%. This difference between the two treatments is not statistically significant. Thus, there is no statistical support for Hypothesis 1 (b).

Overall, our answer to Question 1 in Section 5 is affirmative as transparency makes decision makers better off in state L and has no effect in state R.

Figure 1 displays the proportion of correct choices in state L by period and treatment. The figure shows that there is no discernible time trend in either treatment. We obtain the same conclusion from a random effects panel probit regression of the probability a decision maker chooses l in state L with the round number as the independent variable. We present the regression result for the transparency treatment in Table 7 in the Appendix.[34]

[32] The decision rule implicit in our one-sided Mann-Whitney test is that if the null hypothesis that the two distributions do not differ is rejected, then there is evidence that the random variable "accuracy with transparency" tends to be larger than the random variable "accuracy with nontransparency". While this type of decision rule is very common in scientific research, it is admittedly restrictive in the sense that the researcher focuses on a limited set of distributions. Therefore, when we find a significant difference using a Mann-Whitney test, we also ran a stochastic inequality test which allows us to infer a directional difference without limiting the set of distributions under consideration (Schlag, 2008). According to a one-sided stochastic inequality test, accuracy in state L tends to be higher in the transparency treatment ($p = 0.031$).

[33] In the nontransparency treatment, there are also some decision makers who choose policy l even if the adviser recommends r, and that happens to be the correct decision if the state realization is L (see Section 6.3).

[34] The regression results for the non-transparency treatment are available upon request. Table 7 also presents regression results using the data before role switching and after role switching. The coefficient of the trend variable is significant only in the second half, which suggests that the accuracy of decisions made in state L declines over time in the second half of the transparency treatment.

Figure 1: Decision maker Accuracy in State L

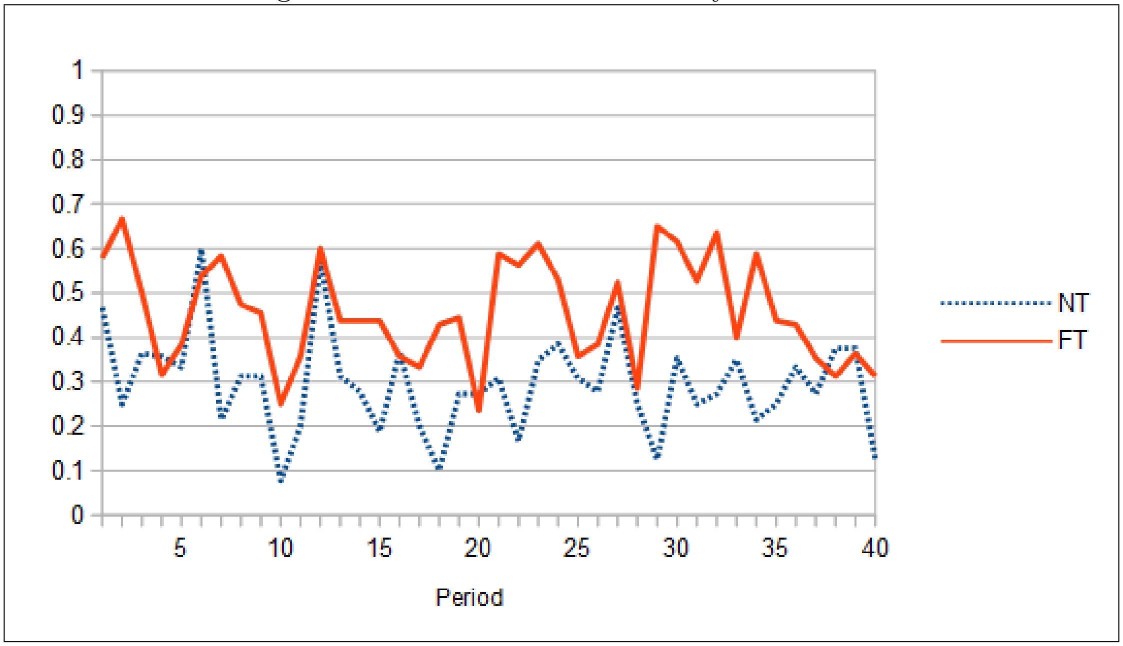


Figure 2 displays the proportion of correct choices in state R by period and treatment. Just like in state L, there is no discernible time trend in either treatment. We obtain the same conclusion from a random effects panel probit regression of the probability a decision maker chooses r in state R as a function of the round number. The regression result for the transparency treatment is given in Table 7.[35]

Next, we have a detailed look at subjects' behavior in order to answer Questions 2 and 3.

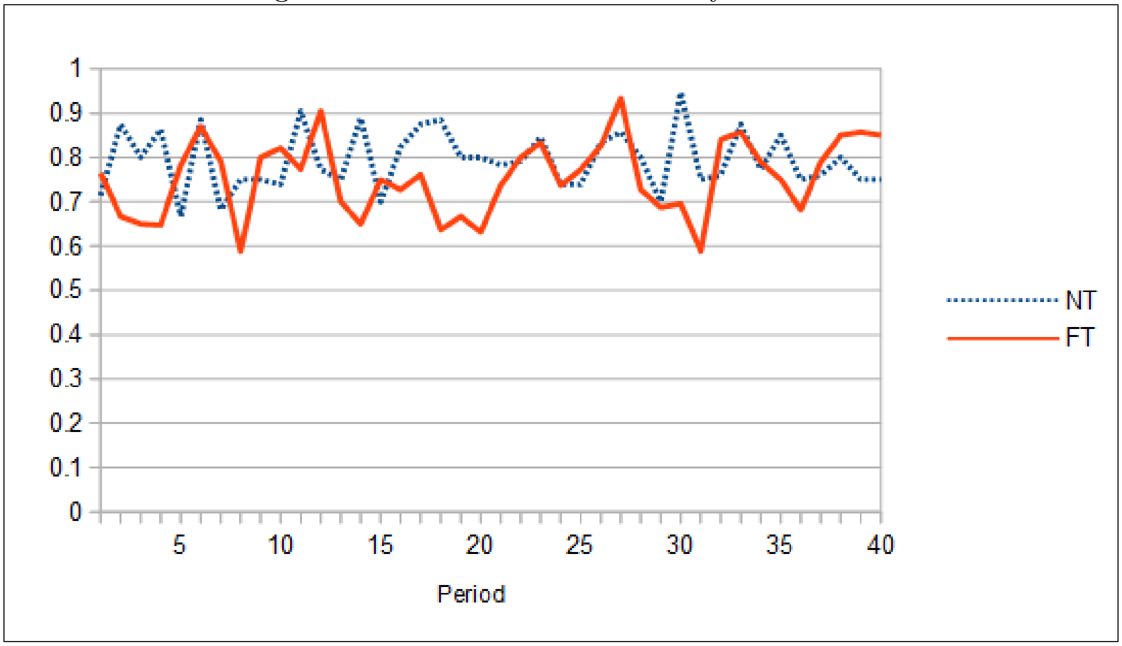
6.2 Adviser Behavior

Hypothesis 2 predicts that transparency increases the fraction of advisers who behave in accordance with the honest equilibrium (i.e., choose l in state L and choose r_R in state R). The relevant data is summarized in Table 3.

We start by analyzing A's behavior in state L. Table 3 shows that the average fraction of advisers who reject the payment and recommend l in state L is 30.3% in the transparency treatment and 22.3% in the non-transparency treatment. The difference across the two treatments is weakly significant (one-sided Mann-Whitney: p-value = 0.075). However, if we consider the data before and after role-switching separately, the treatment difference is

[35] The coefficient of the trend variable is also insignificant when we run separate regressions using the data before role switching and after role switching.

Figure 2: Decision maker Accuracy in State R



statistically significant at 5% level in the first 20 periods (one-sided Mann-Whitney: p-value $= 0.027$) but insignificant in the second (one-sided Mann-Whitney: p-value $= 0.631$).[36] We observe a similar trend in state R. We will discuss the reasons for this trend in adviser behavior shortly.

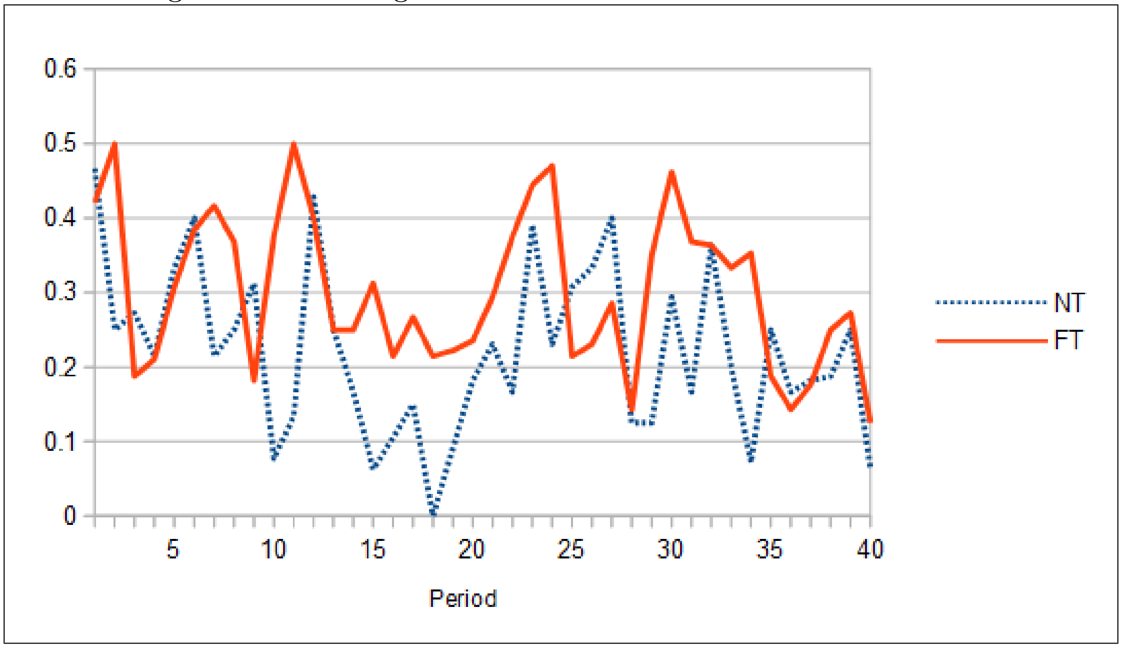
Table 3: Percentage of Advisor Behavior Consistent with Honest Equilibrium

	Transparency	**Non-transparency**	**Mann-Whitney p–value**
State L			
Overall	0.303	0.223	0.075
Before role-switch	0.313	0.217	0.027
After role-switch	0.294	0.230	0.631
State R			
Overall	0.151	0.026	0.036
Before role-switch	0.195	0.028	0.017
After role-switch	0.104	0.026	0.258

Note: One-sided Mann-Whitney tests using matching-group averages as observations (n=12).

[36] A stochastic inequality test finds that subjects are more likely to recommend l in state L in the transparency treatment, but only when considering behaviour before role-switching ($p = 0.078$).

Figure 3: Percentage of Advisors who recommend l in state L



On average, 2.6% of advisers choose r_R in state R without transparency. This increases to 15.1% with transparency. The difference across the two treatments is significant according to a Mann Whitney test (one-sided p-value $= 0.036$). As in state L, the effect of transparency is stronger in the first half of the experiment; Table 3 shows that on average, 19.5% of advisers reject the payment and recommend r in state R with transparency whereas only 2.8% do so without transparency (one-sided p-value $= 0.017$).[37]

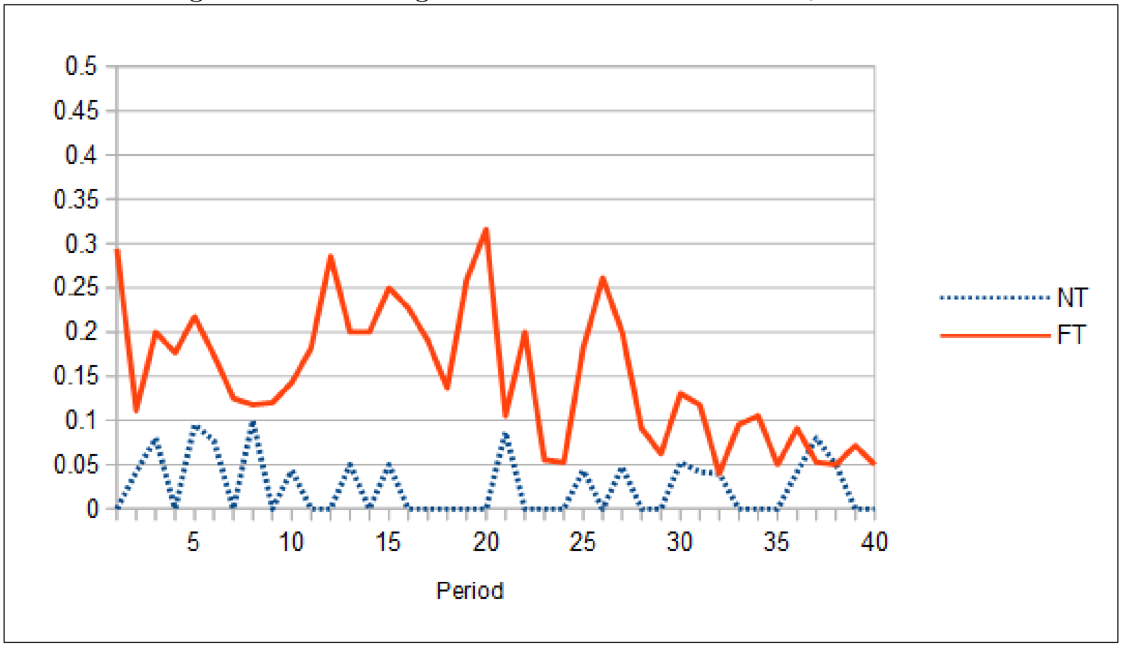
From these findings, we conclude that transparency increases the percentage of advisers who reject the payment and recommend the correct policy in both states—in line with Hypothesis 2. Thus, our answer to Question 2 is also affirmative: Transparency does affect advisers' behavior and makes them more likely to follow the honest equilibrium strategy.

One caveat is that the effects of transparency on adviser behavior weaken over time, as already mentioned. Figures 3 and 4 display the evolution of adviser behavior consistent with the honest equilibrium in state L and state R, respectively. Both figures reveal a downward trend in the transparency treatment; the trend is especially prominent in state R.

What are the plausible explanations for this observation? As we will discuss in the next section in more detail, a substantial fraction of decision makers choose to follow the recommendation of the adviser even if they know that the adviser accepted the payment from

[37] A stochastic inequality test finds that subjects are more likely to reject the payment and recommend r in state R in the transparency treatment, but only when considering behaviour before role-switching ($p = 0.052$).

Figure 4: Percentage of Advisors who choose r_R in state R



the SIG. This behavior is likely to reduce over time the proportion of advisers who behave in accordance with the honest equilibrium. Indeed, our regression analysis in the next section shows that advisers' behavior is shaped by their previous experience with decision makers and that advisers who faced no or limited bias for accepting the side-payment are more likely to accept the payment than those who have been punished for it.

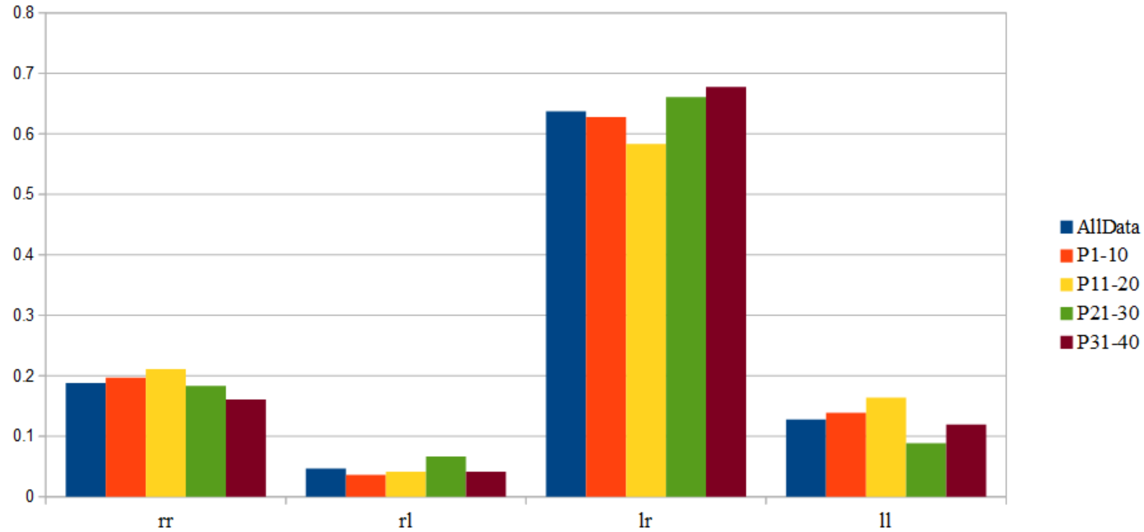
We also note that our one-time role switching may have had a negative influence on the decision makers of the first half of the experiment, who became advisers in the second half. For instance, decision makers in the first half of the transparency treatment were recommended the correct policy less than 1/3 of the time in state L, as Table 3 shows. This type of behavior may have resulted in a resentment among decision makers and facilitated the pervasiveness of the corrupt equilibrium behavior in the second half when the decision makers of the first half became advisers to their previous advisers.

6.3 Decision maker Behavior

As explained in Section 5, we used the strategy method with decision makers in order to elicit their full strategy. This method enables us to have more observations from decision makers at every information set and analyze better whether decision makers use equilibrium strategies.

We start with the non-transparency treatment. The corrupt equilibrium is the unique

Figure 5: Decision maker Strategies in the Nontransparency Treatment



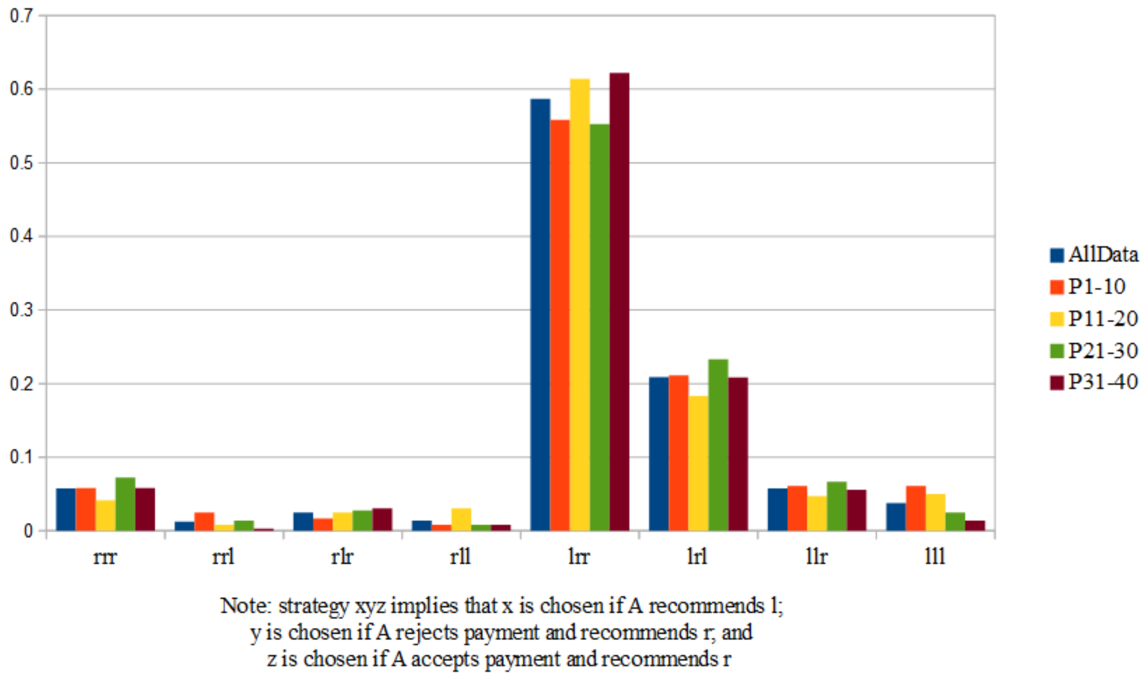
Note: strategy xy implies that x is chosen if A recommends l; and y is chosen if A recommends r

equilibrium in this treatment. To be more precise, the equilibrium is unique on the equilibrium path but there are different off-the-equilibrium path beliefs that support the same equilibrium, none of which can be eliminated using IC or D1. Thus, out of four possible pure strategies, two are consistent with the corrupt equilibrium: (i) always choose r; and (ii) choose r if $m = r$ and l if $m = l$. The strategy "choose r if $m = r$ and l if $m = l$" is by far the most popular strategy—63.8% of the decision makers use it. Figure 5 shows the distribution of decision maker strategies in the transparency treatment. In total, 82.6% of the decision makers use strategies that are consistent with the corrupt equilibrium in the non-transparency treatment.

We next find the best response of the decision maker in the non-transparency treatment given the empirical distribution of the adviser recommendations. This is equivalent to finding the empirical posterior probability of state R given that (i) the adviser recommends r; and (ii) the adviser recommends l. We find that the empirical best response of the decision maker is "choose r if $m = r$ and l if $m = l$," this is indeed the most frequently observed strategy in the data.

Figure 6 shows the distribution of decision maker strategies in the transparency treatment. Of the eight possible decision maker strategies, "choose r if $m \in \{r_A, r_R\}$ and l if $m = l$" is the most popular one—about 59% of decision makers use this strategy. This strategy is consistent with the corrupt equilibrium. More generally, there are four decision maker strategies that are consistent with the corrupt equilibrium in the transparency treatment:

Figure 6: Decision maker Strategies in the Transparency Treatment



(i) always r; (ii) choose r if $m \in \{r_A, r_R\}$ and l if $m = l$; (iii) choose r if $m \in \{r_A, l\}$ and l if $m = r_R$; and (iv) choose r if $m = r_A$ and l if $m \in \{l, r_R\}$.[38] In total, 72.7% of the decision makers use strategies that are consistent with the corrupt equilibrium.

There is only one strategy that is consistent with the honest equilibrium: choose r if $m = r_R$ and l if $m \in \{r_A, l\}$. As we discussed before, the honest equilibrium requires a bias against those advisers who accept the payment. Indeed, we observe that a non-negligible fraction of decision makers are suspicious of advisers who accepted the side payment in the transparency treatments. 21% of decision makers use the honest equilibrium strategy; it is the second most popular strategy after the corrupt equilibrium strategy "choose r if $m \in \{r_A, r_R\}$ and l if $m = l$."

We next find which strategy is the optimal decision maker strategy in the transparency treatment given the empirical distribution of adviser behavior. To that aim, we compute the empirical posterior probability of state R given that (i) A accepts the payment; (ii) A rejects the payment and recommends r; and (iii) A recommends l. It turns out that the optimal decision maker strategy is the corrupt equilibrium strategy, "choose r if $m \in \{r_A, r_R\}$ and l if $m = l$."

[38] The strategies except for the second one may seem unreasonable; however, there is no standard refinement that could rule out such strategies.

Table 4: Random Effects Probit Estimations of Side Payment Acceptance

	Coefficients	
Constant and Independent Variables	**Model 1**	**Model 2**
FA	0.109 (0.0114)***	
NFA	-0.0708 (0.0372)*	
FR	-0.0971 (0.0307)***	
$Accept_{i,t-1}$	0.156 (0.121)	0.0481 (0.264)
$[P(F\|A)]$		1.195 (0.386)***
$[P(F\|R)]$		-0.735 (0.298)**
t	0.00837 (0.00879)	0.0140 (0.00912)
$Constant$	0.384 (0.151)**	0.260 (0.584)

Notes: (1) The dependent variable is the adviser's binary choice between accepting the payment (= 1) and rejecting (= 0). (2) The independent variables in Models 1 and 2 are defined in the main text. (3) Robust standard errors are given in parentheses. (4) * (**; ***) indicates significance at the 10% (5%; 1%) level. (5) Errors are clustered at the session level.

This finding might seemingly imply that it is suboptimal for decision makers to use the honest equilibrium strategy but there is a caveat. The bias against advisers who accept the payment and the awareness of the advisers with respect to such a bias are likely the reasons why decision makers are better off with transparency. Arguably, the behavior of advisers is endogenous and shaped by the strategy of decision makers over the course of the experiment. Put differently, had the decision makers "always followed" the advisers, then we might have obtained a very different empirical distribution of adviser choices and perhaps transparency would not have benefited decision makers.

We now follow up on this line of reasoning and do a regression analysis in order to see whether (i) advisers who were previously (not) followed by the decision maker *due to* rejecting (accepting) the side payment are less likely to accept the payment; and (ii) advisers who were previously (not) followed by the decision maker *despite* accepting (rejecting) the side payment are more likely to accept the payment. We run a random effects panel probit regression of the probability A accepts the payment in the transparency treatment at time t (denoted by $Accept_t$) as a function of (i) the number of times A accepted the payment and was followed until t (denoted by FA); (ii) the number of times A accepted the payment and was *not* followed until t (denoted by NFA); (iii) the number of times A rejected the payment and was followed until t (denoted by FR); (iv) the lagged dependent variable; and (v) the round number (t). Thus, the panel model is given by

$$Accept_{i,t} = \beta_0 + \beta_1 FA_{i,t-1} + \beta_2 NFA_{i,t-1} + \beta_3 FR_{i,t-1} + \beta_4 Accept_{i,t-1} + \beta_5 t + \varepsilon_{i,t} + \gamma_i > 0.$$ [39]

We predict that $\beta_1 > 0$, $\beta_2 < 0$, and that $\beta_3 < 0$. The results of this regression are given in Table 4 under the column titled Model 1. The coefficient signs are as we predicted and the coefficients are significant. Model 2 is a variation on the same theme and involves the explanatory variables (i) the fraction of times the adviser was followed conditional on accepting the payment until t (denoted by $[P(F|A)]$); (ii) the fraction of times he was followed conditional on rejecting the payment until t (denoted by $Pr(F|R)$); (iii) the lagged dependent variable; and (iv) the round number (t). Thus, the panel model is given by

$$Accept_{i,t} = \beta_0 + \beta_1 \left[P(F|A)\right]_{i,t-1} + \beta_2 \left[P(F|R)\right]_{i,t-1} + \beta_3 Accept_{i,t-1} + \beta_4 t + \varepsilon_{i,t} + \gamma_i.$$

We predict that $\beta_1 > 0$ and $\beta_2 < 0$. The results of this regression are given in Table 4 under the column titled Model 2. The coefficient signs are as we predicted and the coefficients are significant.

These regression results support our notion that adviser behavior is endogenous and is shaped by the behavior of decision makers over the course of the experiment. They also suggest that the empirical percentage of decision makers who are biased against advisers that accept the third-party payment is not sufficiently high, and therefore, the percentage of adviser behavior consistent with the honest equilibrium declines over time. As discussed above, our role switching is likely to have contributed to the time trend, as well. Since decision makers in the first half of the transparency treatment were recommended the correct policy less than 1/3 of the time in state L, they may have been primed to accept the side payment frequently when they became advisers in the second half.

7 Voluntary-transparency Condition

In this section, we provide a discussion of the theoretical and experimental results in the voluntary-transparency condition.

7.1 Theory

Equilibria exist both with and without disclosure in the voluntary-transparency condition.[40] Importantly, the main results of interest are identical to the transparency condition: The

[39] To be more precise, the right hand side specifies the underlying latent propensity that $Accept_{i,t} = 1$.

[40] Due to the substantial increase in complexity of the voluntary transparency condition, and the fact that the important intuition regarding equilibrium behaviour is apparent from the analysis of the transparency

corrupt equilibrium exists as well as an honest equilibrium provided that $\alpha \geq \beta(R)$, as in the transparency condition. In fact, there are various corrupt equilibria and various honest equilibria because equilibrium multiplicity in this condition is more severe than in the transparency condition. However, the main features of a corrupt equilibrium and an honest equilibrium are exactly the same as before. On the equilibrium path of every corrupt equilibrium, A accepts the payment in both states and is followed by D. On the equilibrium path of every honest equilibrium, A rejects the payment in both states, recommends the better policy for D and is followed.[41]

7.2 Results

To distinguish between the three treatments, we denote the transparency treatment by T, the voluntary-transparency treatment by VT and the non-transparency treatment by NT. We will make two types of comparisons. First of all, we will compare average outcomes and adviser behaviour in the three treatments; this is important to evaluate the welfare implications of the different regimes. Secondly, to get a deeper understanding of the mechanisms underlying these results, we will disaggregate the data in VT, and compare the outcome and behaviour when A chose transparency with the data from T. Similarly, we will compare the outcome and behaviour in VT when A chose non-transparency with the data from NT.

We start by analyzing the accuracy of decisions in VT and compare it with the other treatments. Table 5 extends Table 2 by adding the data from the VT. Accuracy in state L is 39.6% in VT; this is 9.9% higher than the accuracy in NT, and 6.6% lower than that in T. The difference across VT and NT is statistically significant according to a one-sided Mann-Whitney test but only at the 10% level, and not significant for the latter.[42] In state R, the accuracy of decisions in VT is 75.7%, which is not statistically different from either of the other two treatments.

We now describe adviser behaviour in VT. We start with the decision regarding disclosure. A majority of advisers chose to disclose their decision regarding the third-party payment: 68.6% chose to disclose in state L and 81.4% chose to do so in state R. There

condition, we omit a formal exposition of the voluntary-transparency condition. The equilibrium analysis and theoretical results are available upon request.

[41] Corrupt and honest equilibria differ in the way advisers choose (not) to disclose their decision. For example, A may choose not to reveal his/her decision regarding the side payment in state L but this is still part of an honest equilibrium if A rejects the payment and recommends l in state L. This is theoretically equivalent to revealing that A rejected the side payment because A can recommend l only if it is rejected.

[42] This difference is not significant according to a stochastic inequality test. To keep the text in this section concise, we report p-values for the remaining stochastic inequality tests in Tables 8-10 in the Appendix.

is a small but statistically significant trend towards greater transparency.[43] The prevalence of advisers' preference for transparency as well as the upward time trend in this preference seem to stem from the fact that decision makers have a bias against advisers who choose non-transparency—we will revisit this point shortly. As a result, many advisers choose transparency.

Table 5: Percentage of Correct Decisions by State and Treatment

	Transparency	**Non-transparency**	**Voluntary Transparency**
State L	0.461	0.297	0.396
State R	0.756	0.793	0.757

We now analyze the percentage of adviser behaviour consistent with the honest equilibrium in VT and compare it with the other treatments. Table 6 combines the overall adviser behavior data from Table 3 with the data from VT. In VT, the overall percentage of adviser behaviour consistent with the honest equilibrium—i.e. recommending l in state L, and rejecting the side-payment and recommending r in state R—is 22% in state L (compared with 30.3% in T and 22.3% in NT), and 15% in state R (compared with 15.1% in T and 2.6% in NT). The difference between VT and T is marginally significant in state L according to a two-sided Mann-Whitney test ($p = 0.078$); there is no difference in state R.[44] Conversely, there is significantly more honest equilibrium behaviour in VT than in NT if the state is R (one-sided Mann-Whitney: p-value $= 0.012$), but there is no difference in state L.

Table 6: Percentage of Advisor Behavior Consistent with Honest Equilibrium

	Transparency	**Non-transparency**	**Voluntary Transparency**
State L	0.303	0.223	0.22
State R	0.151	0.026	0.15

In Table 8 in the Appendix, we disaggregate the data in VT to its transparency and non-transparency components given the advisers' choice. In the table, we refer to the data from VT as *endogenous transparency* if A chooses to reveal his decision regarding the payment and T as *exogenous transparency*. In a similar vein, we refer to the data from VT as

[43] The sign and significance of time trend is based on random effects panel probit regressions (with standard errors clustered by matching group, as usual). The results of the regressions are available on request.

[44] We use a two-sided test since we do not have directional predictions regarding the comparison of transparency and voluntary-transparency treatments. However, we use one-sided tests in the comparison of non-transparency and voluntary-transparency treatments since we have directional predictions.

endogenous non-transparency if A chooses not to reveal his decision regarding the payment and NT as *exogenous non-transparency.* Table 8 shows that with transparency, there is no statistically significant difference in either state regardless of whether transparency is endogenous or exogenous. This is not the case with non-transparency, however; accuracy is higher in state L and lower in state R with endogenous non-transparency than with exogenous non-transparency. Both of these differences are statistically significant at the 1% level (see table 8). This outcome can be explained by the way advisers and decision makers behave in VT, which we now explore.

As can be seen from Table 9 in the Appendix, the proportion of subjects recommending l in state L falls from 22% with exogenous non-transparency to 10% with endogenous non-transparency (Mann-Whitney: $p = 0.045$); there is no statistical difference in state R. Put differently, there is much less honesty among advisers in state L when non-transparency is chosen than when it is imposed. In response, decision makers have a bias against advisers who choose non-transparency (as discussed above) and are much less likely to follow a recommendation of r when non-transparency is chosen than when it is imposed: The percentage of subjects following both recommendations falls from 64% to 42% (MW: $p = 0.004$) while the percentage who choose l, regardless of the recommendation rises from 13% to 34% (MW: $p = 0.004$). As a result of this increased tendency of decision makers to choose l when non-transparency is chosen (instead of being exogenous), accuracy increases in state L and decreases in state R.

Table 9 in the Appendix shows that adviser behaviour is very similar with endogenous and exogenous transparency. In accordance with the adviser behaviour, the distribution of decision makers' strategies are also very similar, regardless of whether transparency is exogenous or endogenous (see Table 10 in the Appendix). These two observations also explain why accuracy does not differ across endogenous and exogenous transparency.

Summing up, the voluntary-transparency treatment does not appear to be as effective in improving information transmission and decision making as the transparency treatment due to the following. Advisers who *choose* transparency are neither more nor less honest than those for whom transparency is mandatory. Advisers who choose non-transparency behave less honestly in state L than those for whom non-transparency is exogenous—however, this is compensated for by decision makers who correctly recognize that this choice signals dishonesty. The net effect is that accuracy with voluntary transparency is roughly an average of the accuracy in T and NT, weighted by the proportion of advisers who choose the respective transparency regime; since a large number of advisers choose non-transparency, accuracy in VT suffers relative to an environment in which transparency is mandatory.

8 Conclusion

Our study sheds light on the effects of transparency on strategic information transmission and decision making in a setting where the advisor might be swayed by a third-party who favors the implementation of a specific policy. Our theoretical results show that transparency is ex-ante never harmful and may strictly improve the accuracy of decision making. Experimental results show that transparency makes decision makers better off in the state in which the adviser has an incentive to lie, and has no effect in the other state. With transparency, more advisers reject the side-payment and recommend the correct policy in both states.

There are two important qualifications. First of all, we find that the form of transparency may matter: While transparency clearly improves decision making (relative to non-transparency) the evidence regarding the effect of voluntary transparency is weak. Secondly, positive effects of transparency weaken over time because many decision makers follow advice regardless of whether or not the payment was accepted, and advisers take advantage of this.

Our experiment was neutrally framed and involved minimal communication. In particular, advisers never claimed one policy to be "better" for the decision maker than the other, and they never made a conflict-of-interest statement in the transparency treatments. Such a statement could make the conflicting relationship between the financial interest of the adviser and making a truthful recommendation more salient to the decision maker. The economic literature has increasingly recognized the role of context for decision making (see, for example, Levitt and List (2007) and the references therein). We believe that framing may be particularly relevant in settings where decision makers are consistently naïve or under-suspicious. Another important direction of future research is to endogenize the third party in both the theory and experiment.

Our analysis has relevant implications for the debate on transparency. Our paper documents positive effects of transparency whereas prior experimental research produced bleak results. Prior research modeled conflict of interest as being exogenous. Our results imply that the debate on transparency should take into account whether the conflict of interest between the adviser and the decision maker should be modeled as being exogenous or endogenous. We believe that the latter is the appropriate approach as experts have in most cases the agency to accept or reject side-payments, gifts and bribes.

References

Banks, Jeffrey S., and Joel Sobel. (1987). "Equilibrium Selection in Signaling Games," *Econometrica*, 55(3), 647-661.

Cain, Daylian M., Loewenstein, George, and Don A. Moore. (2005). "The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest," *Journal of Legal Studies*, 34(1), 1–25.

Cain, Daylian M., Loewenstein, George, and Don A. Moore. (2011). "When Sunlight Fails to Disinfect: Understanding the Perverse Effects of Disclosing Conflicts of Interest," *Journal of Consumer Research*, 37(5), 836-857.

Cho, In-Koo, and David M. Kreps (1987). "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102(2), 179-222.

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

Delafontaine, Ramses. (2015). *Historians as expert judicial witnesses in Tobacco litigation: a controversial legal practice*. Springer.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. (2010). "Disclosure by Politicians," *American Economic Journal: Applied Economics*, 2(2), 179-209.

Durbin, Erik, and Ganesh Iyer. (2009). "Corruptible Advice," *American Economic Journal: Microeconomics*, 1(2), 220-242.

Fanelli, D. (2009). "How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data," *PloS one* 4(5), e5738.

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

Gibson, Rajna, Carmen Tanner, and Alexander F. Wagner. (2013). "Preferences for Truthfulness: Heterogeneity among and within Individuals: Dataset," *American Economic Review*, 103(1), 532–548.

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

Hurkens, Sjaak, and Navin Kartik. (2009). "Would I Lie to You? On Social Preferences and Lying Aversion," *Experimental Economics*, 12(2), 180–92.

Inderst, Roman, and Marco Ottaviani. (2009). "Misselling through Agents," *American Economic Review*, 99(3), 883-908.

Inderst, Roman, and Marco Ottaviani. (2012). "Competition through commissions and kickbacks," *American Economic Review*, 102(2), 780-809.

Ismayilov, Huseyn, and Jan Potters. (2013). "Disclosing adviser's interests neither hurts nor helps," *Journal of Economic Behavior and Organization*, 93, 314-320.

Levitt, Steven D., and John A. List. (2007). "What do laboratory experiments measuring social preferences reveal about the real world?" *Journal of Economic Perspectives,* 21(2), 153-174.

Loewenstein, George, Cain, Daylian M., and Sah, Sunita. (2011). "The limits of transparency: Pitfalls and potential of disclosing conflicts of interest," *American Economic Review*, 101(3), 423-428.

Loewenstein, George, Sunita Sah, and Daylian M. Cain. (2012) "The unintended consequences of conflict of interest disclosure," *Journal of the American Medical Association*, 307(7), 669-670.

Koch, Christopher, and Carsten Schmidt. (2010). "Disclosing conflicts of interest–Do experience and reputation matter?," *Accounting, Organizations and Society*, 35(1), 95-107.

Potters, Jan, and Frans Van Winden. (1992). "Lobbying and asymmetric information," *Public Choice*, 74(3), 269-292.

Potters, Jan, and Frans Van Winden. (2000). "Professionals and students in a lobbying experiment: Professional rules of conduct and subject surrogacy," *Journal of Economic Behavior & Organization*, 43(4), 499-522.

Proctor, Robert. (2011). *Golden holocaust: origins of the cigarette catastrophe and the case for abolition.* University of California Press.

Rode, Julian. (2010). "Truth and trust in communication: experiments on the effect of a competitive context," *Games and Economic Behavior*, 68(1), 325–338

Sah, Sunita, and George Loewenstein. (2013). "Nothing to Declare: Mandatory and Voluntary Disclosure Leads Advisers to Avoid Conflicts of Interest," *Psychological Science*, 25(2), 575-584.

Sánchez-Pagés, Santiago, and Marc Vorsatz. (2007). "An Experimental Study of Truth-Telling in a Sender-Receiver Game," *Games and Economic Behavior*, 61(1), 86–112.

Schlag, Karl. (2008). "A New Method for Constructing Exact Tests without Making any Assumptions," *Department of Economics and Business Working Paper 1109, Universitat Pompeu Fabra.*

Sobel, Joel. (1985). "A Theory of Credibility," *Review of Economic Studies*, 52(4), 557-573.

Stossel, Thomas P. (2005). "Regulating Academic–Industrial Research Relationships — Solving Problems or Stifling Progress?" *The New England Journal of Medicine*, 353(10), 1060-1065.

Stossel, Thomas P., and Lance K. Stell. (2011). "Time to 'walk the walk' about Industry Ties to enhance Health," *Nature Medicine*, 17, 437–438.

Weber, Michael A. (2009). "Academic Physicians Confront a Hostile World: The Creation of ACRE," *The Journal of Clinical Hypertension* 11(10), 533–536.

Yale Project on Climate Change Communication. 2013. "Scientific and Public Perspectives on Climate Change." http://environment.yale.edu/climate-communication/files/ClimateNote_Consensus_Gap_May2013_FINAL6.pdf

9 Appendix

Table 7: Random Effects Probit Estimations of Accuracy in the Transparency Treatment

Coefficients	**State L**	**State R**
Period # (complete data)	-0.0098	0.0101
	(0.0074)	(0.0069)
Period # (data in the first half)	-0.0222	0.0003
	(0.0223)	(0.0105)
Period # (data in the second half)	-0.0309**	0.0214
	(0.0140)	(0.0195)

Notes: (1) Standard errors are clustered by matching group. (2) * (**; ***) indicates significance at the 10% (5%; 1%) level.

10 Online Appendix (For Online Publication)

10.1 Proofs

Proof of Proposition 1. To see why only a corrupt equilibrium exists, first note that at least one message chosen by A in equilibrium results in D choosing r with probability one; i.e., there exists $m \in \{l, r\}$ where $\Pr_A(m|S) > 0$ for some $S \in \{L, R\}$ such that $Pr_D(r|m) = 1$. Suppose not. Then, $Pr_D(r|m) < 1$ for any $m \in \{l, r\}$ where $\Pr_A(m|S) > 0$. This can be true

Table 8: Percentage of Correct Decisions by State and Treatment

		Exogenous	**Endogenous**	**p–value**
Transparency	**State L**	0.46	0.36	0.423
	State R	0.76	0.81	0.200
Non-transparency	**State L**	0.30	0.47	0.004 (0.031)
	State R	0.79	0.51	0.007 (0.038)

Notes: (1) We use two-sided Mann-Whitney tests with matching-group averages as observations (n=12). (2) Stochastic inequality test p-values are in parentheses.

Table 9: Percentage of Advisor Behavior Consistent with Honest Equilibrium

		Exogenous	**Endogenous**	**p–value**
Transparency	**State L**	0.30	0.27	0.749
	State R	0.15	0.17	0.873
Non-transparency	**State L**	0.22	0.10	0.045 (0.274)
	State R	0.03	0.08	0.873

Notes: (1) We use two-sided Mann-Whitney tests with matching-group averages as observations (n=12). (2) Stochastic inequality test p-values are in parentheses.

Table 10: Decision Maker Strategies

	Exogenous	**Endogenous**	**p-value**
rrr	0.06	0.15	0.055 (0.156)
rrl	0.01	0.01	0.470
rlr	0.03	0.05	0.629
rll	0.01	0.01	0.336
lrr (CE)	0.59	0.49	0.109
lrl (HE)	0.21	0.15	0.522
llr	0.06	0.10	0.173
lll	0.04	0.04	0.423
rr	0.19	0.14	0.337
rl	0.05	0.1	0.109
lr (CE)	0.64	0.42	0.004 (0.031)
ll	0.13	0.34	0.004 (0.031)

Notes: (1) We use two-sided Mann-Whitney tests with matching-group averages as observations (n=12). (2) Stochastic inequality test p-values are in parentheses.

only if l and r are both chosen in equilibrium with positive probability and $\mu(R|m) \leq 0.5$ for $m \in \{l, r\}$.[45] But $\mu(R|r) \leq 0.5$ and $\mu(R|l) \leq 0.5$ together result in an inconsistency. Since

$$\mu(R|m) = \frac{\Pr_A(m|R)(1-p)}{\Pr_A(m|R)(1-p) + \Pr_A(m|L)p} \leq 0.5,$$

it follows that

$$\Pr_A(m|L) \geq \frac{1-p}{p} \Pr_A(m|R)$$

for $m \in \{r, l\}$. But, $\Pr_A(r|S) + \Pr_A(l|S) = 1$ for $S \in \{L, R\}$. Then,

$$\Pr_A(r|L) + \Pr_A(l|L) = 1 \geq \frac{1-p}{p}(\Pr_A(r|R) + \Pr_A(l|R)) > 1,$$

since $p < 0.5$, a contradiction. Hence, there exists $m \in \{l, r\}$ such that $Pr_A(m|S) > 0$ for some $S \in \{L, R\}$, and $Pr_D(r|m) = 1$. But this message can only be r. Suppose towards a contradiction that it is l; i.e., $Pr_A(l|S) > 0$ for some $S \in \{L, R\}$ and $\Pr_D(r|l) = 1$. The latter requires $\mu(R|l) \geq 0.5$. But if the state is R, then A strictly prefers choosing r_A (thus, $m = r$) rather than l, securing a minimum payoff of $\beta(R) + \gamma$ versus 0. Therefore, $\Pr_A(l|R) = 0$ which contradicts $\mu(R|l) \geq 0.5$. Thus, $\Pr_D(r|r) = 1$. Given that $\Pr_D(r|r) = 1$ and $\beta(L) > \gamma$, D always chooses r_A in equilibrium. Hence, we obtain the desired result.[46]

Proposition 2. *See the statement in the main text for the case where $\alpha + \gamma > \beta(L)$. Assume that $\alpha + \gamma < \beta(L)$. There are (at most) three equilibria that satisfy the IC:*

(1) *Corrupt Equilibrium:* $\Pr_A(r_A|L) = \Pr_A(r_A|R) = 1$; $\Pr_D(r|r_A) = 1$; $\mu(R|r_A) = 1 - p$.

(2) *Type-II Honest Equilibrium (if $\alpha \geq \beta(R)$):* $\Pr_A(r_A|L) = 1$; $\Pr_A(r_R|R) = 1$; $\Pr_D(r|r_R) = 1$; $\Pr_D(l|r_A) = 1$; $\mu(R|r_R) = 1$ and $\mu(L|r_A) = 0$.

(3) *Mixed-strategy Equilibrium (if $\alpha \geq \beta(R)$):* $\Pr_A(r_A|L) = 1$; $\Pr_A(r_A|R) = \frac{p}{1-p}$ and $\Pr_A(r_R|R) = \frac{1-2p}{1-p}$; $\Pr_D(r|r_R) = 1$; $\mu(R|r_R) = 1$; $\Pr_D(r|r_A) = 1 - \beta(R)/\alpha$; $\mu(R|r_A) = 0.5$.

Next, assume that $\alpha + \gamma = \beta(L)$. There always exists a corrupt equilibrium; if, in addition, $\alpha \geq \beta(R)$, then there exists an honest equilibrium, a type-II honest equilibrium and a mixed-strategy equilibrium (as in (3) above); if, in addition, $\alpha = \beta(R)$, then there also exists a continuum of completely mixed equilibria in which A randomizes between r_A and l in state L and between r_A and r_R in state R.

[45] Otherwise, $\Pr_A(m|L) = \Pr_A(m|R) = 1$ for either $m = l$ or $m = r$. This, combined with the fact that $p < 0.5$, implies that D must choose r with probability one, a contradiction.

[46] Off-the-equilibrium path belief $\mu(R|l)$ is not uniquely determined; therefore, $\Pr_D(r|l) = 1$ if $\mu(R|l) > 0.5$, $\Pr_D(r|l) \in [0, 1]$ if $\mu(R|l) = 0.5$ and $\Pr_D(r|l) = 0$ if $\mu(R|l) < 0.5$.

The mixed-strategy equilibrium in (3) generates the same ex-ante expected accuracy as the corrupt equilibrium, and the completely mixed equilibria (which exist if $\alpha + \gamma = \beta(L)$ and $\alpha = \beta(R)$) generate (weakly) higher expected accuracy than the corrupt equilibrium. Honest and Type-II Honest equilibria always generate 100% accuracy. Thus, decision making under transparency (strictly) dominates that under non-transparency (if $\alpha \geq \beta(R)$).

Proof. The proof uses a number of lemmas.

Lemma 3 *For any $m \in \{l, r_A, r_R\}$ chosen by the adviser in equilibrium,*

$$\Pr_D(l|m) = \begin{cases} 1 & \text{if } \Pr_A(m|L) > \frac{1-p}{p} \Pr_A(m|R), \\ \in [0,1] & \text{if } \Pr_A(m|L) = \frac{1-p}{p} \Pr_A(m|R), \\ 0 & \text{if } \Pr_A(m|L) < \frac{1-p}{p} \Pr_A(m|R), \end{cases}$$

Proof. The decision maker strictly prefers choosing l if

$$\mu(L|m)(\bar{\pi} - \underline{\pi}) > (1 - \mu(L|m))(\bar{\pi} - \underline{\pi})$$

for $m \in \{l, r_A, r_R\}$. That is, l is strictly preferred if the adviser chooses m and $\mu(L|m) > 0.5$. In equilibrium, $\mu(L|m)$ must accord with Bayes' law; i.e.

$$\mu(L|m) = \frac{p \Pr_A(m|L)}{p \Pr_A(m|L) + (1-p) \Pr_A(m|R)}.$$

This implies that the decision maker chooses l with probability one if

$$\Pr_A(m|L) > \frac{1-p}{p} \Pr_A(m|R).$$

The rest of the proof follows similar lines and is therefore omitted.

Lemma 4 *At least one action chosen by the adviser in equilibrium is followed by r with probability one, i.e., there exists an $m \in \{l, r_A, r_R\}$ with $\Pr_A(m|S) > 0$ for some $S \in \{L, R\}$ in equilibrium such that $\Pr_D(r|m) = 1$.*

Proof. Suppose that $\Pr_D(r|m) < 1$ for every m with $\Pr_A(m|S) > 0$, $S \in \{L, R\}$ in equilibrium. Then by Lemma 3, $\Pr_A(m|L) \geq \frac{1-p}{p} \Pr_A(m|R)$ $\forall m \in \{l, r_A, r_R\}$ (if m is not played in equilibrium then the inequality is satisfied trivially). This implies that

$$
\begin{aligned}
1 &= \Pr_A(l|L) + \Pr_A(r_A|L) + \Pr_A(r_R|L) \\
&\geq \tfrac{1-p}{p}(\Pr_A(l|R) + \Pr_A(r_A|R) + \Pr_A(r_R|R)) \\
&= \tfrac{1-p}{p} \\
&> 1,
\end{aligned}
$$

as $p < 1/2$, a contradiction.

Lemma 5 *In equilibrium, at least one of* r_A *and* r_R *is played, then followed by* r *with probability one.*

Proof. By Lemma 4, there exists an $m \in \{l, r_A, r_R\}$ with $\Pr_A(m|S) > 0$ for some $S \in \{L, R\}$ in equilibrium such that $\Pr_D(r|m) = 1$. Suppose towards a contradiction that it is l in some equilibrium; i.e., $\Pr_A(l|S) > 0$ for some $S \in \{L, R\}$ and $\Pr_D(r|l) = 1$. The adviser's payoff from choosing l in state R is zero, however a payoff of at least $\beta(R) + \gamma$ can be guaranteed by choosing r_A, so $\Pr_A(l|R) = 0$. But if l is being chosen only in state L, then $\mu(L|l) = 1$ and $\Pr_D(r|l) = 0$, a contradiction.

Lemma 6 *In equilibrium, the adviser never chooses* l *in* R. *If the adviser chooses* l *in equilibrium, then it is always followed by* l.

Proof. By the previous lemma, in state R, the adviser will be able to earn either $\alpha + \gamma$ by deviating to r_R, or $\alpha + \beta(R) + \gamma$ by deviating to r_A, both of which are greater than the highest possible payoff from l (α). Therefore if the adviser chooses l in equilibrium, he chooses it only in state L, so that $\mu(L|l) = 1$, and $\Pr_D(l|l) = 1$.

Lemma 7 *Suppose that* $\alpha + \gamma > \beta(L)$. *If* $\Pr_A(r_A|S) > 0$ *for some* $S \in \{L, R\}$ *in an equilibrium robust to IC, then (i)* $\Pr_A(r_A|L) = 1$ *and* $\Pr_A(r_A|R) > 0$ *must hold if* $\beta(R) + \gamma \neq \beta(L)$ *(i.e., there exists a corrupt equilibrium and a mixed-strategy equilibrium); (ii)* $\Pr_A(r_A|L) = \Pr_A(r_A|R) = 1$ *must hold if* $\beta(R) + \gamma > \beta(L)$ *(i.e., only the corrupt equilibrium exists); (iii)* $\Pr_A(r_A|R) > 0$ *and* $\Pr_A(r_A|L) > 0$ *if* $\beta(R) + \gamma = \beta(L)$ *(i.e., in addition to the equilibria spelled out in (i) and (ii), there exists a continuum of mixed-strategy equilibria in which* A *randomizes between* r_A *and* l *in state* L *and between* r_A *and* r_R *in state* R*).*

Proof. Suppose that r_A is chosen in equilibrium; i.e., $\Pr_A(r_A|S) > 0$ for some $S \in \{L, R\}$.

(1) First, we show that this implies that $\Pr_A(r_A|L) > 0$ must hold. Assume towards a contradiction that r_A is chosen in equilibrium but only in state R; i.e., $\Pr_A(r_A|R) > 0$ and $\Pr_A(r_A|L) = 0$. Using Bayes' law, $\mu(R|r_A) = 1$ in equilibrium and thus, $\Pr_D(r|r_A) = 1$.

But then A deviates to r_A in state L since $\beta(L) > \gamma$, a contradiction. Note that this result actually holds regardless of the assumption that $\alpha + \gamma > \beta(L)$.

(2) Next, we show that if A chooses r_A in equilibrium in some S, then it must be chosen with positive probability in both states. Suppose not. By (1) above, $\Pr_A(r_A|L) > 0$. Then, by hypothesis, $\Pr_A(r_A|R) = 0$ must hold; i.e., A chooses r_A in equilibrium only in state L. But this implies that $\mu(L|r_A) = 1$ using Bayes' law and that $\Pr_D(l|r_A) = 1$. Thus, by Lemma 5, r_R is played in equilibrium and $\Pr_D(r|r_R) = 1$. As a result, the equilibrium payoff in state R is at least $\alpha + \gamma$ and recommending l must be equilibrium dominated in state R. By IC, $\mu(L|l) = 1$ and thus, A can deviate to recommending l in state L and obtain at least $\alpha + \gamma > \beta(L)$, a contradiction. As a result, $\Pr_A(r_A|R) > 0$ must hold if $\Pr_A(r_A|L) > 0$ in equilibrium. This in turn implies that if r_A is chosen in equilibrium, then it must be chosen with positive probability in both states.

(3) We now show that $\Pr_A(r_A|L) = 1$ and $\Pr_A(r_A|R) > 0$ must hold if r_A is chosen in equilibrium and $\beta(R) + \gamma \neq \beta(L)$. Suppose not so that $\Pr_A(r_A|L) \in (0, 1)$ (as implied by (2) above). This means that A randomizes between l and r_A in state L. Note that A cannot randomize between r_R and r_A in state L. Because if A randomizes between r_R and r_A in state L, then by Lemma 5, $\Pr_D(r|r_R) = 1$ in equilibrium—otherwise, A always chooses r_A. It follows that recommending l is equilibrium dominated in R. But then, recommending l is a profitable deviation in state L since the payoff from r_R in state L is only α. Thus, $\Pr_A(r_A|L) \in (0, 1)$ implies that A randomizes between l and r_A in state L. In equilibrium, $\Pr_A(r_R|R) > 0$ must hold, otherwise $\Pr_D(r|r_A) = 1$ by Lemma 5, which is a contradiction. Put differently, A must randomize between r_A and r_R in state R. We next prove that it cannot be the case that A randomizes between r_A and l in state L *and* between r_A and r_R in state R unless $\beta(R) + \gamma = \beta(L)$—we discuss this case below. To see why, note that $\Pr_D(r|r_A)$ is given by

$$\Pr_D(r|r_A)\alpha + \beta(L) = \alpha + \gamma$$

since A is indifferent between r_A and l in state L and by Lemma 6, $\Pr_D(l|l) = 1$. Thus,

$$\Pr_D(r|r_A) = 1 - \frac{\beta(L) - \gamma}{\alpha}$$

in equilibrium. Since $\Pr_D(r|r_A) < 1$, $\Pr_D(r|r_R) = 1$ must hold by Lemma 5. Thus, A is indifferent between r_A and r_R in state R and between l and r_A in state L only if the knife-edge case

$$\beta(R) + \gamma = \beta(L)$$

holds. As a result, if r_A is chosen in equilibrium, then $\Pr_A(r_A|L) = 1$ and $\Pr_A(r_A|R) > 0$ must hold as long as $\beta(R) + \gamma \neq \beta(L)$. This proves part (i). If however $\beta(R) + \gamma = \beta(L)$

holds, then apart from the equilibria in which $\Pr_A(r_A|L) = 1$ and $\Pr_A(r_A|R) > 0$, there also exists a continuum of mixed-strategy equilibria such that A randomizes between r_A and l in state L *and* between r_A and r_R in state R. $\Pr_A(r_A|R)/\Pr_A(r_A|L) = p/(1-p)$ and $\Pr_D(r|r_R) = 1$, $\Pr_D(l|l) = 1$ and $\Pr_D(r|r_A) = 1 - (\beta(L) - \gamma)/\alpha$ hold in such equilibria. This is the case described in part (iii) of Lemma 7.

(4) Part (3) implies that if r_A is played in equilibrium and $\beta(R) + \gamma \neq \beta(L)$, then this is part of the corrupt equilibrium with$\Pr_A(r_A|R) = \Pr_A(r_A|L) = 1$, and it can also be part of a mixed-strategy equilibrium in which $\Pr_A(r_A|R) \in (0,1)$ and $\Pr_A(r_R|R) \in (0,1)$.[47] However, if $\beta(R) + \gamma > \beta(L)$ and r_A is chosen in equilibrium, then $\Pr_A(r_A|L) = \Pr_A(r_A|R) = 1$ must hold; i.e., a mixed-strategy equilibrium in which $\Pr_A(r_A|L) = 1$ and $\Pr_A(r_A|R) \in (0,1)$ is not robust to IC. To see why, first note that if $\Pr_A(r_A|L) = 1$ and $\Pr_A(r_A|R) \in (0,1)$, then we must have $\Pr_A(r_R|R) \in (0,1)$ (see Footnote 44). By Lemma 5, $\Pr_D(r|r_R) = 1$ must hold (otherwise, A chooses r_A with probability one in both states). This means that l is equilibrium dominated in state R, and since $\Pr_D(r|r_A)\alpha + \beta(L) < \alpha + \gamma$ (by the indifference condition in state R and $\beta(R) + \gamma > \beta(L)$) recommending l is a profitable deviation in state L. This proves part (ii). Note that the mixed-strategy equilibrium *is robust to IC* if $\beta(R) + \gamma \leq \beta(L)$: Even if l is equilibrium dominated in state R, recommending l is no longer a profitable deviation in state L given that $\beta(R) + \gamma \leq \beta(L)$.

Finally, we can prove Proposition 2. First, assume that $\alpha + \gamma > \beta(L)$. We prove the result in this case in 4 steps.

(1) Given Lemma 7 and Intuitive Criterion, A must choose in state L either l with probability one or r_A with probability one if $\beta(R) + \gamma \neq \beta(L)$; i.e., either $\Pr_A(l|L) = 1$ or $\Pr_A(r_A|L) = 1$ if $\beta(R) + \gamma \neq \beta(L)$. Suppose not. Then, by Lemma 7, $\Pr_A(r_A|L) = \Pr_A(r_A|R) = 0$ must hold in equilibrium. Moreover, $\Pr_A(r_R|L) > 0$ must hold since $\Pr_A(l|L) < 1$ (by contradictory hypothesis) and $\Pr_A(r_A|L) = 0$. Lemma 5 implies that $\Pr_D(r|r_R) = 1$. Thus, $\Pr_A(r_R|R) = 1$. As a result, $\Pr_A(r_R|L) = 1$ must hold because if $\Pr_A(l|L) > 0$, then $\mu(L|l) = 1$ and A cannot be indifferent between l and r_R in state L. Yet, IC rules out an equilibrium in which $\Pr_A(r_R|L) = \Pr_A(r_R|R) = 1$. A can deviate to l in state L, which is a profitable deviation given the belief $\mu(L|l) = 1$ (because l is equilibrium-dominated in state R). Hence, if $\beta(R) + \gamma \neq \beta(L)$ either $\Pr_A(r_A|L) = 1$ or $\Pr_A(l|L) = 1$ in any equilibrium that satisfies IC.

(2) If $\Pr_A(l|L) = 1$ in an equilibrium, then $\Pr_A(r_R|R) = 1$ must hold—this constitutes the honest equilibrium and it exists provided that $\alpha \geq \beta(R)$ holds. To see why $\Pr_A(l|L) = 1$

[47] To see why $\Pr_A(r_A|L) = 1$ and $\Pr_A(r_A|R) \in (0,1)$ imply that $\Pr_A(r_R|R) \in (0,1)$, note that otherwise $\Pr_D(r|r_A) = 1$ by Lemma 5, a contradiction.

implies that $\Pr_A(r_R|R) = 1$, first note that $\Pr_A(r_A|L) = 0$ implies that $\Pr_A(r_A|R) = 0$ by Lemma 7. Moreover, $\Pr_A(l|R) = 0$ must hold. Since $\Pr_A(r_A|R) = 0$ and $\Pr_A(r_A|L) = 0$, it follows that r_R is played and $\Pr_D(r|r_R) = 1$, by Lemma 5. Given this, choosing l in state R is inferior to choosing r_R. Thus, $\Pr_A(r_R|R) = 1$. The honest equilibrium uses the following out-of-equilibrium belief: If the adviser deviates to r_A, then the decision maker holds the belief that $\mu(L|r_A) \geq 0.5$ and chooses l (with sufficiently high probability). This belief system is not ruled out by the Intuitive Criterion because r_A is not equilibrium dominated in L. Since $\alpha + \gamma > \beta(L)$ by assumption, A does not deviate from l in state L. Also, A does not deviate from r_R in state R provided that $\alpha \geq \beta(R)$.

(3) If $\Pr_A(r_A|L) = 1$ in equilibrium, then either
(i) $\Pr_A(r_A|R) = 1$; or
(ii) $\Pr_A(r_A|R) \in (0, 1)$ and $\Pr_A(r_A|R) + \Pr_A(r_R|R) = 1$—provided that $\beta(R) + \gamma \leq \beta(L)$.
The former is the corrupt equilibrium. Note that IC puts no restriction on the out-of-equilibrium beliefs; $\mu(R|r_R) \in [0, 1]$ and $\mu(R|l) \in [0, 1]$. The case in (ii) is the mixed-strategy equilibrium. In this equilibrium, $\Pr_D(r|r_A)$ is derived from the indifference condition

$$\Pr_D(r|r_A)\alpha + \beta(R) + \gamma = \alpha + \gamma.$$

Thus, $\Pr_D(r|r_A) = 1 - \beta(R)/\alpha$. Since $\Pr_D(r|r_A) \in (0, 1)$, $\mu(R|r_A) = 0.5$ must hold. This, in turn, requires that $\Pr_A(r_A|R) = \frac{p}{1-p}$ and $\Pr_A(r_R|R) = \frac{1-2p}{1-p}$. As explained in part (4) in the proof of Lemma 7, the mixed-strategy equilibrium is robust to IC if $\beta(L) \geq \beta(R) + \gamma$. IC puts no restriction on the out-of-equilibrium belief $\mu(R|l)$—i.e., $\mu(R|l) \in [0, 1]$. It follows that off-the-equilibrium path, $\Pr_D(r|l) = 1$ if $\mu(R|l) > 0.5$, $\Pr_D(r|l) \in [0, 1]$ if $\mu(R|l) = 0.5$ and $\Pr_D(r|l) = 0$ if $\mu(R|l) < 0.5$.

(4) If $\beta(R) + \gamma = \beta(L)$, there exist (in addition to the above equilibria) a continuum of mixed-strategy equilibria in which A randomizes between r_A and l in state L *and* between r_A and r_R in state R—see part (3) in the proof of Lemma 7.

Next, assume that $\alpha + \gamma < \beta(L)$. Note that in this case, A always chooses r_A in state L, which simplifies the analysis. As usual, the corrupt equilibrium always exists. Moreover, if $\alpha < \beta(R)$, then there exists only the corrupt equilibrium because accepting the payment is the dominant strategy in both states. However, if $\alpha \geq \beta(R)$, then there also exists a type-II honest equilibrium in which A chooses r_A in state L and r_R in state R. Thus, the equilibrium is fully informative, and $\Pr_D(r|r_R) = \Pr_D(l|r_A) = 1$. Moreover, if $\alpha \geq \beta(R)$, there is also the mixed strategy equilibrium with $\Pr_A(r_A|R) \in (0, 1)$ and $\Pr_A(r_A|R) + \Pr_A(r_R|R) = 1$. The mixed-strategy equilibrium is exactly the same as in (3) above: $\Pr_D(r|r_A) = 1 - \beta(R)/\alpha$, $\Pr_A(r_A|R) = \frac{p}{1-p}$ and $\Pr_A(r_R|R) = \frac{1-2p}{1-p}$.

Finally, assume that $\alpha + \gamma = \beta(L)$. As usual, the corrupt equilibrium always exists. If $\alpha < \beta(R)$, then there exists only the corrupt equilibrium. Since $\alpha < \beta(R)$, A chooses r_A in state R with probability one. By Lemma 5, $\Pr_D(r|r_A) = 1$ must hold since $\alpha < \beta(R)$ implies that $\Pr_A(r_R|R) = 0$. Given that $\Pr_D(r|r_A) = 1$, A will always choose r_A in both states. If, however, $\alpha > \beta(R)$, there are the honest equilibrium, the type-II honest equilibrium and the mixed strategy equilibrium in which $\Pr_A(r_A|L) = 1$, $\Pr_A(r_A|R) \in (0, 1)$ and $\Pr_A(r_A|R)+\Pr_A(r_R|R) = 1$, in addition to the corrupt equilibrium. Finally, if $\alpha = \beta(R)$, there exists, in addition to the list above, a continuum of mixed-strategy equilibria in which A randomizes between r_A and l in state L *and* between r_A and r_R in state R.

The accuracy comparison across the transparency and non-transparency conditions is straightforward using the equilibrium indifference conditions and mixing probabilities and therefore, it is omitted.

We now prove our claim in the main text that transparency does not make a differency if $\beta(L) \leq \gamma$.[48] Transparency does not make a difference in that case because there is always a fully informative equilibrium in both conditions. First assume that $\beta(L) < \gamma$. In that case, the adviser always chooses l in state L in any equilibrium that satisfies the IC. Suppose not. As shown in the proof of Proposition 1, r_A is always chosen in equilibrium, and then followed by r with probability one in the non-transparency condition. As shown in Lemma 5, at least one of r_A and r_R is played, then followed by r with probability one in the transparency condition (Lemma 5 does not require the assumption that $\beta(L) \geq \gamma$). Thus, l is equilibrium dominated in state R in both conditions. But given the belief $\mu(R|l) = 0$, the adviser will always recommend l in state L, a contradiction. In state R, r_A is played in equilibrium, and then followed by r with probability one in both conditions. Note that in the transparency condition, r_R is never chosen in an equilibrium that satisfies the IC. This is because r_A is equilibrium dominated in state L (as $\beta(L) < \gamma$). Given the belief $\mu(L|r_A) = 0$, r_A is the best choice in state R. As a result, if $\beta(L) < \gamma$, then the equilibrium that satisfies the IC is unique and fully informative under both conditions. If $\beta(L) = \gamma$, then the fully informative equilibrium described above goes through but there are also other equilibria in both conditions, including a corrupt equilibrium.

Proposition 3 *Assume that $\alpha + \gamma > \beta(L)$. If $\beta(R) + \gamma \leq \beta(L)$, then all three equilibria stated in Proposition 2 exist and satisfy D1.*

Proof. We first provide the definition of D1.

[48] For conciseness, we prove the claim under the assumption that $p < 1/2$. The proof under the assumption that $p \geq 1/2$ follows similar lines and is available upon request.

Definition 3 *Let $D_S(m)$ be the set of best responses by D that make A "strictly" prefer m over the equilibrium message in state S, and let $D_S^0(m)$ be the set of best responses by D that make A "weakly" prefer m over the equilibrium message in state S. An equilibrium satisfies D1 if $\mu(S|m) = 0$ holds for every out-of-equilibrium message m such that $D_S^0(m) \subseteq D_{S'}(m) \neq \varnothing$.*

First, we show that the mixed strategy equilibrium is robust to D1. Since $D_R^0(l) \subseteq D_L(l)$, $\mu(R|l) = 0$ must hold according to D1. Given this belief, recommending l is a profitable deviation in state L. However, the mixed strategy equilibrium exists only if $\beta(L) \geq \beta(R) + \gamma$, and therefore, even if $D_R^0(l) \subseteq D_L(l)$ holds, recommending l is *not* a profitable deviation in state L. Hence, the desired result. Next, we show that the honest equilibrium is robust to D1 if and only if $\beta(L) \geq \beta(R) + \gamma$. To see why, note that the honest equilibrium relies on off-the-equilibrium path belief, $\mu(L|r_A) \geq 0.5$. Since $\beta(L) < \beta(R) + \gamma$ implies that $D_L^0(r_A) \subseteq D_R(r_A)$, $\mu(L|r_A) = 0$ is a requirement of D1, a contradiction. However, D1 does not contradict the belief $\mu(L|r_A) \geq 0.5$ if $\beta(L) \geq \beta(R) + \gamma$ because $D_L^0(r_A) \subseteq D_R(r_A)$ no longer holds. Finally, the corrupt equilibrium is robust to D1 regardless of $\beta(L)$, $\beta(R)$ and γ since A obtains the highest possible payoff in either state in the corrupt equilibrium (given our assumption that $\beta(L) \geq \gamma$) and has no incentive to deviate no matter what off-the-equilibrium path beliefs are.

10.2 Instructions

In this section, we provide the instructions for the non-transparency treatment and the transparency treatment, respectively.