

2018-07

# Disclosure and choice

---

Elchanan Ben-Porath, Eddie Dekel, Barton L. Lipman. 2018. "Disclosure and Choice." *The Review of Economic Studies*, Volume 85, Issue 3, 1 July 2018, Pages 1471–1501. <https://doi.org/10.1093/restud/rdx064>  
<https://hdl.handle.net/2144/40543>

*"Downloaded from OpenBU. Boston University's institutional repository."*

# Disclosure and Choice<sup>1</sup>

Elchanan Ben-Porath <sup>2</sup>      Eddie Dekel <sup>3</sup>      Barton L. Lipman<sup>4</sup>

First Draft  
December 2014

Current Draft  
September 2017

<sup>1</sup>We thank Rick Green, Mark Machina, Phil Reny, various seminar audiences, and Botond Koszegi and three anonymous referees for helpful comments and the National Science Foundation, grant SES-0820333 (Dekel), and the US-Israel Binational Science Foundation for support for this research.

<sup>2</sup>Department of Economics and Center for Rationality, Hebrew University. Email: benporat@math.huji.ac.il.

<sup>3</sup>Economics Department, Northwestern University, and School of Economics, Tel Aviv University. Email: dekel@northwestern.edu.

<sup>4</sup>Department of Economics, Boston University. Email: blipman@bu.edu.

## Abstract

An agent chooses among projects with random outcomes. His payoff is increasing in the outcome and in an observer's expectation of the outcome. With some probability, the agent will be able to disclose some information about the true outcome to the observer. We show that choice is inefficient in general. We illustrate this point with a characterization of the inefficiencies that result when the agent can perfectly disclose the outcome with some probability and can disclose nothing otherwise as in Dye (1985a). In this case, the agent favors riskier projects even with lower expected returns. On the other hand, if information can also be disclosed by a challenger who prefers lower beliefs of the observer, the chosen project is excessively risky when the agent has better access to information, excessively risk-averse when the challenger has better access, and efficient otherwise. We also characterize the agent's worst-case equilibrium payoff. We give examples of alternative disclosure technologies illustrating other forms the inefficiencies can take. For example, in a two-dimensional setting, we demonstrate a "hitting for the fences" effect where the agent systematically focuses on the "harder" dimension at the expense of success on the easier.

# 1 Introduction

Consider an agent who makes productive decisions and also decisions about how much to disclose about the outcomes of these choices. The productive decisions are not observed directly and the outcome is only observed after some delay. The agent’s payoff depends on the outcome of the productive decisions but also on the beliefs of an observer regarding the outcome prior to its observation. We give several examples of this situation below.

Intuitively, the agent’s control of information flows gives him an incentive to deviate from efficient productive decisions. For example, he may engage in excessive risk-taking. After all, he can (at least to some extent for some period of time) hide bad outcomes and disclose only good ones. This creates an option value which encourages risk-taking. More broadly, he has an incentive to make production choices that are more likely to give him an opportunity to disclose information that makes him look good even if these choices are less likely to generate good outcomes.

We show that this incentive harms the agent in the sense that he would be better off if he had no control over information. The reason is that the agent has an incentive to try to choose a project that makes the outcome look better than it is. In equilibrium, though, the observer cannot be fooled, so the agent simply hurts himself.

To illustrate these effects of strategic disclosure, we model disclosure as in Dye (1985a): with some probability the agent can disclose the exact realization and otherwise cannot disclose anything. We show that under these conditions, the agent would be better off if he could not affect disclosure. More specifically, with any disclosure process under which the probability that information is disclosed is independent of the information being disclosed, the payoff to the agent is the expected value of the project with the highest expected value.<sup>1</sup> We refer to this payoff as the *first best*. In contrast, when the agent has control of disclosure, he has an incentive to engage in excessive risk-taking, leading to a utility loss relative to the first best which can be “large” in a sense to be made precise.

We now give examples of this setting.

First, consider the manager of a firm. His actions determine a probability distribution over the firm’s profits. In the short run, he can choose to release privately observed information about profits. The observer is the stock market whose beliefs about the firm’s profits determine the stock price of the firm. The manager’s payoff is a convex combination of the short-run and long-run stock price, where the latter is the realized profits — the true value of the firm. Here the first-best project is the one which maximizes the expected value of the firm.

---

<sup>1</sup>For example, if all information that can be disclosed is disclosed (e.g., due to mandatory disclosure requirements), then the disclosure process satisfies this assumption.

One way to interpret this model is to assume that a typical stockholder in the firm has a liquidity shock with some probability which forces him to sell his share. If not forced to sell, the stockholder will have the same information as the market about the value of the stock and so will be indifferent between selling or holding his share. Suppose for simplicity that he always holds his share in this event. Then if the manager chooses actions to maximize the stockholders expected utility, he will maximize a convex combination of the short-run stock price and the realized profits. In that sense, we can reinterpret this example as assuming that the manager acts to maximize the utility of a representative stockholder. Thus the inefficiency we identify is not due to the textbook moral hazard problem (*e.g.*, Mas-Colell, Whinston, and Green (1995), Chapter 14).

Second, suppose the agent is an incumbent politician and the observer is a representative voter. The productive activity chosen by the incumbent is a policy which affects the utility of the voter. Before the outcome of the policy is observed, the incumbent comes up for reelection. As part of his campaign, he may release information regarding the progress of his policies. The probability the voter retains the incumbent is strictly increasing in the voter's beliefs about the utility he will receive from the incumbent's policy choice. One can think of this as retrospective voting or can assume that if the incumbent is not reelected, his policy will be replaced by that of a challenger. The incumbent desires to be reelected and also cares about the true utility of the voters. In this setting, the first-best project is that which maximizes the expected utility of the voters.

Third, an entrepreneur chooses a project which he may need to sell part of to a venture capitalist at the interim stage. The funding he receives is increasing in the beliefs of potential buyers about the value of the project. He may have private information he could disclose at the interim stage regarding how well the project is progressing. Again, the first-best project is the one with the highest expected value.<sup>2</sup>

Fourth, consider a firm with multiple divisions, each of which could potentially head up a prestigious project. The agent is the first division to have an opportunity to lead and the observer is senior management. The agent has to decide among several ways to try to achieve success on the project, where each method corresponds to a probability distribution over profits from the project. The agent may have private information about the progress of the project that he could disclose at the interim stage. If senior management believes the project has not been handled sufficiently well at the interim stage, it transfers control to another division.

In some of these settings, it is natural to consider a challenger to the agent who might also have access to information he can disclose. For example, in the case of an incumbent politician, it is natural to suppose that a challenger running against him might be able to disclose information about the incumbent's policies. Similarly, in the example of a

---

<sup>2</sup>We thank David Kreps for this example.

firm deciding whether to retain the current project manager or opt for an alternative, the alternative manager might have information about what is happening which he could disclose.

Again using the Dye evidence structure, we will show that in the extreme case where all disclosure is by the challenger, the agent has an incentive to behave in a risk-averse manner. In effect, the option value lies entirely with his opponent, so he wishes to minimize risk to reduce the value of this (negative) option. When both the agent and the challenger can disclose, the effect of disclosure on action choice depends on which is more likely to obtain evidence he can disclose. If the agent has more access to information in this sense than the challenger, excessively risky decisions are made, while if the challenger has more access, then excessively risk-averse choices result. Only when information is exactly balanced are production decisions first-best.<sup>3</sup>

While it is an empirical question whether these effects are large in reality, we show that they can be quite large by characterizing the worst possible equilibrium payoff for the agent relative to the first-best payoff. For example, we show that there are parameters for which there is an equilibrium where the agent's payoff is arbitrarily close to 50% of the first-best payoff, but it is impossible for his payoff to be lower than this. (We also characterize the worst-case payoffs with a challenger.) As we show, one advantage of characterizing worst-case payoffs is that this minimum has more intuitive comparative static properties than a characterization of equilibria for fixed parameters.

In the next section, we illustrate the basic ideas with a simple example. In Section 3, we give an overview of the most general version of the Dye model we study. As we show in Section 6, the analysis of the general version can be reduced to the special cases where only the agent has access to information to disclose and where only the challenger has such access. In light of this and the fact that these special cases are simpler, we begin with them in Sections 4 and 5 respectively. In Section 7, we discuss two alternative models, illustrating how the nature of the inefficiencies depends on the disclosure technology. Specifically, we show that if projects differ in the extent to which they yield disclosable evidence, then there is a bias in favor of more transparent projects (those more likely to yield evidence). In a model where projects yield outcomes in more than one dimension, we show that there is a "hitting-for-the-fences" bias: the agent will choose projects that are more likely to succeed in the dimension on which all projects are least likely to succeed. Finally, Section 8 concludes.

The remainder of this introduction is a brief survey of the related literature. There is a large literature on disclosure, beginning with Grossman (1981) and Milgrom (1981). These papers established a key result which is useful for some of what follows. They con-

---

<sup>3</sup>Given the continuity of the model, if information is "close" to balanced, then production decisions are "close" to the first best.

sider a model where an agent wishes to persuade an observer, but only through disclosure — the agent *cannot* affect the underlying distribution over outcomes. They assume the agent is known to have information and show that “unraveling” leads to the conclusion that the unique equilibrium is for the agent to always disclose his information. Roughly, the reasoning is that the agent with the best possible information will disclose, rather than pool with any lower types. Hence the agent with the second–best possible information cannot pool with the better information and so will also disclose, etc. Subsequent important contributions including Verrecchia (1983), Dye (1985a), Jung and Kwon (1988), Fishman and Hagerty (1990), Okuno–Fujiwara, Postlewaite, and Suzumura (1990), Shin (1994, 2003), Lipman and Seppi (1995), Glazer and Rubinstein (2004, 2006), Forges and Koessler (2005, 2008), Archarya, DeMarzo, and Kremer (2011), and Guttman, Kremer, and Skrzypacz (2014) add features to the model which block this unraveling result and explore the implications. To explore the effect of disclosure on productive activities by the agent, we also need a model of disclosure in which unraveling does not occur. We primarily focus on the approach initially developed by Dye (1985a) and Jung and Kwon (1988) for this purpose.

While the literature on disclosure is large, relatively little attention has been paid to the interaction of disclosure and production decisions and the papers that do consider this take very different approaches from ours.<sup>4</sup> Some papers consider “real effects” of disclosure through its effect on the discloser’s competitors (Verrecchia (1983) or Dye (1985b)) or effects that work through how disclosure affects the informativeness of stock prices (Diamond and Verrecchia (1991), Bond and Goldstein (2015), or Gao and Liang (2013)). In Dye and Sridhar (2002), disclosure generates information for the manager through the market’s response to the disclosure. Wen (2013) considers a model where a firm can only disclose if it invests, so that it may undertake unprofitable investments in order to have the opportunity to disclose. While these factors have effects on the firm’s productive decisions, they are very different effects than the incentive issues we study.

There are at least two other literatures where an agent’s productive decisions have informational consequences that influence those decisions. First, in the career concerns literature initiated by Holmstrom (1999), an agent whose abilities are unknown to the market (and possibly to himself) chooses actions whose outcomes are observed by the market and used to form beliefs about his abilities. See Chen (2015) for a recent contribution to and summary of this literature. Second, there are several papers following Stein (1989) in assuming that the manager may have an incentive to divert future cash flows to the present in order to mislead the market about the long–run value of the firm. In this setting, the nature of mandatory disclosure rules (e.g., the frequency of disclosure and the kind of information which must be disclosed) have welfare implications through

---

<sup>4</sup>Numerous papers in the accounting literature have observed that the Dye disclosure model makes the firm’s payoff convex in cash flows, but, to the best of our knowledge, none have noted the implications of this for risk–taking incentives. See, for example, Ostaszewski and Gietzmann (2007).

the effect on the manager’s diversion of cash flows or other investment distortions. See, for example, Kanodia and Mukherji (1996), Kanodia, Sapiro, and Venugopalan (2004), Edmans, Heinle, and Huang (2013), Gigler, Kanodia, Sapiro, and Venugopalan (2013), or the broader overview in Kanodia and Sapiro (2015).<sup>5</sup>

In both of these literatures, the inefficiencies demonstrated are related to the inefficiency we study in that all are generated by an agent’s concern both for the true outcome of his decisions but also the perceptions of an observer. The agent’s desire to influence the latter causes him to take actions which would be suboptimal if he cared only about the former. The key difference between these papers and our work is that we focus on how the agent’s *control* of disclosure affects his incentives. In the career concerns and short-termism literatures, the manager/agent cannot control information except through his productive actions.<sup>6</sup> In our model, the agent controls both factors and the key is the interaction between them.

A different approach to incentive effects associated with strategic disclosure is taken by Beyer and Guttman (2012) who consider a model in which disclosure interacts with investment and financing decisions. Their paper is primarily focused on the signaling effects stemming from private information about the exogenous quality of investment opportunities. Thus both the nature and source of the inefficiency are very different from what we consider.

Finally, we note that the analysis of DeMarzo, Kremer, and Skrzypacz (2017), while motivated differently from ours, has interesting connections with what we do. To clarify the relationship of the models, consider the example above where the agent is the manager of a firm and the observer is the market. In our model, the manager chooses a project which has stochastic outcomes and later may have the option of disclosing information about those outcomes. In DeMarzo, Kremer, and Skrzypacz, the manager does not know and cannot affect the true profits of the firm, but can carry out statistical tests to learn about the profits and then can decide whether to disclose the outcomes of these tests. This is equivalent to specializing our model to the case where all projects available to the manager yield the same expected value of profits. They provide a characterization of the test chosen by the manager and use this characterization to derive results that go beyond the analysis conducted here. For example, they characterize the informativeness of the tests chosen by the manager and how this relates to the best test from the point of view of the observer.

---

<sup>5</sup>These papers can be seen as part of a broader literature on moral hazard in corporate finance and accounting. As in our paper, the manager, even if he represents the interests of current shareholders, has an incentive to take actions to try to “fool” the market or other investors but, of course, is correctly interpreted in equilibrium. As a result, he is worse off than if he could have committed to efficient choices in the first place. See, for example, the risk shifting problem discussed in Jensen and Meckling (1976).

<sup>6</sup>Rodina (2016) considers the case where the principal can control the information.

## 2 Illustrative Example

We begin with an illustrative example to highlight the intuition of our results. This example is for a special case of the environment, where the agent has no challenger and cares only about the observer's beliefs. We explain the model in more detail in the next section, stating here only what is needed for the example. Specifically, we analyze the perfect Bayesian equilibria of a three-stage game. In the first stage, the agent chooses a project to undertake where a project corresponds to a lottery over outcomes in  $\mathbf{R}_+$ . In the second stage, with probability  $q_1$ , the agent receives evidence revealing the exact realization from the project. If he receives evidence, he can either disclose it or withhold it. (If he has no evidence, he cannot show anything.)

The observer does not see the project chosen by the agent or whether he has evidence; the observer sees only the evidence, if any, which is presented. In the third stage, the observer forms a belief  $b$  about the outcome of the project which equals the expectation of the outcome conditional on all public information. Thus if evidence was presented in the second stage, the observer's belief must equal the outcome shown. The agent's payoffs equal the observer's belief,  $b$ .

Consider the following example. Assume  $q_1 \in (0, 1)$ , so the agent may or may not have information. Also, assume that there are only two projects,  $F$  and  $G$ , where  $G$  is a degenerate distribution yielding  $x = 4$  with probability 1 and  $F$  gives 0 with probability 1/2 and 6 with probability 1/2. Recall that the agent's ex ante payoff is the expectation of the observer's belief  $b$ . In equilibrium, the observer will make correct inferences about the outcome of the project given what is or is not disclosed, so the expectation of the observer's belief must equal the expectation of  $x$  under the project chosen by the agent. Hence if we have an equilibrium in which  $F$  is chosen, then the agent's ex ante payoff must be 3, while if we have an equilibrium in which  $G$  is chosen, the agent's ex ante payoff must be 4. In this sense,  $G$  is the best project for the agent. For this reason, we say  $G$  is the *first-best* project and that 4 is the agent's first-best payoff.

Despite the fact that the agent would like to commit to  $G$ , it is not an equilibrium for him to choose it. To see this, suppose the observer expects the agent to choose this project. Then if the agent discloses nothing, the observer believes this is only because the agent did not receive any information (an event with positive probability in the hypothetical equilibrium as  $q_1 < 1$ ) and so believes  $x = 4$ . Given this, suppose the agent deviates to project  $F$ . Since the project choice is not seen by the observer, the observer's beliefs cannot change in response. If the outcome of project  $F$  is observed by the agent to be 0, he can simply not disclose this and the observer will continue to believe that  $x = 4$ . If the outcome is observed to be 6, the agent can disclose this, changing the observer's belief to  $x = 6$ . Hence the agent's payoff to deviating is a convex combination of 4 and 6 and hence is strictly larger than 4. (Specifically, it is

$(1 - q_1)(4) + q_1[(1/2)(4) + (1/2)(6)] > 4$ .) So it is not an equilibrium for the agent to choose project  $G$ . One can show that if  $0 < q_1 \leq 1/2$ , then the unique equilibrium in this example is for the agent to choose project  $F$ .<sup>7</sup> Thus the agent is worse off than in the first-best. His inability to commit leads him to deviate from projects that are efficient but not “showy” enough. Since such deviations are anticipated in equilibrium, he ends up choosing an inefficient project and suffering the consequences.

In this example, the agent’s expected payoff as a proportion of his first-best payoff is  $3/4$ . An implication of Theorem 2 is that, for all  $q_1$  and all sets of feasible projects, the agent’s equilibrium payoff must be at least half the first-best utility and that this bound can be essentially achieved (that is, we can find parameters for which there is an equilibrium payoff as close as we want to this bound).

### 3 Model

In this section, we present the most general version of the model we consider and explain the basic structure of equilibria. In the following sections, we discuss the inefficiencies of the equilibria.

Now the game has three players — the *agent*, the *challenger*, and the *observer*. As in the example, there are three stages. In the first stage, the agent chooses a *project* to undertake. Each project corresponds to a lottery over outcomes. The set of feasible lotteries is denoted  $\mathcal{F}$  where each  $F \in \mathcal{F}$  is a (cumulative) distribution function over  $\mathbf{R}_+$ . For simplicity, we assume the supports of the feasible distributions are bounded from below by 0 and from above by  $\bar{x}$ . That is, we assume that there exists  $\bar{x} < \infty$  such that  $F(\bar{x}) = 1$  for all  $F \in \mathcal{F}$ . We assume the set  $\mathcal{F}$  is finite with at least two elements.<sup>8</sup>

In the second stage, there is a random determination of whether the agent or challenger has evidence demonstrating the outcome of the project. As in Dye (1985a), we assume that evidence, if it exists, proves exactly what the outcome of the project is — there is no “partial” evidence. In the conclusion, we comment briefly on how the results change when partial evidence is possible. Let  $q_1$  denote the probability that the agent has evidence and  $q_2$  the probability that the challenger has evidence. We assume that the events that the agent has evidence and that the challenger has evidence are independent of one another and that both are independent of the project chosen by the agent and its realization.<sup>9</sup> If a player has evidence, then he can either present it, demonstrating

---

<sup>7</sup>If  $q_1 \in (1/2, 1)$ , the unique equilibrium is mixed.

<sup>8</sup>The assumption that  $\mathcal{F}$  is finite is a simple way to ensure equilibrium existence. Also, it is not difficult to allow for unbounded supports as long as all relevant expectations exist.

<sup>9</sup>As shown in Section 6, our results do not rely on the first of these independence assumptions. We

conclusively the outcome of the project, or he can withhold it. If he has no evidence, he cannot show anything. The decisions by the agent and challenger regarding whether to show their evidence (if they have any) are made simultaneously.<sup>10</sup> Neither the agent nor the challenger sees whether the other has evidence. The observer does not see the project chosen by the agent nor whether he or the challenger has evidence — the observer sees only the evidence, if any, which is presented and by whom.

In the third stage, the observer forms a belief  $b$  about the outcome of the project which equals the expectation of  $x$  conditional on all public information.<sup>11</sup> Thus if evidence was presented in the second stage, the observer’s belief must equal the outcome shown since evidence is conclusive.

Finally, the outcome of the project is realized and observed. The payoffs are as follows. Let  $x$  be the realization of the project and  $b$  the observer’s belief in the third stage. The agent’s payoff is  $\alpha x + (1 - \alpha)b$  where  $\alpha \in [0, 1]$ .<sup>12</sup> The challenger’s payoff is  $-b$ . Because the challenger cannot affect  $x$ , the results would be the same if we assumed the challenger’s payoff is  $\beta x + (1 - \beta)(-b)$  for  $\beta \in [0, 1]$ , for example.

Note that the game is completely specified by a feasible set of projects  $\mathcal{F}$  and the values of  $\alpha$ ,  $q_1$ , and  $q_2$ . For this reason, we sometimes write an instance of this game as a tuple  $(\mathcal{F}, \alpha, q_1, q_2)$ . Throughout, we consider perfect Bayesian equilibria (Fudenberg and Tirole, 1991).

Before characterizing equilibria of this game, we consider the benchmark case where the information seen by the observer is not strategically determined. In other words, suppose the observer sees the realization of the project at stage 2 with probability  $q \in [0, 1]$ , unaffected by any actions of the agent or challenger. Except for the degenerate case where  $\alpha = q = 0$ , the optimal project choice by the agent is any project  $F$  which maximizes the expectation of  $x$  with respect to  $F$ , denoted  $E_F(x)$ . We refer to such a project  $F$  as a *first-best project*. To see this, let  $\hat{x}$  denote the belief of the observer if he does not see any evidence. Then if the agent chooses project  $F$ , his expected payoff in equilibrium is

$$\alpha E_F(x) + (1 - \alpha) [q E_F(x) + (1 - q)\hat{x}].$$

---

use it only for notational convenience. We relax the other independence assumption in Section 7.1.

<sup>10</sup>As will be clear from the analysis, the results also hold if the players move sequentially.

<sup>11</sup>For expositional simplicity, we do not explicitly model the payoffs of the observer as they are irrelevant for the equilibrium analysis. Among other formulations, one could assume that the observer chooses an action  $b$  and has payoff  $-(x - b)^2$ . Obviously, the observer would then choose  $b$  equal to the conditional expected value of  $x$ . The examples in the introduction suggest various other payoff functions for the observer.

<sup>12</sup>The linearity of the agent’s payoff in  $x$  and the belief of the observer is not without loss of generality. Similar forces exist with nonlinear payoff functions, but nonlinearity creates additional, potentially quite different, tradeoffs.

Obviously, unless  $\alpha = q = 0$ , the agent's payoff is maximized at any project  $F$  which maximizes  $E_F(x)$ . As we show later in Section 7.1, this conclusion holds even if  $q$  depends on the chosen project, as long as it does not depend on the realized  $x$ .

As the example in Section 2 showed, equilibria are typically not first-best when disclosure is chosen by the agent strategically. If the observer expects the agent to choose a first-best project, he may have an incentive to deviate to a less efficient project which has a better chance of a very good outcome, preventing his choice of the first-best from being an equilibrium. Of course, in equilibrium, his choice is anticipated, so he ends up worse off.

Now we turn to the general structure of equilibria in this model. So suppose we have an equilibrium where the agent uses a mixed strategy  $\sigma$  where  $\sigma(F)$  is the probability the agent chooses project  $F$ . Again, let  $\hat{x}$  denote the belief of the observer if he is not shown any evidence at stage 2. If  $q_1$  and  $q_2$  are both strictly less than 1, then this information set must have a strictly positive probability of being reached.

Given  $\hat{x}$ , it is easy to determine the optimal disclosure strategies for the agent and the challenger. Suppose the agent obtains proof that the outcome is  $x$ . Clearly, he is better off with this revealed if  $x > \hat{x}$  and better off with it not revealed if  $\hat{x} > x$ . It is easy to use this to show that in any equilibrium, the agent discloses  $x$  if  $x > \hat{x}$ .<sup>13</sup> If  $x < \hat{x}$ , the agent discloses only if the challenger is disclosing with probability 1 so that the agent's choice is irrelevant. Finally, if  $x = \hat{x}$ , the equilibrium is entirely unaffected by the disclosure choice so, for simplicity, we assume the agent discloses in this situation.<sup>14</sup> Hence without loss of generality, we can take the agent's strategy to be to disclose  $x$  iff  $x \geq \hat{x}$ . Similar comments apply to the challenger, so we can take his strategy to be to disclose  $x$  iff  $x \leq \hat{x}$ .

In light of this, we can write the agent's payoff as a function of the project  $F$  and  $\hat{x}$

---

<sup>13</sup>Clearly, if the probability the challenger would reveal this information is less than 1, then the agent is strictly better off revealing than not revealing. So suppose the challenger reveals this information with probability 1 — that is,  $q_2 = 1$  and the challenger's strategy given  $x$  is to disclose it. Since the challenger would not want to reveal this information, the only way this could be optimal for the challenger is if the agent is also disclosing it, rendering the challenger indifferent between disclosing and not. Hence, either way, the agent must disclose this information with probability 1.

<sup>14</sup>It is obvious that a player's choice when he observes  $x = \hat{x}$  is irrelevant if this is a measure zero event. However, even with discrete distributions, this remains true. First, obviously, a player's payoff is unaffected by what he does when indifferent. Second, if either the agent or challenger is indifferent, the other is as well, so the agent's choice doesn't affect the challenger or conversely. Finally, the indifferent player's choice does not affect the observer's posterior beliefs since this is a matter of whether we include a term equal to the average in the average or not — it cannot affect the calculation.

as

$$V_A(F, \hat{x}) = \alpha \mathbf{E}_F(x) + (1 - \alpha) [(1 - q_1)(1 - q_2)\hat{x} + q_1(1 - q_2)\mathbf{E}_F \max\{x, \hat{x}\} + q_2(1 - q_1)\mathbf{E}_F \min\{x, \hat{x}\} + q_1 q_2 \mathbf{E}_F(x)]. \quad (1)$$

We can complete the characterization of equilibria as follows. First, given  $\hat{x}$ , we have

$$V_A(F, \hat{x}) = \max_{G \in \mathcal{F}} V_A(G, \hat{x}) \text{ for all } F \text{ such that } \sigma(F) > 0.$$

That is, the agent's mixed strategy is optimal given the disclosure behavior described above and the observer's choice of  $\hat{x}$ .

Second, given  $\sigma$ ,  $\hat{x}$  must be the expectation of  $x$  conditional on no evidence being presented and given the disclosure strategies and the observer's belief that the project was chosen according to distribution  $\sigma$ . The most convenient way to state this is to use the law of iterated expectations to write it as

$$\sum_{F \in \mathcal{F}} \sigma(F) \mathbf{E}_F(x) = \sum_{F \in \mathcal{F}} \sigma(F) [(1 - q_1)(1 - q_2)\hat{x} + q_1(1 - q_2)\mathbf{E}_F \max\{x, \hat{x}\} + q_2(1 - q_1)\mathbf{E}_F \min\{x, \hat{x}\} + q_1 q_2 \mathbf{E}_F(x)]. \quad (2)$$

The left-hand side is the expectation of  $x$  given the mixed strategy used by the agent in selecting a project. The right-hand side is the expectation of the observer's expectation of  $x$  given the disclosure strategies and the agent's mixed strategy for selecting a project.

Substituting from equation (2) into equation (1) yields the conclusion that the agent's equilibrium expected payoff is  $\sum_{F \in \mathcal{F}} \sigma(F) \mathbf{E}_F(x)$ . Thus the agent's payoff in any equilibrium must be weakly below the first-best payoff.

Also, if  $\alpha = q_1 = q_2 = 0$ , then  $V_A(F, \hat{x}) = \hat{x}$ . In this case, the agent's actions do not affect his payoff, so he is indifferent over all projects. Henceforth, we refer to a game  $(\mathcal{F}, \alpha, q_1, q_2)$  with  $\alpha = q_1 = q_2 = 0$  as *degenerate* and call the game *nondegenerate* otherwise.

## 4 Agent Only

In this section, we focus on the case where the challenger is effectively not present. Specifically, we consider the model of the previous section for the special case where  $q_2 = 0$ . This is of interest in part because there is no obvious counterpart of the challenger in some natural examples which otherwise fit the model well. Also, as we will see in

Section 6, the general model can be reduced either to this special case or the special case discussed in the next section where only the challenger may have information.

When  $q_2 = 0$ , equation (1) defining  $V_A(F, \hat{x})$  reduces to

$$V_A(F, \hat{x}) = \alpha E_F(x) + (1 - \alpha)[(1 - q_1)\hat{x} + q_1 E_F \max\{x, \hat{x}\}]. \quad (3)$$

Thus the agent chooses the project  $F$  to maximize  $E_F[\alpha x + (1 - \alpha)q_1 \max\{x, \hat{x}\}]$  for a certain value of  $\hat{x}$ . If  $\hat{x}$  were exogenous and we simply considered  $\alpha x + (1 - \alpha)q_1 \max\{x, \hat{x}\}$  to be the agent's von Neumann–Morgenstern utility function, we would conclude that the agent is risk loving since this is a convex function of  $x$  (as long as  $(1 - \alpha)q_1 > 0$ ). The results we show below build on this observation, making more precise the way this incentive to take risks is manifested in the agent's equilibrium choices.

To clarify the sense in which the agent's choices are risk seeking, we first recall some standard concepts.

**Definition 1.** *Given two distributions  $F$  and  $G$  over  $\mathbf{R}_+$ ,  $G$  dominates  $F$  in the sense of second-order stochastic dominance, denoted  $G$  SOSD  $F$ , if for all  $z \geq 0$ ,*

$$\int_0^z F(x) dx \geq \int_0^z G(x) dx.$$

*We say that  $F$  is riskier than  $G$  if  $G$  SOSD  $F$  and  $E_F(x) = E_G(x)$ .*

It is well-known that if  $G$  SOSD  $F$ , then every risk averse agent prefers  $G$  to  $F$ . If  $F$  is riskier than  $G$ , then every risk-loving agent prefers  $F$  to  $G$  and every risk neutral agent is indifferent between the two.<sup>15</sup>

As shown above, for any  $\hat{x}$ , given the equilibrium disclosure behavior, the agent's payoff function is convex in  $x$ . This implies that if the distribution  $F$  is riskier than  $G$ , then the pure strategy  $F$  must yield a weakly higher payoff for the agent.

Of course, this observation does not imply that there are no equilibria in which the agent chooses  $G$ . Note that the agent's payoff function is piecewise linear. Hence if the distributions of  $F$  and  $G$  differ only within a given linear segment, the agent would see  $F$  and  $G$  as equivalent, so he could certainly choose  $G$  in equilibrium. But then a slight first-order stochastic dominance improvement in either  $F$  or  $G$  would lead the agent to strictly prefer the improved project. This means that we could not say that the agent sacrifices efficiency to choose a riskier project.

---

<sup>15</sup>The reason that the mean condition has to be added for the second two comparisons is that if  $G$  SOSD  $F$ , then the mean of  $G$  must be weakly larger than the mean of  $F$ . Clearly, if it is strictly larger, then  $G$  could be better than  $F$  even for a risk-loving agent.

To characterize the situations where such a sacrifice is made, we need a result where, at equal means, the agent *strictly* prefers the riskier project. Then a small improvement in the efficiency of the less risky project will not lead him to deviate, so inefficiencies occur due to the agent's preference for risk.

Below, we show that under mild conditions, a strengthening of the risk comparison creates the strict comparison we seek, ruling out use of the less risky project in *any* equilibrium. Under this condition, even if we slightly improve the less risky project, the agent continues to choose the less efficient but riskier project.

The stronger notion of riskier is given in the following definition.

**Definition 2.** *Given two distributions  $F$  and  $G$  over  $[a, b]$  with  $0 < a < b$ ,  $G$  strongly dominates  $F$  in the sense of second-order stochastic dominance, denoted  $G$  SSOSD  $F$ , if for all  $z \in (a, b)$ ,*

$$\int_0^z F(x) dx > \int_0^z G(x) dx.$$

*We say that  $F$  is strongly riskier than  $G$  if  $G$  SSOSD  $F$  and  $E_F(x) = E_G(x)$ .*

One can show that if  $F$  is strongly riskier than  $G$ , then for every continuous and increasing utility function  $u$  with uniformly bounded directional derivatives,  $F$  yields strictly higher expected utility than  $G$  if the agent is risk loving and not risk neutral, while  $G$  yields strictly higher expected utility than  $F$  if the agent is risk averse and not risk neutral.

Under three conditions, this notion gives the desired strict comparison. The first is that  $\alpha < 1$  so that the agent cares about the observer's belief, not just the realization of  $x$ . The second is that  $q_1 \in (0, 1)$ . We rule out  $q_1 = 0$  so that the agent has a chance to disclose information and rule out  $q_1 = 1$  so that he has the ability to withhold information as well. Third, we need to ensure that  $\hat{x}$ , the observer's belief if there is no disclosure, is between the lower and upper bound of the supports of the distributions being compared. The simplest way to achieve this is to assume that all projects have the same lower and upper bounds.

**Theorem 1.** *Suppose  $q_2 = 0$ ,  $\alpha < 1$ , and  $q_1 \in (0, 1)$ . Suppose all distributions in  $\mathcal{F}$  have supports with minimum  $\underline{x} \geq 0$  and maximum  $\bar{x} > \underline{x}$ . If there are distributions  $F, G \in \mathcal{F}$  such that  $F$  is strongly riskier than  $G$ , then  $G$  is chosen with zero probability in every equilibrium.<sup>16</sup>*

*Proof.* Suppose to the contrary that there is an equilibrium in which the agent chooses  $G$  with strictly positive probability. Then the payoff to  $G$  must exceed the payoff to  $F$ .

---

<sup>16</sup>This result also holds in a model of project choice with disclosure modeled as in Verrecchia (1983) if the cost of disclosure is small enough.

Using equation (3), this implies

$$\alpha E_G(x) + (1 - \alpha)q_1 E_G \max\{x, \hat{x}\} \geq \alpha E_F(x) + (1 - \alpha)q_1 E_F \max\{x, \hat{x}\}.$$

Since  $F$  is strongly riskier than  $G$ , they have the same mean. Hence, given  $\alpha < 1$  and  $q_1 > 0$ , this reduces to

$$E_G \max\{x, \hat{x}\} \geq E_F \max\{x, \hat{x}\}.$$

Note that

$$E_F \max\{x, \hat{x}\} = F(\hat{x})\hat{x} + \int_{\hat{x}}^{\bar{x}} x dF(x).$$

Integration by parts shows that

$$\int_0^{\hat{x}} F(x) dx = F(\hat{x})\hat{x} - \int_0^{\hat{x}} x dF(x) = F(\hat{x})\hat{x} - E_F(x) + \int_{\hat{x}}^{\bar{x}} x dF(x),$$

so

$$E_F \max\{x, \hat{x}\} = E_F(x) + \int_0^{\hat{x}} F(x) dx.$$

Hence we must have

$$E_G(x) + \int_0^{\hat{x}} G(x) dx \geq E_F(x) + \int_0^{\hat{x}} F(x) dx.$$

Again, since  $F$  is strongly riskier than  $G$ , we have  $E_G(x) = E_F(x)$  implying

$$\int_0^{\hat{x}} G(x) dx \geq \int_0^{\hat{x}} F(x) dx.$$

Since all projects in  $\mathcal{F}$  have minimum value  $\underline{x} \geq 0$  and maximum value  $\bar{x}$ , we must have  $\hat{x} \in [0, \bar{x}]$ . It is easy to show that we cannot have an equilibrium with  $\hat{x} = 0$  or  $\hat{x} = \bar{x}$ , so  $\hat{x} \in (0, \bar{x})$ . But this contradicts  $F$  being strongly riskier than  $G$ . ■

Theorem 1 compares distributions with the same means, but the strictness of the agent's preference implies that he will accept a slightly lower mean in order to obtain more risk.<sup>17</sup>

As an extreme illustration, we generalize the example of Section 2 as follows. Suppose  $\alpha = 0$  and let  $G$  be a degenerate distribution yielding  $x^*$  with probability 1. There is a pure strategy equilibrium in which the agent chooses  $G$  if and only if there is no other feasible distribution that has any chance of producing a larger outcome. That is, this is an equilibrium iff there is no  $F \in \mathcal{F}$  with  $F(x^*) < 1$ . The conclusion that  $G$  is an equilibrium if  $F(x^*) = 1$  for all  $F \in \mathcal{F}$  is obvious, so consider the converse. Suppose we have an equilibrium in which the agent chooses  $G$  but  $F(x^*) < 1$ . Because the agent is

---

<sup>17</sup>See DeMarzo, Kremer, and Skrzypacz's (2017) Proposition III for a related result.

expected to choose  $G$ , we have  $\hat{x} = x^*$ . But then the agent could deviate to  $F$  and with some (perhaps very small probability) will be able to show a better outcome than  $x^*$ , yielding a payoff strictly above  $x^*$ . If he cannot, he shows nothing and receives payoff  $x^*$ . Hence his expected payoff must be strictly larger than  $x^*$ , a contradiction. Note that the mean of  $x$  under  $F$  could be arbitrarily smaller than the mean under  $G$ .

While the mean of  $F$ , the distribution to which the agent deviates, can be arbitrarily smaller than the mean of  $G$ , this does not say that the agent’s payoff loss in equilibrium is arbitrarily large. Since  $F$  may not itself be an equilibrium choice by the agent, such a conclusion would not follow from the observation above. Below, we give tight lower bounds on the ratio of the agent’s equilibrium payoff to his best feasible payoff which show that the equilibrium payoff loss is not, in fact, arbitrarily large. For example, one simple implication of this result is that, except in the degenerate case where  $\alpha = q_1 = 0$ , the agent’s equilibrium payoff must always be at least half of his first–best payoff.

The more general result characterizes the ratio of the worst equilibrium payoff for the agent to the first–best payoff.<sup>18</sup> More precisely, given a game  $(\mathcal{F}, \alpha, q_1, q_2)$ , let

$$U^{FB}(\mathcal{F}) = \max_{F \in \mathcal{F}} E_F(x).$$

So  $U^{FB}$  is the first–best payoff for the agent. Let  $\mathcal{U}(\mathcal{F}, \alpha, q_1, q_2)$  denote the set of equilibrium payoffs for the agent in the game. We construct a function  $R(\alpha, q_1, q_2)$  with the following properties. First, for every  $\mathcal{F}$ , for every  $U \in \mathcal{U}(\mathcal{F}, \alpha, q_1, q_2)$ ,

$$U \geq R(\alpha, q_1, q_2)U^{FB}(\mathcal{F}).$$

That is,  $R(\alpha, q_1, q_2)$  is a lower bound on the proportion of the first–best payoff that can be obtained in equilibrium — i.e., on  $U/U^{FB}$  for any equilibrium for any feasible set  $\mathcal{F}$ .

Second, this bound is tight in the sense that for every  $\varepsilon > 0$ , there exists  $\mathcal{F}$  and  $U \in \mathcal{U}(\mathcal{F}, \alpha, q_1, q_2)$  such that

$$U \leq R(\alpha, q_1, q_2)U^{FB}(\mathcal{F}) + \varepsilon.$$

We therefore sometimes refer to  $R$  as the “worst–case payoff” for the agent.

In this section, we focus on games with  $q_2 = 0$ , so we only characterize the function for this special case here, giving the more general characterization later.<sup>19</sup>

---

<sup>18</sup>This is essentially the inverse of what is sometimes called the Price of Anarchy. See, for example, Koutsoupias and Papadimitriou (1999), who coined the term, or Roughgarden (2005).

<sup>19</sup>The exact statements of the lower bounds in Theorems 2 and 4 exploit our normalization that the outcome from any project is non–negative. However, it is straightforward to adapt these bounds to the more general case where there is some (not necessarily positive) lower bound for all supports. Specifically, suppose  $\underline{x}$  is a lower bound for all supports. When  $\underline{x} = 0$ , our theorems characterize a function  $R$  such that  $U \geq RU^{FB}$  and this bound is tight. When  $\underline{x} \neq 0$ , what we are establishing is that  $U \geq RU^{FB} + (1 - R)\underline{x}$  and that this bound is tight. Note that this implies that if  $\underline{x} \downarrow -\infty$ , then the outcome can be arbitrarily worse than the first–best. We thank Bruno Strulovici for raising this issue.

**Theorem 2.** *For any nondegenerate game,*

$$R(\alpha, q_1, 0) = \frac{\alpha + (1 - \alpha)q_1}{\alpha + (1 - \alpha)q_1(2 - q_1)}.$$

Also,  $R(0, 0, 0) = 0$ . Hence for  $\alpha > 0$ ,

$$\min_{q_1 \in [0, 1]} R(\alpha, q_1, 0) = \frac{1 + \sqrt{\alpha}}{2}.$$

We offer several comments on this result. First, there is a discontinuity in the function  $R$  at the degenerate case where  $\alpha = q_1 = q_2 = 0$ . In the degenerate game, the agent's payoff is  $\hat{x}$ , but his actions cannot affect this. Hence for *any*  $F \in \mathcal{F}$ , it is an equilibrium for the agent to choose  $F$  since no deviation from this  $F$  will change his expected payoff. Thus equilibrium payoffs can be substantially worse than in any nondegenerate game. Consequently, our remaining remarks focus on the nondegenerate case.

Second, it is easy to see that  $R(\alpha, q_1, 0)$  is increasing in  $\alpha$  and equals 1 at  $\alpha = 1$ . Hence, as one would expect, if  $\alpha = 1$ , we obtain the first-best. In this case, the agent does not care about the observer's belief, only the true realization of  $x$ , and so is led to maximize it (in expectation).

Third, it is not hard to show that  $R(\alpha, q_1, 0)$  is not monotonic in  $q_1$  except when  $\alpha = 0$  or (trivially)  $\alpha = 1$ . Specifically, given any  $\alpha \in (0, 1)$ , the unique value of  $q_1$  which minimizes the bound is  $q_1 = \sqrt{\alpha}/[1 + \sqrt{\alpha}]$ , which is interior.

This non-monotonicity stems from the fact that when  $\alpha > 0$ , we obtain the first-best at both  $q_1 = 0$  and at  $q_1 = 1$ . That is,  $R(\alpha, 0, 0) = R(\alpha, 1, 0) = 1$  for all  $\alpha > 0$ . When  $q_1 = 0$ , the agent cannot influence the observer's beliefs and so cares only about the true value of  $x$ . Hence he chooses the project which maximizes its expectation. When  $q_1 = 1$ , he is known to always have information. So the standard unraveling argument implies that he must reveal the information always. Hence he cannot be strategic about disclosure and therefore will again maximize the expected value of  $x$ .

Figure 1 illustrates Theorem 2. It shows  $R(\alpha, q_1, 0)$  as a function of  $q_1$  for various values of  $\alpha$ .

The proof of Theorem 2 is a little tedious and so is relegated to the Appendix. To provide some intuition, we prove a simpler result here, namely, that for  $\alpha = 0$ , the agent's payoff in any pure strategy equilibrium must be at least half the first-best in any nondegenerate game. That is, we prove the last statement of the theorem for  $\alpha = 0$  with a restriction to pure strategies.

So fix any feasible set of projects  $\mathcal{F}$ , any  $q_1 \in (0, 1]$ , and an equilibrium in which the agent chooses project  $F \in \mathcal{F}$ . Fix the  $\hat{x}$  of the equilibrium and let  $G$  be any first-best

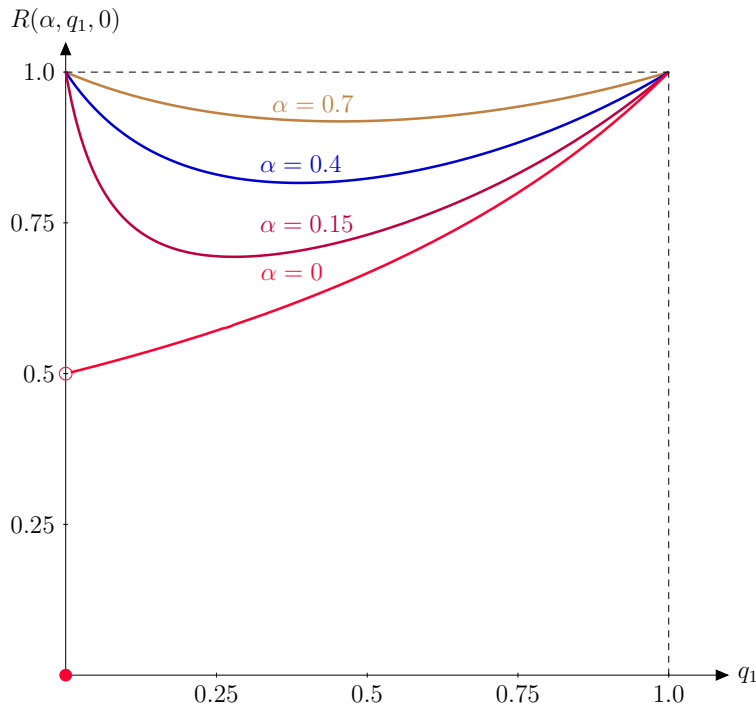


Figure 1: “Worst Case” as a Function of  $q_1$ .

project. From equation (3), when  $q_1 > 0$ , the optimality of  $F$  implies

$$\mathbb{E}_F \max\{x, \hat{x}\} \geq \mathbb{E}_G \max\{x, \hat{x}\}.$$

This implies

$$\mathbb{E}_F(x) + \hat{x} \geq \mathbb{E}_F \max\{x, \hat{x}\} \geq \mathbb{E}_G \max\{x, \hat{x}\} \geq \mathbb{E}_G(x)$$

where the first inequality follows from the fact that the  $x$ 's are always nonnegative. In equilibrium, nondisclosure must always be “bad news” in the sense that  $\mathbb{E}_F(x) \geq \hat{x}$ , implying  $2\mathbb{E}_F(x) \geq \mathbb{E}_F(x) + \hat{x} \geq \mathbb{E}_G(x)$ , so that the agent's equilibrium payoff,  $\mathbb{E}_F(x)$ , must be at least half of the first-best payoff, as claimed.

To show that this bound is approximately achievable, consider the following example. Let  $\alpha = 0$ . Suppose  $\mathcal{F} = \{F, G\}$  where  $F$  is a discrete distribution putting probability  $1 - p$  on 0 and  $p$  on  $1/p$  for some  $p \in (0, 1)$ , so  $\mathbb{E}_F(x) = 1$ . Let  $G$  be a degenerate distribution giving probability 1 to  $x = x^*$ . We construct an equilibrium where  $F$  is chosen by the agent, so the agent's equilibrium payoff,  $U$ , is 1. We focus on the case where  $x^* > 1$ , so  $U^{FB} = x^*$ . If the observer expects the agent to choose  $F$  with probability 1, then by equation (2),  $\hat{x}$  solves

$$(1 - q_1)\hat{x} + q_1 [(1 - p)\hat{x} + 1] = 1$$

so

$$\hat{x} = \frac{1 - q_1}{1 - q_1 p}.$$

This is an equilibrium iff  $E_G \max\{x, \hat{x}\} \leq E_F \max\{x, \hat{x}\}$  or

$$\begin{aligned} \max\{x^*, \hat{x}\} &\leq (1 - p)\hat{x} + 1 \\ &= \frac{(1 - p)(1 - q_1)}{1 - q_1 p} + 1 = \frac{2 - q_1 - p}{1 - q_1 p}. \end{aligned}$$

It is easy to see that  $\hat{x} < 1$  while, by assumption,  $x^* > 1$ . So we have an equilibrium iff

$$x^* \leq \frac{2 - q_1 - p}{1 - q_1 p}.$$

Let  $x^*$  equal the right-hand side. Then we have an equilibrium where the agent's payoff is 1, but the first-best payoff is  $x^*$ . By taking  $q_1$  and  $p$  arbitrarily close to 0, we can make  $x^*$  arbitrarily close to 2, so the agent's payoff is arbitrarily close to half the first-best payoff.<sup>20</sup>

The implication of Theorem 2 that the worst-case payoffs are increasing as the agent cares more about the true  $x$  and less about the observer's belief  $b$  is intuitive, but it is important to note that this result does not carry over to equilibrium payoffs in general. In Appendix C, we give an example which illustrates several senses in which equilibrium payoffs can decrease as  $\alpha$  increases for fixed  $\mathcal{F}$ . In the example, there is a mixed strategy equilibrium with payoffs that are decreasing in  $\alpha$ . Also, this equilibrium is the worst equilibrium for the agent for some parameters, showing that the worst equilibrium payoff for a fixed  $\mathcal{F}$  can decrease with  $\alpha$ . Finally, the payoff in the worst pure strategy equilibrium is also decreasing in  $\alpha$  for a certain range, showing that this result is not an artifact related to mixed-strategy equilibria.

## 5 Challenger Only

In this section, we consider the case where  $q_1 = 0$  and  $q_2$  may be strictly positive. In this case, the agent's payoff as a function of  $\hat{x}$  and his chosen project  $F$  is

$$V_A(F, \hat{x}) = \alpha E_F(x) + (1 - \alpha) [(1 - q_2)\hat{x} + q_2 E_F \min\{x, \hat{x}\}].$$

Analogously to our discussion in Section 4, we see that given  $\hat{x}$ , it is as if the agent has a von Neumann–Morgenstern utility function of  $\alpha x + (1 - \alpha)q_2 \min\{x, \hat{x}\}$ . If  $(1 - \alpha)q_2 > 0$ ,

---

<sup>20</sup>Note that by taking  $p$  arbitrarily close to zero, we can make  $x^*$  arbitrarily close to  $2 - q_1$ , showing the agent's payoff can be arbitrarily close to  $1/(2 - q_1)$  times the first-best payoff, exactly the bound in Theorem 2 when  $\alpha = 0$ .

this function is concave, so the agent's choices are effectively risk averse. Given this, the agent must at least weakly prefer  $G$  to  $F$  whenever  $F$  is riskier than  $G$ . We can strengthen this observation similarly to the way we strengthened the analogous observation in Section 4 to obtain the following analog to Theorem 1.

**Theorem 3.** *Suppose  $q_1 = 0$  and  $(1 - \alpha)q_2 > 0$ . Suppose all distributions in  $\mathcal{F}$  have supports with minimum  $\underline{x} \geq 0$  and maximum  $\bar{x} > \underline{x}$ . If there are distributions  $F, G \in \mathcal{F}$  such that  $F$  is strongly riskier than  $G$ , then  $F$  is chosen with zero probability in every equilibrium.*

*Proof.* The proof parallels the proof of Theorem 1 with min replacing max and concave replacing convex. ■

We can also characterize  $R$  for this case.<sup>21</sup> More specifically, we have the following analog to Theorem 2:

**Theorem 4.** *For all nondegenerate games, we have*

$$R(\alpha, 0, q_2) = \frac{\alpha}{\alpha + (1 - \alpha)q_2}.$$

Hence for  $\alpha > 0$ ,

$$\min_{q_2 \in [0,1]} R(\alpha, 0, q_2) = \alpha$$

and for  $q_2 > 0$ ,

$$\min_{\alpha \in [0,1]} R(\alpha, 0, q_2) = 0.$$

Figure 2 illustrates this result. It shows  $R(\alpha, 0, q_2)$  as a function of  $q_2$  for the same values of  $\alpha$  as used in Figure 1.

Theorem 4 has some features in common with Theorem 2. In particular, both results show that the outcome must be first-best when  $\alpha = 1$  or when  $\alpha > 0$  and there is zero probability of disclosure (i.e.,  $q_2 = 0$ ). In both cases, the worst case improves as  $\alpha$  increases.

On the other hand, this result also shows several differences from Theorem 2. First, this result implies that the worst case over  $\alpha$  when  $q_1 > 0$  and  $q_2 = 0$  is better than the worst case when  $q_1 = 0$  and  $q_2 > 0$ . In other words,

$$\min_{\alpha \in [0,1]} R(\alpha, q_1, 0) > \min_{\alpha \in [0,1]} R(\alpha, 0, q_2)$$

---

<sup>21</sup>While the proofs of Theorems 1 and 3 are essentially identical, those of Theorems 2 and 4 are not.

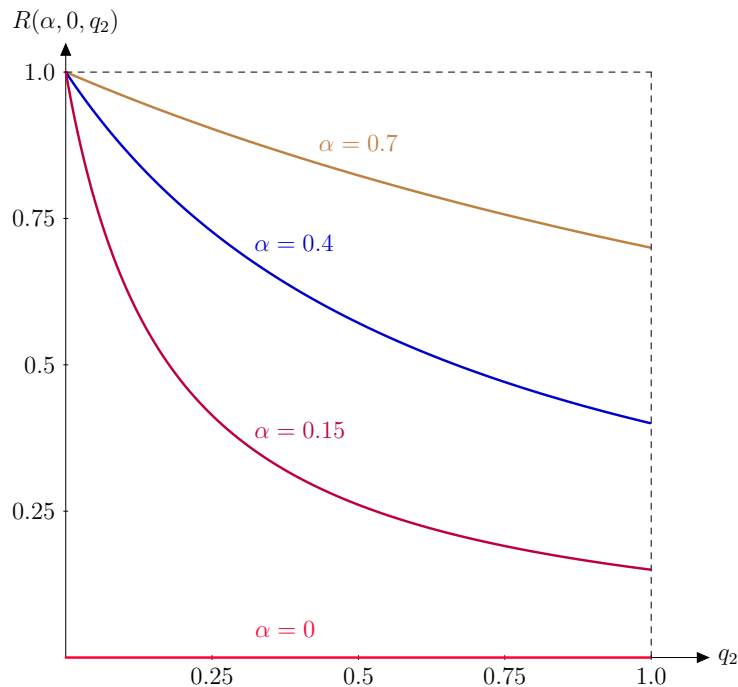


Figure 2: “Worst Case” as a Function of  $q_2$ .

for  $q_1 > 0$  and  $q_2 > 0$ . The left-hand side is  $1/2$ , while the right-hand side is 0. Since the lower bound is zero and payoffs are non-negative, this implies that in the case where only the challenger can disclose, the agent could be arbitrarily worse off than at the first-best.

Second, recall that for  $\alpha \in (0, 1)$ , the worst-case payoff in Theorem 2 was first decreasing, then increasing in  $q_1$ , equalling the first-best at both  $q_1 = 0$  and  $q_1 = 1$ . Here the worst case is always decreasing in  $q_2$ . In particular, we obtain the first-best at  $q_2 = 0$  but not at  $q_2 = 1$ . This may seem unintuitive since at  $q_2 = 1$ , the challenger is known to have information and therefore the standard unraveling argument would seem to suggest he must reveal it. Hence, one might expect, it is as if the observer always saw the true  $x$  and so the outcome would seem to necessarily be first-best.

The following example shows why we do not necessarily obtain the first best at  $q_2 = 1$  and gives some broader intuition for Theorem 4. Suppose  $q_2 = 1$  but  $\alpha = 0$ . In this case, the lower bound given in Theorem 4 holds trivially since it only says that the agent’s equilibrium payoff must be non-negative. To see that we can have equilibria with payoffs arbitrarily close to zero, suppose that  $\mathcal{F} = \{F, G\}$  where  $F$  gives  $\varepsilon \in (0, 50)$  with probability 1, while  $G$  gives 0 with probability  $1/2$  and 100 with probability  $1/2$ . Obviously,  $G$  is the first-best project. But there is an equilibrium in which the agent chooses  $F$  and obtains a payoff of  $\varepsilon$ . To see this, suppose  $F$  is the project the observer

expects the agent to choose. Then if the challenger presents no evidence, the observer believes the outcome to have been  $\varepsilon$  since this is the only feasible outcome under  $F$ . Because of this, the agent has no incentive to deviate to  $G$ . If he does deviate and the outcome is 0, the challenger can show this and the agent is hurt. If the outcome is 100, the challenger can hide this and the observer thinks the outcome was  $\varepsilon$ , so the agent does not gain. Since we only assume  $\varepsilon \in (0, 50)$ , this shows that the agent’s equilibrium payoff can be arbitrarily close to 0.

Intuitively, it is true that if the challenger always learns the outcome of the project, we get unraveling and all information is revealed *along the equilibrium path* — i.e., when the agent chooses the equilibrium project. On the other hand, we do not necessarily get unraveling if the agent deviates to an unexpected project and this fact is what creates the possibility of inefficient equilibria.

On the other hand, the efficient outcome is also an equilibrium if  $q_2 = 1$ .<sup>22</sup> To see this, fix any first–best project  $F$  and suppose the agent is expected to choose this project. Let  $x^*$  denote the supremum of the support of  $F$  and set  $\hat{x} = x^*$ . That is, assume that if the challenger does not reveal  $x$ , the observer believes the realization is the largest possible value under  $F$ . It is easy to see that this is what unraveling implies given that the agent chooses  $F$ . So this is an equilibrium as long as the agent has no incentive to deviate to a different project. By choosing  $F$ , the agent’s payoff is  $E_F(x)$ . If he deviates to any other feasible project  $G$ , his expected payoff is

$$\alpha E_G(x) + (1 - \alpha) E_G \min\{x, x^*\} \leq E_G(x) \leq E_F(x).$$

So the agent has no incentive to deviate.

## 6 Agent and Challenger

Now we consider the case where both the agent and the challenger may have information to disclose in the second stage. The following result shows that the analysis reduces to either the case where only the agent has evidence or the case where only the challenger has evidence, depending on whether  $q_1$  or  $q_2$  is larger.

---

<sup>22</sup>It is also worth noting that the efficient outcome is the only equilibrium when  $q_2 = 1$  if all projects have the same support. That is, in the equal support case,  $R(\alpha, 0, 1) = 1$ . We thank Georgy Egorov for pointing this out. An implication of this is that Theorem 4, unlike all of our other results, does not hold as stated if we add the assumption that all projects have the same support since we can no longer achieve the stated worst–case for  $q_2 = 1$  under this assumption. On the other hand, the only change needed for the equal support case is at  $q_2 = 1$  — for  $q_2 < 1$ , the statement of Theorem 4 is correct even in the equal support case.

**Theorem 5.** Fix  $(\mathcal{F}, \alpha, q_1, q_2)$ . If  $q_1 \geq q_2$ , then the set of equilibria is the same as for the game  $(\mathcal{F}, \hat{\alpha}, \hat{q}_1, 0)$  where

$$\hat{\alpha} = \alpha + (1 - \alpha)q_2$$

and

$$\hat{q}_1 = \frac{q_1 - q_2}{1 - q_2}.$$

If  $q_1 \leq q_2$ , then the set of equilibria is the same as for the game  $(\mathcal{F}, \hat{\alpha}, 0, \hat{q}_2)$  where

$$\hat{\alpha} = \alpha + (1 - \alpha)q_1$$

and

$$\hat{q}_2 = \frac{q_2 - q_1}{1 - q_1}.$$

**Corollary 1.** For any nondegenerate game with  $q_1 = q_2$ , the outcome is first-best.

To see why Theorem 5 implies the corollary, suppose we have a nondegenerate game, so it is *not* the case that  $\alpha = q_1 = q_2 = 0$ . By Theorem 5, if  $q_1 = q_2$ , the outcome is the same in the game with  $\hat{\alpha} = \alpha + (1 - \alpha)q_2 > 0$  and  $\hat{q}_1 = \hat{q}_2 = 0$ . As shown in Theorem 2, the outcome must be first-best in this case.

*Proof of Theorem 5.* Fix  $(\mathcal{F}, \alpha, q_1, q_2)$  and an equilibrium. Let  $\hat{x}$  be the observer's belief if no evidence is presented. First, assume  $q_1 \geq q_2$ . Recall that the agent chooses  $F$  to maximize

$$\begin{aligned} & \alpha E_F(x) + (1 - \alpha) [(1 - q_1)(1 - q_2)\hat{x} + q_2(1 - q_1)E_F \min\{x, \hat{x}\} \\ & \quad + q_1(1 - q_2)E_F \max\{x, \hat{x}\} + q_1 q_2 E_F(x)]. \end{aligned}$$

Note that

$$E_F \min\{x, \hat{x}\} + E_F \max\{x, \hat{x}\} = E_F [\min\{x, \hat{x}\} + \max\{x, \hat{x}\}] = E_F(x) + \hat{x}.$$

Hence

$$E_F \min\{x, \hat{x}\} = E_F(x) + \hat{x} - E_F \max\{x, \hat{x}\}. \quad (4)$$

Substituting, we can rewrite the agent's payoff as

$$[\alpha + (1 - \alpha)q_2]E_F(x) + (1 - \alpha) [(1 - q_1)\hat{x} + (q_1 - q_2)E_F \max\{x, \hat{x}\}]. \quad (5)$$

Let  $\hat{\alpha} = \alpha + (1 - \alpha)q_2$ , so  $1 - \hat{\alpha} = (1 - \alpha)(1 - q_2)$ . We can rewrite the above as

$$\hat{\alpha} E_F(x) + (1 - \hat{\alpha})(1 - \alpha) \left[ \frac{1 - q_1}{(1 - \alpha)(1 - q_2)} \hat{x} + \frac{q_1 - q_2}{(1 - \alpha)(1 - q_2)} E_F \max\{x, \hat{x}\} \right].$$

Let  $\hat{q}_1 = (q_1 - q_2)/(1 - q_2)$ , so  $1 - \hat{q}_1 = (1 - q_1)/(1 - q_2)$ . Then this is

$$\hat{\alpha} E_F(x) + (1 - \hat{\alpha}) [(1 - \hat{q}_1)\hat{x} + \hat{q}_1 E_F \max\{x, \hat{x}\}].$$

This is exactly the agent's payoff when the observer's inference in response to no evidence is  $\hat{x}$  in the game  $(\mathcal{F}, \hat{\alpha}, \hat{q}_1, 0)$ . Hence the agent's best response to  $\hat{x}$  in the game  $(\mathcal{F}, \alpha, q_1, q_2)$  is the same as in the game  $(\mathcal{F}, \hat{\alpha}, \hat{q}_1, 0)$ .

To see that the observer's belief given a mixed strategy by the agent also does not change, note that we can rewrite equation (2) as

$$\begin{aligned} \sum_{F \in \mathcal{F}} \sigma(F) \mathbf{E}_F(x) &= \sum_{F \in \mathcal{F}} \sigma(F) \left\{ \alpha \mathbf{E}_F(x) + (1 - \alpha) \left[ (1 - q_1)(1 - q_2) \hat{x} \right. \right. \\ &\quad \left. \left. + q_1(1 - q_2) \mathbf{E}_F \max\{x, \hat{x}\} + q_2(1 - q_1) \mathbf{E}_F \min\{x, \hat{x}\} + q_1 q_2 \mathbf{E}_F(x) \right] \right\}. \end{aligned}$$

We can rewrite the term in brackets in the same way we rewrote the agent's payoff above to obtain

$$\sum_{F \in \mathcal{F}} \sigma(F) \mathbf{E}_F(x) = \sum_{F \in \mathcal{F}} \sigma(F) \left\{ \hat{\alpha} \mathbf{E}_F(x) + (1 - \hat{\alpha}) \left[ (1 - \hat{q}_1) \hat{x} + \hat{q}_1 \mathbf{E}_F \max\{x, \hat{x}\} \right] \right\},$$

which is the same equation that would define  $\hat{x}$  given  $\sigma$  in the game  $(\mathcal{F}, \hat{\alpha}, \hat{q}_1, 0)$ .

A similar substitution and rearrangement shows the result for  $q_2 \geq q_1$ . ■

This result also holds for arbitrary correlation between the event that the agent receives evidence and the event that the challenger does. To see this, let  $p_b$  be the probability that both have evidence,  $p_1$  the probability that only the agent has evidence,  $p_2$  the probability that only the challenger has evidence, and  $p_n$  the probability that neither has evidence. So we now reinterpret  $q_1$  to be the marginal probability that the agent has evidence — that is,  $q_1 = p_1 + p_b$  — and reinterpret  $q_2$  analogously. It is easy to see that our argument that the challenger will reveal any  $x$  he observes with  $x \leq \hat{x}$  and that the agent will reveal any  $x \geq \hat{x}$  does not rely on any correlation assumption. Hence the agent's payoff as a function of  $F$  and  $\hat{x}$  is now

$$\alpha \mathbf{E}_F(x) + (1 - \alpha) \left[ p_n \hat{x} + p_2 \mathbf{E}_F \min\{x, \hat{x}\} + p_1 \mathbf{E}_F \max\{x, \hat{x}\} + p_b \mathbf{E}_F(x) \right].$$

If we again substitute from equation (4), we obtain

$$\alpha \mathbf{E}_F(x) + (1 - \alpha) \left[ (p_b + p_2) \mathbf{E}_F(x) + (p_n + p_2) \hat{x} + (p_1 - p_2) \mathbf{E}_F \max\{x, \hat{x}\} \right].$$

But  $p_2 + p_b = q_2$ ,  $p_n + p_2 = 1 - p_b - p_1 = 1 - q_1$ , and  $p_1 - p_2 = q_1 - q_2$ . Substituting these expressions, we can rearrange to obtain equation (5) and complete the proof exactly as above.

We can use Theorem 5 to extend Theorems 2 and 4 to this setting. To see this, note that the former theorem tells us that the worst possible payoff for the agent in  $(\mathcal{F}, \alpha, q_1, 0)$  is the first-best payoff times

$$\frac{\alpha + (1 - \alpha)q_1}{\alpha + (1 - \alpha)q_1(2 - q_1)}.$$

Reinterpret this as our “translation” of a game  $(\mathcal{F}, \alpha, q_1, q_2)$  where  $q_1 > q_2$ . In other words, we can treat this lower bound as

$$\frac{\hat{\alpha} + (1 - \hat{\alpha})\hat{q}_1}{\hat{\alpha} + (1 - \hat{\alpha})\hat{q}_1(2 - \hat{q}_1)}$$

where  $\hat{\alpha} = \alpha + (1 - \alpha)q_2$  and  $\hat{q}_1 = (q_1 - q_2)/(1 - q_2)$ . We can substitute in and rearrange to obtain a lower bound as a function of  $(\alpha, q_1, q_2)$  when  $q_1 > q_2$  of

$$\frac{(1 - q_2)[\alpha + (1 - \alpha)q_1]}{\alpha + (1 - \alpha)q_1(2 - q_1) - q_2}.$$

Similar reasoning gives a lower bound when  $q_2 > q_1$  of

$$\frac{\alpha + (1 - \alpha)q_1}{\alpha + (1 - \alpha)q_2}.$$

These bounds reinforce the message of Theorem 5 in that both expressions equal 1 when  $q_1 = q_2$  if either  $\alpha > 0$  or  $q_1 > 0$ . Thus for any nondegenerate game, we obtain the first-best when  $q_1 = q_2$ .

It is intuitive and not hard to see that the properties of  $R$  discussed earlier for the cases  $q_1 = 0$  and  $q_2 = 0$  hold in general. Specifically, the worst-case payoff is increasing in  $\alpha$  and hence is minimal at  $\alpha = 0$ . If  $q_2 > q_1$ , then it is decreasing in  $q_2$ , while if  $q_1 > q_2$ , it is non-monotonic in  $q_1$ . In addition, we now can see that if  $q_i > q_j$ , then  $R$  is continuously increasing in  $q_j$  up to the first best when  $q_j = q_i$ . That is, making the less informed player more equally informed is beneficial. Hence the worst case is that the less informed player has no information at all.

## 7 Alternatives

The simple model explored above shows that the ability of the agent to control the flow of information can give him incentives to take actions which create positive appearances even if these conflict with creating positive outcomes. Since he cannot systematically fool the observer, these incentives end up hurting the agent. In particular, with the Dye model of disclosure, the agent has an incentive to take excessive risk since he can (temporarily) hide bad outcomes. To the extent that hostile forces control the flow of information, the agent has the opposite incentive, namely to avoid risk to an excessive degree.

In this section, we use two examples of alternative forms of disclosure to show how the nature of the inefficiency created by the agent’s control of information depends on

the technology of disclosure. First, we show the bias created when projects vary in the probability that they generate disclosable evidence. Second, we illustrate how the inefficiencies change with the possibility of a particular form of partial disclosure, rather than the all-or-nothing disclosure of Dye. As these examples indicate, the inefficiency generated by strategic disclosure can take many forms.

## 7.1 Varying Transparency Across Projects

In this section, we consider a variation on our model where the challenger never has evidence and the probability the agent has evidence depends on the project he selects. Here we denote a project by the pair  $(F, q_F)$  where  $F$  is a probability distribution over outcomes  $x$  and  $q_F$  is the probability the agent receives evidence he can disclose. In this case, the inefficiency is a distortion toward projects that are more transparent in the sense of being more likely to yield disclosable information.

In particular, if two nondegenerate projects are identical except that one has a larger probability that the agent receives evidence, then the project with the smaller probability of receiving evidence must have zero probability in any equilibrium, a result analogous to Theorem 1. We also give worst-case results analogous to Theorem 2.

First, we show that when the agent cannot control disclosure, he chooses a first-best project in any equilibrium. More specifically, suppose that if the agent chooses project  $F$ , then the observer sees the outcome of the project at stage 2 with probability  $q_F$ . Here the agent's project choice affects the information the observer sees, but the agent does not have a separate disclosure decision. Thus it is not obvious whether the project's effect on observability gives the agent an incentive to choose inefficiently. We reestablish our benchmark for this model by showing that it does not.

To see this, fix any mixed strategy equilibrium  $\sigma$ . Because disclosure is not controlled by the agent, the belief of the observer if he does not see the outcome in stage 2,  $\hat{x}$ , is a weighted average of  $E_F(x)$  for the  $F$ 's in the support of  $\sigma$  where the weights depend on the  $q_F$ 's. With this in mind, suppose  $F$  and  $G$  are projects in the support of  $\sigma$  with  $E_F(x) < E_G(x)$ . If such projects exist, then we can take them to satisfy  $E_F(x) < \hat{x} < E_G(x)$ . Since both are in the support of the mixed strategy, the agent must be indifferent between them, so

$$\alpha E_F(x) + (1 - \alpha)(1 - q_F)\hat{x} + (1 - \alpha)q_F E_F(x) = \alpha E_G(x) + (1 - \alpha)(1 - q_G)\hat{x} + (1 - \alpha)q_G E_G(x),$$

implying

$$\alpha E_F(x) + (1 - \alpha)q_F (E_F(x) - \hat{x}) = \alpha E_G(x) + (1 - \alpha)q_G (E_G(x) - \hat{x}).$$

But  $E_F(x) < E_G(x)$  and  $E_F(x) - \hat{x} < 0 < E_G(x) - \hat{x}$ , a contradiction. Hence  $E_F(x) = E_G(x) = \hat{x}$  for all  $F$  and  $G$  in the support of  $\sigma$ . So if  $F$  is in the support of  $\sigma$  and  $F'$  is not, optimality implies

$$E_F(x) \geq \alpha E_{F'}(x) + (1 - \alpha)(1 - q_{F'})E_F(x) + (1 - \alpha)q_{F'}E_{F'}(x),$$

which implies  $E_F(x) \geq E_{F'}(x)$ . Hence every project with positive probability in equilibrium must be first best when disclosure is nonstrategic.

We now characterize the inefficiencies due to strategic disclosure in this setting. In particular, we show two results which are analogs for Theorems 1 and 2.

**Theorem 6.** *Suppose there are feasible projects  $(F, q_F)$  and  $(G, q_G)$  where  $F = G$ ,  $F$  is nondegenerate, and  $q_G > q_F$ . Then if  $\alpha < 1$ , project  $(F, q_F)$  is chosen with zero probability in any equilibrium.*

*Proof.* Suppose to the contrary that there is an equilibrium in which  $(F, q_F)$  is chosen with strictly positive probability. Then we must have

$$\begin{aligned} \alpha E_F(x) + (1 - \alpha)(1 - q_F)\hat{x} + (1 - \alpha)q_F E_F \max\{x, \hat{x}\} \\ \geq \alpha E_G(x) + (1 - \alpha)(1 - q_G)\hat{x} + (1 - \alpha)q_G E_G \max\{x, \hat{x}\} \end{aligned}$$

where  $\hat{x}$  is the observer's belief if the agent does not disclose any evidence. Since  $F = G$ , this implies

$$q_F[E_F \max\{x, \hat{x}\} - \hat{x}] \geq q_G[E_F \max\{x, \hat{x}\} - \hat{x}].$$

Since  $q_G > q_F$ , this requires  $\hat{x} \geq \bar{x}_F$  where  $\bar{x}_F$  is the upper bound of the support of  $F$ .

Given this, the payoff to  $F$  in equilibrium is  $\alpha E_F(x) + (1 - \alpha)\hat{x} < \hat{x}$ . The inequality follows from  $\hat{x} \geq \bar{x}_F$  and is strict because  $F$  is nondegenerate by assumption. But it is easy to show that the agent's equilibrium payoff is

$$\sum_{F'} \sigma(F') E_{F'}(x) \geq \hat{x},$$

where  $\sigma$  is the agent's mixed strategy. Hence the agent's equilibrium payoff strictly exceeds the payoff to project  $(F, q_F)$ , a contradiction. ■

The following analog of Theorem 2 shows that the worst case is given by the trivial lower bound which says that the agent's equilibrium payoff must weakly exceed the payoff to deviating to the first-best project and having the observer believe that  $x = 0$ .

**Theorem 7.** *For any set of feasible projects, any  $\alpha \in [0, 1]$ , and any equilibrium, the agent's payoff is at least  $\alpha$  times the first-best payoff. Furthermore, there exists a set of feasible projects and an equilibrium such that the agent's payoff equals  $\alpha$  times the first-best.*

*Proof.* To show the bound, fix any set of feasible projects, any  $\alpha$ , and any equilibrium. Let  $U$  be the agent's payoff in the equilibrium and let  $\hat{x}$  be the belief in response to no disclosure in the equilibrium. Let  $(F, q_F)$  be any first-best project. Then

$$\begin{aligned} U &\geq \alpha E_F(x) + (1 - \alpha)q_F E_F \max\{x, \hat{x}\} + (1 - \alpha)(1 - q_F)\hat{x} \\ &\geq \alpha E_F(x) \end{aligned}$$

where the second inequality uses the fact that  $x \geq 0$  with probability 1. Hence  $U$  is at least  $\alpha$  times the first-best payoff.

To see that this is attainable, fix any  $y \geq 0$  and any  $U \in [\alpha y, y)$ . Let the feasible set of projects consist of two projects,  $(F, 0)$  and  $(G, 1)$  where  $F$  yields  $y$  with probability 1 and  $G$  yields  $2U$  with probability  $1/2$  and 0 otherwise. Clearly,  $(F, 0)$  is the first-best project. However, it is easy to see that it is an equilibrium for the agent to choose project  $(G, 1)$ . To see this, suppose it is the project the observer expects. Then  $\hat{x}$  must satisfy

$$U = \frac{1}{2}\hat{x} + \frac{1}{2}(2U),$$

so  $\hat{x} = 0$ . Hence if the agent were to deviate to project  $(F, 0)$ , his payoff would be  $\alpha y + (1 - \alpha)(0)$ . Since  $U \geq \alpha y$ , the agent has no incentive to deviate from  $(G, 1)$ , so this is an equilibrium. In particular, this construction gives an equilibrium even when  $U = \alpha y$ , showing there is an equilibrium with payoff equal to  $\alpha$  times the first-best. ■

## 7.2 Hitting for the Fences

Our second example illustrates possible inefficiencies under partial disclosure. As in Shin (2003), suppose the agent's choice of a project affects two outcomes. For example, if the agent is a political leader, then his choices may affect the economy and also foreign affairs. We refer to the two outcomes as the outcomes on two different issues. For simplicity, suppose that on each issue, the outcome is either a success or a failure, where success corresponds to an outcome of 1 and failure to a 0. So the total value of the projects is simply the number of successes achieved across issues, either 0, 1, or 2.

Suppose there are two possible projects,  $F$  and  $G$ . Let  $f_i$  denote the probability of success on issue  $i$  under project  $F$  and define  $g_i$  analogously. We assume these realizations are independent across issues. Finally, suppose that one issue is harder than the other in the sense that it has a lower probability of success regardless of the project. Specifically, assume  $f_1 > f_2$  and  $g_1 > g_2$ , so it is more difficult to succeed on issue 2.

Our assumptions on disclosure generalize those above. Specifically, with probability  $q$ , the agent is able to disclose the outcome on a given issue, where these events are

independent across issues, independent of the outcome on the issue, and independent of the agent’s project choice. Again, we consider only disclosure by the agent, not by a challenger. As before, one can show that the agent chooses the first–best project if disclosure is nonstrategic (in the sense that the observer sees the outcome on a given issue with probability  $q$ , independent of the outcome on that issue).

When the agent controls disclosure, the bias we get in this setting is a *hitting–for–the–fences effect*. More specifically, if the two projects are equally efficient in the sense that they have the same expected total outcome, then in the unique equilibrium (subject to refinement issues discussed below), the agent chooses whichever project gives the higher chance of success on the harder issue. For concreteness, suppose  $f_1 > g_1 > g_2 > f_2$ . Then  $G$  has the better chance of success on issue 2, the harder issue, so our claim is that the agent chooses  $G$  in the unique equilibrium.

To see the intuition, note that in an equilibrium, the agent will disclose the outcome on an issue if it is a success and he is able to disclose it. He will never disclose a failure on an issue. Thus if he does not disclose anything on an issue, the observer knows that either it was a failure or the agent cannot disclose it. If the agent does not disclose an outcome on issue 1, the observer recognizes that success is relatively likely on issue 1 anyway, so, relative to issue 2, this is more likely to reflect an inability to disclose, not unwillingness to do so. Thus the lack of disclosure is not as harmful to the agent on issue 1 as it is on issue 2. But this means he is more concerned about being able to show a success on issue 2 and hence will focus his efforts there. Then he will prefer project  $G$  to project  $F$  since  $G$  gives the better chance of success on issue 2. Thus even if success is extremely unlikely on issue 2 regardless of the project, the comparison on issue 2 still determines the agent’s choice.

Intuitively, this effect looks like the bias toward riskiness shown in Section 4 since focusing on issue 2 seems riskier. As we show below, our assumptions do imply that the  $G$  is the riskier project. However, this selection of the riskier project only occurs when one issue is harder than the other. If one issue is harder under one project but less hard under the other, then the agent randomizes between projects, even if one project is riskier than the other.

To formalize this intuition, we assume that  $F$  and  $G$  have the same expected value — that is, that  $f_1 + f_2 = g_1 + g_2$ . Since the means are the same and the projects are not identical, one project has a strictly higher success probability on issue 1 and the other has a strictly higher success probability on issue 2. Without loss of generality, assume  $f_1 > g_1$  and  $f_2 < g_2$ . Note for future use that equal means,  $f_1 > g_1$ , and  $g_2 > f_2$  imply that we must have either

$$f_1 > \max\{g_1, g_2\} \geq \min\{g_1, g_2\} > f_2$$

or

$$g_2 > \max\{f_1, f_2\} \geq \min\{f_1, f_2\} > g_1.$$

As suggested above, it seems natural that the agent discloses all successes and no failures. It is easy to see that this will form an equilibrium. For any issue on which the observer is not shown an outcome, he infers (correctly) that either the agent has nothing to disclose on that issue or the outcome was a failure. This is worse for the agent than disclosing a success on that issue but better than disclosing a failure, so the optimal response is disclosing all successes and no failures. While this equilibrium seems natural, others are possible analogously to Appendix A of Shin (2003). We define an equilibrium to be natural if the agent discloses all available successes and no failures and focus only on natural equilibria.

**Theorem 8.** *Fix any  $\alpha \in [0, 1)$  and any  $q_1 \in (0, 1)$ . If  $f_1 > g_1 \geq g_2 > f_2$ , the unique natural equilibrium is for the agent to choose project  $G$  with probability 1. If  $g_2 > f_2 \geq f_1 > g_1$ , the unique natural equilibrium is for the agent to choose project  $F$  with probability 1. Finally, if  $f_1 > g_2 > g_1 > f_2$  or if  $g_2 > f_1 > f_2 > g_1$ , there is a unique natural equilibrium in which the agent chooses a nondegenerate mixed strategy.*

Note that if issue  $i$  is always harder than issue  $j$  in the sense that  $f_j > f_i$  and  $g_j > g_i$ , then we must be in one of the first two cases described in the theorem. Hence the agent chooses the project with the greater probability of success on issue  $i$ , the “hitting for the fences” effect.

*Proof.* Fix a natural equilibrium. Define  $\hat{x}_i$  to be the observer’s expected outcome on issue  $i$  given that the agent does not disclose anything on issue  $i$ . Since the equilibrium is natural, this is well-defined.

The agent prefers  $F$  to  $G$  iff

$$\sum_i [\alpha E_F(x_i) + (1 - \alpha)q_1 E_F \max\{x_i, \hat{x}_i\}] \geq \sum_i [\alpha E_G(x_i) + (1 - \alpha)q_1 E_G \max\{x_i, \hat{x}_i\}].$$

Since the means of the projects are the same and since  $(1 - \alpha)q_1 > 0$ , this holds iff

$$\sum_i E_F \max\{x_i, \hat{x}_i\} \geq \sum_i E_G \max\{x_i, \hat{x}_i\}$$

or

$$f_1 + (1 - f_1)\hat{x}_1 + f_2 + (1 - f_2)\hat{x}_2 \geq g_1 + (1 - g_1)\hat{x}_1 + g_2 + (1 - g_2)\hat{x}_2,$$

where we use  $0 \leq \hat{x}_i \leq 1$ . Using equal means again, we can rewrite this as

$$g_1\hat{x}_1 + g_2\hat{x}_2 \geq f_1\hat{x}_1 + f_2\hat{x}_2$$

or

$$(g_2 - f_2)\hat{x}_2 \geq (f_1 - g_1)\hat{x}_1.$$

By equal means,  $g_2 - f_2 = f_1 - g_1$ . By assumption, this is strictly positive. Hence the agent weakly prefers  $F$  to  $G$  iff  $\hat{x}_2 \geq \hat{x}_1$ .

Given this, when is it an equilibrium for the agent to choose  $F$ ? When the agent chooses  $F$ ,  $\hat{x}_i$  is defined by

$$f_i = (1 - q_1)\hat{x}_i + q_1[f_i + (1 - f_i)\hat{x}_i]$$

so

$$\hat{x}_i = \frac{f_i(1 - q_1)}{1 - q_1 f_i}.$$

Since this is increasing in  $f_i$ , we have  $\hat{x}_2 \geq \hat{x}_1$  iff  $f_2 \geq f_1$ . Therefore, we have an equilibrium in which the agent chooses  $F$  iff  $g_2 > f_2 \geq f_1 > g_1$ .

The analogous reasoning shows that it is an equilibrium for the agent to choose  $G$  iff  $f_1 > g_1 \geq g_2 > f_2$ . Clearly, these parameter conditions are mutually exclusive.

Finally, when is it an equilibrium for the agent to use a nondegenerate mixed strategy? Let  $\sigma$  be the probability on  $F$ . Let  $s_i^\sigma = \sigma f_i + (1 - \sigma)g_i$ . Then the same reasoning as above shows that

$$\hat{x}_i = \frac{s_i^\sigma(1 - q_1)}{1 - q_1 s_i^\sigma}.$$

Indifference between projects implies  $\hat{x}_1 = \hat{x}_2$  or  $s_1^\sigma = s_2^\sigma$ . That is,

$$\sigma f_1 + (1 - \sigma)g_1 = \sigma f_2 + (1 - \sigma)g_2$$

or

$$\sigma[f_1 - g_1 + g_2 - f_2] = g_2 - g_1.$$

Recall that  $f_1 > g_1$  and  $g_2 > f_2$ , so the term multiplying  $\sigma$  is strictly positive. Hence  $\sigma > 0$  iff  $g_2 > g_1$ . Note that  $\sigma < 1$  iff  $f_1 > f_2$ . This implies that we have a nondegenerate mixed equilibrium iff either  $f_1 > g_2 > g_1 > f_2$  or  $g_2 > f_1 > f_2 > g_1$ . ■

As mentioned above, there is a relationship between the “hitting for the fences” effect and a riskiness effect. Specifically, with or without an assumption on whether one issue is always harder than the other, we can compare  $F$  and  $G$  in terms of riskiness. We prove the following result in Appendix D.

**Theorem 9.**  *$G$  is strongly riskier than  $F$  iff  $f_2 > \max\{g_1, g_2\} \geq \min\{g_1, g_2\} > f_1$ .*

Of course, we can reverse the roles of  $F$  and  $G$  above to characterize when  $F$  is strongly riskier than  $G$ .

Comparing Theorems 8 and 9, we see that if issue 2 is the harder issue regardless of the project, then  $G$  is the unique equilibrium and the strongly riskier project. Similarly, if issue 1 is always the harder issue, then  $F$  is the unique equilibrium and the strongly riskier project. On the other hand, if neither issue is always harder, then one of the projects is riskier than the other, but the unique equilibrium has the agent mixing and thus putting positive probability on the less risky project. For example, if  $f_1 > g_2 > g_1 > f_2$ , then  $G$  is the strongly riskier project by Theorem 9. But Theorem 8 shows that there is a unique mixed equilibrium in this case. Hence in this setting of partial disclosure, the conclusion of Theorem 1 does not hold:  $G$  is strongly riskier than  $F$  but the agent chooses  $F$  with positive probability in equilibrium. There is an inefficiency here in that we could slightly worsen  $F$  or  $G$  and the agent will still put positive probability on both, but this is a different inefficiency from the one discussed in Theorem 1.

## 8 Discussion

We conclude with some brief comments on omitted factors that might be of interest to explore further. One natural factor to consider is the possibility of “noise” in the disclosure process. It is natural to wonder if our results are robust to the possibility that the evidence disclosed by either the agent or challenger is a noisy signal of  $x$  rather than the realization  $x$  itself.

To see why one might suspect nonrobustness, consider the model where only the agent may have evidence and suppose that there are two projects,  $F$  and  $G$ , where  $F$  yields  $x = 2$  with certainty and  $G$  gives  $x = 0$  or  $x = 3$ , each with probability  $1/2$ . For any sufficiently small  $\alpha$  and any  $q_1 \in (0, 1)$ , in the model without noise, it is never an equilibrium for the agent to choose  $F$ . However, now suppose that the evidence the agent might obtain in the disclosure stage is noisy. Specifically, suppose there is a set of signals, say  $\mathcal{S}$ , and that the distribution over signals received by the agent is a full support distribution which depends on the true outcome. That is, if the true outcome is  $x$ , then the distribution over signals is  $\psi(\cdot | x)$  and this distribution has full support on  $\mathcal{S}$  for any  $x$ . Then it is *always* an equilibrium for the agent to choose  $F$ . If the observer expects the agent to choose  $F$ , then he expects  $x$  to equal 2 and his belief will not change regardless of the signal the agent shows him, if any. Hence the agent has no incentive to deviate.

On the other hand, it is easy to see that this example relies critically on the degeneracy of the chosen project. In fact, if we assume that all projects have the same support (that is, the same set of possible outcomes), then the discontinuity at zero noise disappears. To see this, think of the observer as having a prior belief about the outcome given by the project he expects the agent to choose. For any full support “prior,” sufficiently precise

signals will generate a posterior belief close to the true realization of the outcome. Thus if all projects have the same support, the fact that the observer’s prior would be, in a sense, wrong when the agent deviates will not prevent the observer from assigning probability close to the 1 to the true outcome if the agent discloses a sufficiently precise signal. Consequently, the set of equilibria for “small noise” and for “zero noise” will necessarily be “close.”<sup>23</sup> While our analysis is therefore robust with respect to small amounts of noise under this full-support assumption, the introduction of noise may introduce new issues and effects worth exploring.

Another direction to consider is to return to our discussion of transparency above and consider the possibility that the agent can take actions which determine the probability that he or the challenger receive evidence. There are a number of delicate modeling questions here. Are the agent’s actions regarding transparency observable? If so, he may have the ability to commit to a  $q_1$ . In this case, at least if these actions are costless, he would commit to  $q_1 = 1$  and achieve the first-best outcome.<sup>24</sup> If his actions aren’t observable but are costless, he still has an incentive to choose  $q_1 = 1$  since this ensures he can disclose if he wishes to do so. On the other hand, if his actions are unobserved and costly, things are more complex, particularly if the challenger can also choose actions which affect his probability of receiving evidence.

Finally, given the severe inefficiency of equilibria in this environment, it is natural to ask whether players would find ways to improve the outcomes by some richer incentive devices. In some cases, this seems difficult or impossible — e.g., in the voting example. There it seems that the best one can do is to give equal access to information to the challenger and incumbent (something that presumably a free press can help maintain). In other environments, contracting may help. For example, suppose the agent is the manager of a firm and the observer is the stock market. Then it seems natural to expect the firm’s stockholders to alter the agent’s compensation in order to induce more efficient behavior. Intuitively, the model implies that inefficiency results in part from the fact that the manager’s payoff is increasing in the “short run” stock price — i.e., the stock price before the outcome of the project is revealed to all. If his payoff instead depended only on the “long run” stock price — i.e., the realization of  $x$  — the outcome would be first-best.

As has been noted in the literature,<sup>25</sup> there are good reasons for expecting managerial compensation to depend positively on both short-run and long-run stock prices. First, if the long run is indeed long, the manager requires compensation in the short run too.

---

<sup>23</sup>It is worth noting that we could also add noise to the model in way which obviously has no effect on our results. Specifically, suppose that the realized outcome is the signal drawn in the disclosure phase (whether this is observed or not) plus an independent, mean zero, random variable. In this case, the best estimate of the outcome conditional on the disclosure of a signal realization of  $x$  is simply  $x$ , so none of our analysis changes at all. We thank Andy Skrzypacz for pointing this out.

<sup>24</sup>We thank David Kreps for pointing this out.

<sup>25</sup>See, for example, the discussion in Stein (1989) or Edmans, Heinle, and Huang (2013).

Due to limited liability, it seems implausible that he can be forced to repay short-run compensation if the realization of the project turns out to be poor in the long run. Second, there is an issue as to whether stockholders can commit to not rewarding short-run stock prices. To see the point, suppose that stockholders may need to sell their holdings in the short run and hence care about the short-run stock price.<sup>26</sup> If the manager has positive news in the second period, then they would be better off at this point if he would disclose it. Hence even if the original contract for the manager did not reward him for a high short-run stock price, the stockholders would have an incentive to renegotiate the contract after the project choice is made. Of course, if the manager anticipates this, it is as if the original contract depended on the short-run price. Optimal contracting in such an environment is a natural next step to consider.

---

<sup>26</sup>This formulation is common in the literature. See, for example, Diamond and Verrecchia (1991) or Gigler, Kanodia, Saprà, and Venugopalan (2013).

# Appendix

## A Proof of Theorem 2

Consider any game  $(\mathcal{F}, \alpha, q_1, 0)$ . Since the conclusion that  $R(0, 0, 0) = 0$  was shown in the text, we focus here only on nondegenerate games so either  $\alpha > 0$  or  $q_1 > 0$  (or both).

It is easy to see that  $R(1, q_1, 0) = 1$ . If  $\alpha = 1$ , the agent's payoff from choosing  $F$  is  $E_F(x)$ , independently of the strategy of the observer. Hence he must maximize this and so his payoff must be the first-best. For the rest of this proof, assume  $\alpha < 1$ .

It is also not hard to show that  $R(\alpha, 1, 0) = 1$ . To see this, suppose  $q_1 = 1$  but we have an equilibrium in which the agent's payoff is strictly below the first-best. Then the agent could deviate to any first-best project and always disclose the outcome. Since  $q_1 = 1$ , this ensures the agent a payoff equal to the first-best, a contradiction. Since equilibria always exist, we see that  $R(\alpha, 1, 0) = 1$ . For the rest of this proof, we assume  $q_1 < 1$ .

For a fixed  $\hat{x}$ , the agent's payoff to choosing  $F$  is

$$\alpha E_F(x) + (1 - \alpha)[(1 - q_1)\hat{x} + q_1 E_F \max\{x, \hat{x}\}]. \quad (6)$$

As shown in the text,  $E_F \max\{x, \hat{x}\} = E_F(x) + \int_0^{\hat{x}} F(x) dx$ , so we can rewrite this as

$$(\alpha + (1 - \alpha)q_1)E_F(x) + (1 - \alpha)(1 - q_1)\hat{x} + (1 - \alpha)q_1 \int_0^{\hat{x}} F(x) dx.$$

Fix an equilibrium mixed strategy for the agent  $\sigma$  and the associated  $\hat{x}$ . Let  $U = \sum_{F' \in \mathcal{F}} \sigma(F') E_{F'}(x)$ , so this is the agent's expected payoff in the equilibrium. Let  $F$  be any project in the support of the agent's mixed strategy such that  $E_F(x) \leq U$  and let  $G$  be any other feasible project. Then we must have

$$\begin{aligned} (\alpha + (1 - \alpha)q_1)E_G(x) + (1 - \alpha)q_1 \int_0^{\hat{x}} G(x) dx \\ \leq (\alpha + (1 - \alpha)q_1)E_F(x) + (1 - \alpha)q_1 \int_0^{\hat{x}} F(x) dx. \end{aligned}$$

Since  $G(x) \geq 0$ , this implies

$$(\alpha + (1 - \alpha)q_1)E_G(x) \leq (\alpha + (1 - \alpha)q_1)E_F(x) + (1 - \alpha)q_1 \int_0^{\hat{x}} F(x) dx.$$

Define  $z = \int_0^{\hat{x}} F(x) dx / \hat{x}$ . It is not hard to use equation (2) to show that  $q_1 < 1$  implies  $\hat{x} > 0$ , so this is well-defined.<sup>27</sup> Since  $F(x) \in [0, 1]$ , we must have  $z \in [0, 1]$ . Then we can rewrite this equation as

$$(\alpha + (1 - \alpha)q_1)E_G(x) \leq (\alpha + (1 - \alpha)q_1)E_F(x) + (1 - \alpha)q_1z\hat{x}. \quad (7)$$

Since  $F$  is in the support of the agent's equilibrium mixed strategy, we must have

$$(\alpha + (1 - \alpha)q_1)E_F(x) + (1 - \alpha)(1 - q_1)\hat{x} + (1 - \alpha)q_1z\hat{x} = U,$$

so

$$\hat{x} = \frac{U - (\alpha + (1 - \alpha)q_1)E_F(x)}{(1 - \alpha)[1 - q_1 + zq_1]}.$$

Substituting into equation (7) gives

$$(\alpha + (1 - \alpha)q_1)E_G(x) \leq (\alpha + (1 - \alpha)q_1)E_F(x) + q_1z \left[ \frac{U - (\alpha + (1 - \alpha)q_1)E_F(x)}{1 - q_1 + zq_1} \right]. \quad (8)$$

Recall that  $U \geq E_F(x)$ , so  $U \geq (\alpha + (1 - \alpha)q_1)E_F(x)$ . Hence  $q_1 \geq 0$  implies that the right-hand side is weakly increasing in  $z$ . Hence

$$(\alpha + (1 - \alpha)q_1)E_G(x) \leq (\alpha + (1 - \alpha)q_1)E_F(x) + q_1 [U - (\alpha + (1 - \alpha)q_1)E_F(x)]$$

or

$$[\alpha + (1 - \alpha)q_1]E_G(x) \leq Uq_1 + E_F(x)(\alpha + (1 - \alpha)q_1)(1 - q_1).$$

Since the term multiplying  $E_F(x)$  is positive, the fact that  $E_F(x) \leq U$  implies

$$(\alpha + (1 - \alpha)q_1)E_G(x) \leq U [q_1 + (\alpha + (1 - \alpha)q_1)(1 - q_1)].$$

Hence, taking  $G$  to be a first-best project,

$$U \geq U^{FB} \left[ \frac{\alpha + (1 - \alpha)q_1}{\alpha + (1 - \alpha)q_1(2 - q_1)} \right].$$

To show that this bound is tight, consider the following example. Suppose  $\mathcal{F} = \{F, G\}$ . Assume  $F$  is a distribution putting probability  $1 - p$  on 0 and  $p$  on  $U/p$ , so  $E_F(x) = U$ , for some  $p \in (0, 1)$  and  $U > 0$ . Let  $G$  be a distribution putting probability

---

<sup>27</sup>To see this, suppose  $\hat{x} = 0$ . Then equation (2) implies that either  $q_1 = 1$  or  $E_F(x) = 0$  for all  $F$  in the support of the agent's mixed strategy. Since  $q_1 < 1$  by assumption, this implies  $U = 0$ . But this is not possible. The agent can deviate to any project with a strictly positive mean (since there are at least two projects, such a project must exist) and always show the outcome. Since either  $\alpha > 0$  or  $q_1 > 0$  or both, the agent would gain by such a deviation.

1 on  $x^*$  for some  $x^* > U$ . Note that  $E_F(x) = U < x^* = E_G(x)$ , so  $U^{FB} = x^*$ . We will characterize a situation where  $F$  is a pure strategy equilibrium and show that this establishes the bound. Note that if  $F$  is chosen with probability 1 in equilibrium, then we must have  $\hat{x} < U < x^*$ . Hence  $\int_0^{\hat{x}} G(x) dx = 0$  and  $\int_0^{\hat{x}} F(x) dx = (1-p)\hat{x}$ . Hence  $F$  is optimal for the agent iff equation (7) holds at  $E_G(x) = x^*$ ,  $E_F(x) = U$ , and  $z = 1-p$ . We can also solve for  $\hat{x}$  exactly as above with  $z = 1-p$  and  $E_F(x) = U$ . Therefore, from equation (8), this is an equilibrium iff

$$(\alpha + (1-\alpha)q_1)x^* \leq U \left[ \alpha + (1-\alpha)q_1 + q_1(1-p) \left( \frac{1 - (\alpha + (1-\alpha)q_1)}{1 - q_1 + (1-p)q_1} \right) \right].$$

Tedious algebra leads to

$$U \geq x^* \left( \frac{(\alpha + (1-\alpha)q_1)(1 - q_1 + (1-p)q_1)}{(\alpha + (1-\alpha)q_1)(1 - q_1) + (1-p)q_1} \right).$$

Fix  $p$  and choose  $x^*$  so that this holds with equality. (It is immediate that the resulting  $x^*$  is necessarily larger than  $U$ , as assumed.) For  $p$  arbitrarily close to 0, we obtain an example where

$$U \approx U^{FB} \left( \frac{\alpha + (1-\alpha)q_1}{(\alpha + (1-\alpha)q_1)(1 - q_1) + q_1} \right) = U^{FB} \left( \frac{\alpha + (1-\alpha)q_1}{\alpha + (1-\alpha)q_1(2 - q_1)} \right).$$

Hence

$$R(\alpha, q_1, 0) = \frac{\alpha + (1-\alpha)q_1}{\alpha + (1-\alpha)q_1(2 - q_1)}.$$

It is not hard to show that  $1/R$  is concave in  $q_1$  and that the first-order condition for maximization of  $1/R$  holds uniquely at

$$q_1 = \frac{\sqrt{\alpha}}{1 + \sqrt{\alpha}}.$$

Thus  $R$  is uniquely minimized at this  $q_1$ . Substituting this value of  $q_1$  into  $R$  and rearranging yields

$$\min_{q_1 \in [0,1]} R(\alpha, q_1, 0) = \frac{1 + \sqrt{\alpha}}{2},$$

as asserted. ■

## B Proof of Theorem 4

Again, nondegeneracy implies that either  $\alpha > 0$  or  $q_2 > 0$  or both. Just as in the proof of Theorem 2, the result that we obtain the first-best when  $\alpha = 1$  is straightforward, so

we assume throughout this proof that  $\alpha < 1$ . The case of  $\alpha = 0$  is also straightforward. To see this, suppose there is a distribution  $F \in \mathcal{F}$  which is degenerate at 0. Suppose the observer believes the agent chooses this distribution and the challenger never shows any strictly positive  $x$ . Then since  $\alpha = 0$ , no deviation by the agent can achieve a strictly positive payoff. No matter what the agent does, the observer's belief is that  $x = 0$ , so the agent's payoff is zero. Hence this is an equilibrium, establishing that  $R(0, 0, q_2) = 0$  for any  $q_2$ . Hence for the rest of this proof, we assume  $\alpha \in (0, 1)$ .

Given that  $q_1 = 0$ , we can write the agent's payoff given  $\hat{x}$  and a choice of project  $F$  as

$$\alpha E_F(x) + (1 - \alpha)(1 - q_2)\hat{x} + (1 - \alpha)q_2 E_F \min\{x, \hat{x}\}.$$

Since  $E_F \min\{x, \hat{x}\} = \int_0^{\hat{x}} [1 - F(x)] dx$ , we can rewrite this as

$$\alpha E_F(x) + (1 - \alpha)(1 - q_2)\hat{x} + (1 - \alpha)q_2 \int_0^{\hat{x}} [1 - F(x)] dx.$$

So fix an equilibrium mixed strategy for the agent  $\sigma$  and the associated  $\hat{x}$ . Again, let  $U$  be the agent's expected payoff — that is,  $U = \sum_{F' \in \mathcal{F}} \sigma(F') E_{F'}(x)$ . Let  $F$  be a project in the support of the agent's mixed strategy satisfying  $E_F(x) \leq U$  and let  $G$  be any other feasible project. Then we must have

$$\alpha E_G(x) + (1 - \alpha)q_2 \int_0^{\hat{x}} [1 - G(x)] dx \leq \alpha E_F(x) + (1 - \alpha)q_2 \int_0^{\hat{x}} [1 - F(x)] dx.$$

Since  $G(x) \leq 1$ , this implies

$$\alpha E_G(x) \leq \alpha E_F(x) + (1 - \alpha)q_2 \int_0^{\hat{x}} [1 - F(x)] dx.$$

Define  $z = \int_0^{\hat{x}} [1 - F(x)] dx / \hat{x}$ . One can use equation (2) and  $\alpha > 0$  to show that  $\hat{x} > 0$  so this is well-defined.<sup>28</sup> As in the proof of Theorem 2,  $F(x) \in [0, 1]$  implies  $z \in [0, 1]$ . Then we can rewrite this equation as

$$\alpha E_G(x) \leq \alpha E_F(x) + (1 - \alpha)q_2 z \hat{x}. \tag{9}$$

Because  $F$  is in the support of the agent's equilibrium mixed strategy, we must have

$$\alpha E_F(x) + (1 - \alpha)(1 - q_2)\hat{x} + (1 - \alpha)q_2 z \hat{x} = U,$$

---

<sup>28</sup>To see this, suppose  $\hat{x} = 0$ . From equation (2), this implies that

$$\sum_{F' \in \mathcal{F}} \sigma(F') E_{F'}(x) = q_2 \sum_{F' \in \mathcal{F}} \sigma(F') E_{F'} \min\{x, 0\} = 0.$$

Hence the agent's mixed strategy must put probability 1 on a degenerate distribution at 0 and so  $U = 0$ . Since  $\alpha > 0$ , the agent can deviate to any other project (which must have a strictly positive mean) and be strictly better off even if the challenger never discloses anything.

so

$$\hat{x} = \frac{U - \alpha E_F(x)}{(1 - \alpha)(1 - q_2 + zq_2)}.$$

Substituting into equation (9) gives

$$\alpha E_G(x) \leq \alpha E_F(x) + q_2 z \left[ \frac{U - \alpha E_F(x)}{1 - q_2 + zq_2} \right]. \quad (10)$$

By assumption,  $U \geq E_F(x)$ , so  $U \geq \alpha E_F(x)$ . Hence the right-hand side is weakly increasing in  $z$ , so this implies

$$\alpha E_G(x) \leq \alpha E_F(x) + q_2 [U - \alpha E_F(x)]$$

or

$$\alpha E_G(x) \leq q_2 U + \alpha(1 - q_2)E_F(x) \leq U[q_2 + \alpha(1 - q_2)] = U[\alpha + (1 - \alpha)q_2].$$

Hence, taking  $G$  to be a first-best project,

$$U \geq U^{FB} \left[ \frac{\alpha}{\alpha + (1 - \alpha)q_2} \right].$$

To see that the bound is tight, let  $F$  be a degenerate distribution at  $x^*$  and suppose we have an equilibrium where the agent chooses  $F$ . Clearly, then,  $\hat{x} = U = x^*$ . Let  $G$  put probability  $1 - p$  on 0 and  $p$  on  $y/p$  where  $y > x^*$  for some  $p \in (0, 1)$ . Note that  $E_G(x) = y$ . Assume  $F$  and  $G$  are the only feasible projects. Then this is an equilibrium if

$$\alpha y + (1 - \alpha)q_2[(1 - p)(0) + p\hat{x}] \leq (\alpha + (1 - \alpha)q_2)\hat{x}.$$

Since  $\hat{x} = U$ , we can rewrite this as

$$\alpha y \leq U[\alpha + (1 - \alpha)(1 - p)q_2].$$

Fix any  $p \in (0, 1)$  and choose  $y$  so that this holds with equality. Since the resulting  $y$  satisfies  $y \geq U$ , we have  $U^{FB} = y$ . So this gives an example where

$$U = U^{FB} \left[ \frac{\alpha}{\alpha + (1 - \alpha)(1 - p)q_2} \right].$$

As  $p \downarrow 0$ , the right-hand side converges to  $\alpha/[\alpha + (1 - \alpha)q_2]$ . Hence we can get arbitrarily close to the stated bound, so

$$R(\alpha, 0, q_2) = \frac{\alpha}{\alpha + (1 - \alpha)q_2}.$$

The last two statements of the theorem follow directly. ■

## C Comparative Statics Example

Suppose there are three feasible projects,  $F_1$ ,  $F_2$ , and  $F_3$ . Project  $F_i$  gives a “high outcome”  $h_i$  with probability  $p_i$  and a “low outcome”  $\ell_i$  otherwise. The specific values of  $h_i$ ,  $\ell_i$ , and  $p_i$  are given in the table below.

	$h_i$	$\ell_i$	$p_i$	$\mu_i$
$F_1$	964	532	1/2	748
$F_2$	5904/7	0	7/8	738
$F_3$	1737/2	171	4/5	729

In the table,  $\mu_i = E_{F_i}(x)$ . Note that  $F_1$  is the first-best project,  $F_2$  is second best, and  $F_3$  worst. Simple calculations show the range of  $\alpha$ 's for which it is a pure strategy equilibrium for the agent to choose  $F_i$  for each  $i$ . For each of the three projects, there is a nonempty range of  $\alpha$ 's where it is chosen in equilibrium. Similarly, for each pair of projects, there is a nonempty range of  $\alpha$ 's where that pair is the support of the agent's mixed strategy.

In the case where the agent randomizes between projects  $F_1$  and  $F_2$  or between  $F_1$  and  $F_3$ , the agent's equilibrium payoff decreases with  $\alpha$ . On the other hand, the equilibrium payoff when the agent randomizes between  $F_2$  and  $F_3$  is increasing in  $\alpha$ .

To see the intuition, consider the case where the agent randomizes between  $F_1$  and  $F_2$ . As  $\alpha$  increases, if  $\hat{x}$  is fixed, the agent would switch to  $F_1$  since he now cares more about the outcome of the project and  $F_1$  has the higher expected outcome. So  $\hat{x}$  must adjust to deter this deviation. Which way do we need to adjust  $\hat{x}$  to make the agent indifferent again? Note that  $F_2$  has a much higher chance of having a good outcome to show than  $F_1$ . Thus if  $\hat{x}$  declines, this pushes the agent toward  $F_2$ . Hence the adjustment that restores indifference is reducing  $\hat{x}$ . To reduce  $\hat{x}$ , we must make the observer more pessimistic about the outcome. This means we must reduce the probability that the agent picks  $F_1$ , lowering the agent's equilibrium payoff. Similarly, note that  $F_3$  gives its high outcome with higher probability than  $F_1$ , so similar intuition applies here. On the other hand, in comparing  $F_2$  and  $F_3$ , it is  $F_2$ , the better of the two projects, which has the higher chance of the high outcome. Hence the opposite holds in this case.

The figure below shows the equilibrium payoffs as a function of  $\alpha$ . Note that, as asserted, the equilibrium payoffs for two of the three mixed strategy equilibria are decreasing in  $\alpha$ . Note also that the payoff to the worst equilibrium is decreasing in  $\alpha$  for  $\alpha$  between  $1/4$  and  $1/3$ . Finally, note that if we focus only on pure strategy equilibria, the worst equilibrium payoff is decreasing in  $\alpha$  as we move from the range where  $\alpha \in [5/24, 1/3]$  to  $\alpha > 1/3$ .

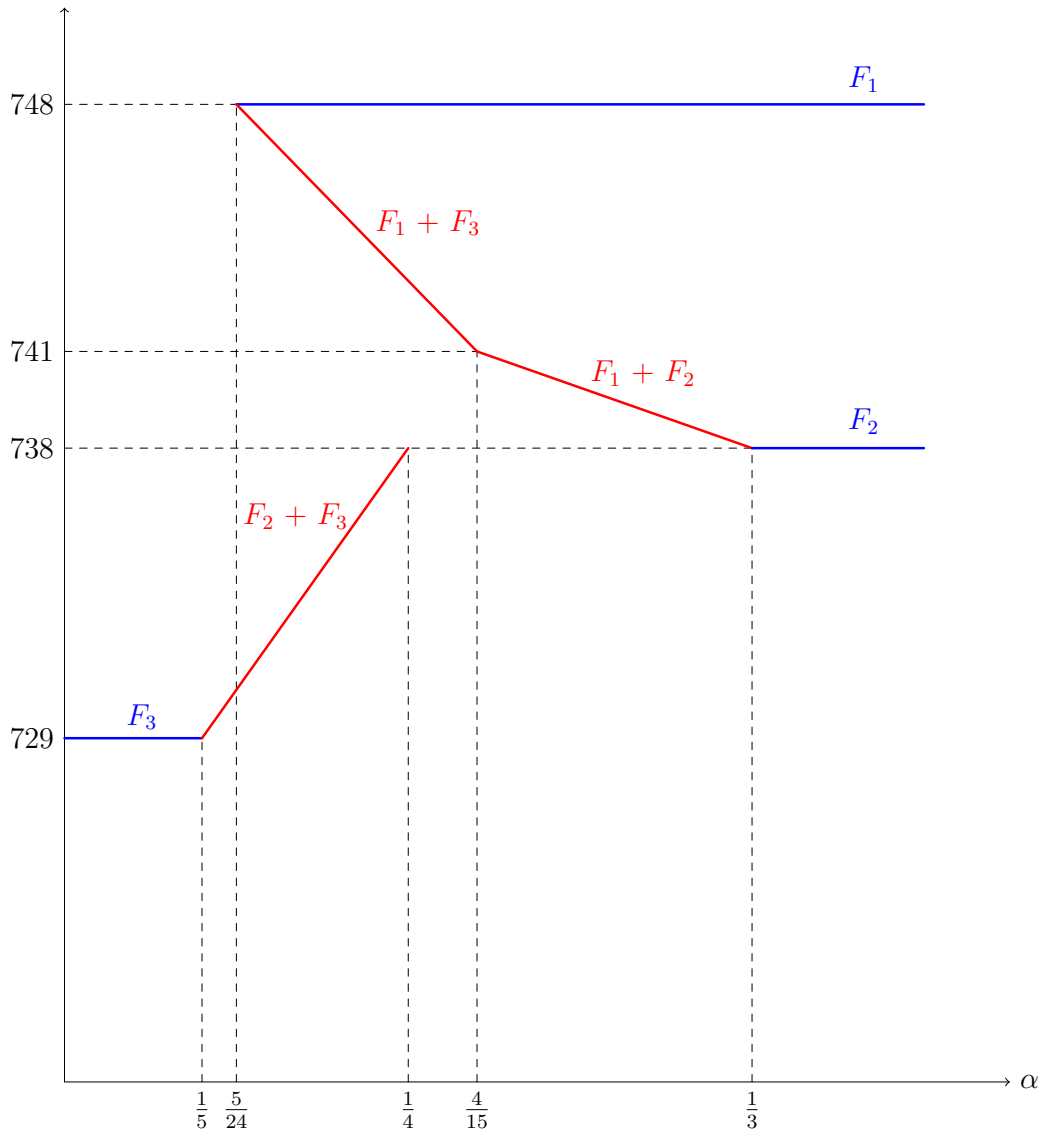


Figure 3: Equilibrium Payoffs

## D Proof of Theorem 9

Project  $F$  corresponds to the distribution that puts probability  $(1 - f_1)(1 - f_2)$  on 0,  $f_1 f_2$  on 2, and  $f_1 + f_2 - 2f_1 f_2$  on 1. We write the probability  $F$  puts on  $i$  successes as  $f(i)$  and analogously for  $G$ . So

$$\int_0^z F(x) dx > \int_0^z G(x) dx$$

for all  $z \in (0, 2)$  iff  $f(0) > g(0)$  and

$$f(0) + (z - 1)[f(0) + f(1)] > g(0) + (z - 1)[g(0) + g(1)], \forall z \in (1, 2). \quad (11)$$

Note that

$$\int_0^2 F(x) dx = 2f(0) + f(1) = 2[1 - f(1) - f(2)] + f(1) = 2 - \mathbb{E}_F(x).$$

Since  $F$  and  $G$  are assumed to have equal means, this implies that we have an equality in equation (11) at  $z = 2$ . Hence equation (11) holds iff  $f(0) > g(0)$ .

Hence  $F$  is strongly riskier than  $G$  if

$$1 - f_1 - f_2 + f_1 f_2 > 1 - g_1 - g_2 + g_1 g_2.$$

By equal means,  $f_1 + f_2 = g_1 + g_2$ , so this holds iff

$$f_1 f_2 > g_1 (f_1 + f_2 - g_1)$$

iff

$$(f_1 - g_1)(f_2 - g_1) > 0.$$

By assumption,  $f_1 > g_1$ , so this holds iff  $f_2 > g_1$ . Given equal means, this holds iff

$$g_2 > \max\{f_1, f_2\} \geq \min\{f_1, f_2\} > g_1.$$

The case where  $G$  is strongly riskier is analogous. ■

## References

- [1] Acharya, V., P. DeMarzo, and I. Kremer, “Endogenous Information Flows and the Clustering of Announcements,” *American Economic Review*, **101**, 2011, 2955–2979.
- [2] Beyer, A., and I. Guttman, “Voluntary Disclosure, Manipulation, and Real Effects,” *Journal of Accounting Research*, **50**, June 2012, 1141–1177.
- [3] Bond, P., and I. Goldstein, “Government Intervention and Information Aggregation by Prices,” *Journal of Finance*, **70**, December 2015, 2777–2812.
- [4] Chen, Y., “Career Concerns and Excessive Risk Taking,” *Journal of Economics and Management Strategy*, **24**, Spring 2015, 110–130.
- [5] DeMarzo, P., I. Kremer, and A. Skrzypacz, “Test Design and Disclosure,” working paper, 2017.
- [6] Diamond, D., and R. Verrecchia, “Disclosure, Liquidity, and the Cost of Capital,” *Journal of Finance*, **46**, September 1991, 1325–1359.
- [7] Dye, R. A., “Disclosure of Nonproprietary Information,” *Journal of Accounting Research*, **23**, 1985a, 123–145.
- [8] Dye, R. A., “Strategic Accounting Choice and the Effect of Alternative Financial Reporting Requirements,” *Journal of Accounting Research*, **23**, 1985b, 544–574.
- [9] Dye, R. A., and S. S. Sridhar, “Resource Allocation Effects of Price Reactions to Disclosures,” *Contemporary Accounting Research*, **19**, Fall 2002, 385–410.
- [10] Edmans, A., M. Heinle, and C. Huang, “The Real Costs of Disclosure,” Wharton working paper, May 2013.
- [11] Fishman, M. and K. Hagerty, “The Optimal Amount of Discretion to Allow in Disclosure,” *Quarterly Journal of Economics*, **105**, May 1990, 427–444.
- [12] Forges, F. and F. Koessler, “Communication Equilibria with Partially Verifiable Types,” *Journal of Mathematical Economics*, **41**, November 2005, 793–811.
- [13] Forges, F. and F. Koessler, “Long Persuasion Games,” *Journal of Economic Theory*, **143**, November 2008, 1–35.
- [14] Fudenberg, D., and J. Tirole, “Perfect Bayesian Equilibrium and Sequential Equilibrium,” *Journal of Economic Theory*, **53**, April 1991, 236–260.
- [15] Gao, P., and P. Liang, “Informational Feedback Effect, Adverse Selection, and the Optimal Disclosure Policy,” *Journal of Accounting Research*, **51**, December 2013, 1133–1158.

- [16] Gigler, F., C. Kanodia, H. Sapiro, and R. Venugopalan, “How Frequent Financial Reporting Can Cause Managerial Short-Termism: An Analysis of the Costs and Benefits of Increasing Reporting Frequency,” University of Minnesota working paper, September 2013.
- [17] Glazer, J., and A. Rubinstein, “On Optimal Rules of Persuasion,” *Econometrica*, **72**, November 2004, 1715–1736.
- [18] Glazer, J., and A. Rubinstein, “A Study in the Pragmatics of Persuasion: A Game Theoretical Approach,” *Theoretical Economics*, **1**, December 2006, 395–410.
- [19] Grossman, S. J., “The Informational Role of Warranties and Private Disclosures about Product Quality,” *Journal of Law and Economics*, **24**, 1981, 461–483.
- [20] Guttman, I., I. Kremer, and A. Skrzypacz, “Not Only What but also When: A Theory of Dynamic Voluntary Disclosure,” *American Economic Review*, **104**, August 2014, 2400–2420.
- [21] Holmstrom, B., “Managerial Incentive Problems: A Dynamic Perspective,” *Review of Economic Studies*, **66**, January 1999, 169–182.
- [22] Jensen, M., and W. Meckling, “Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure,” *Journal of Financial Economics*, **3**, October 1976, 305–360.
- [23] Jung, W., and Y. Kwon, “Disclosure When the Market is Unsure of Information Endowment of Managers,” *Journal of Accounting Research*, **26**, 1988, 146–153.
- [24] Kanodia, C., and A. Mukherji, “Real Effects of Separating Investment and Operating Cash Flows,” *Review of Accounting Studies*, **1**, 1996, 51–71.
- [25] Kanodia, C., and H. Sapiro, “A Real Effects Perspective to Accounting Measurement and Disclosure: Implications and Insights for Future Research,” University of Minnesota working paper, September 2015.
- [26] Kanodia, C., H. Sapiro, and R. Venugopalan, “Should Intangibles Be Measured: What Are the Economic Trade-Offs?” *Journal of Accounting Research*, **42**, March 2004, 89–120.
- [27] Koutsoupias, E., and C. Papadimitriou, “Worst-case Equilibria,” in *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science*, 1999, pp. 404–413.
- [28] Lipman, B., and D. Seppi, “Robust Inference in Communication Games with Partial Provability,” *Journal of Economic Theory*, **66**, August 1995, 370–405.

- [29] Milgrom, P., “Good News and Bad News: Representation Theorems and Applications,” *Bell Journal of Economics*, **12**, 1981, 350–391.
- [30] Mas-Colell, A., M. Whinston, and J. Green, *Microeconomic Theory*, New York: Oxford University Press, 1995.
- [31] Okuno-Fujiwara, M., A. Postlewaite, and K. Suzumura, “Strategic Information Revelation,” *Review of Economic Studies*, **57**, January 1990, 25–47.
- [32] Ostaszewski, A., and M. Gietzmann, “Value Creation with Dye’s Disclosure Option: Optimal Risk–Shielding with an Upper Tailed Disclosure Strategy,” *Review of Quantitative Finance and Accounting*, **31**, July 2008, 1–27.
- [33] Rodina, D., “Information Design and Career Concerns,” Northwestern University working paper, November 2016.
- [34] Roughgarden, T., *Selfish Routing and the Price of Anarchy*, MIT Press, 2005.
- [35] Shin, H. S., “The Burden of Proof in a Game of Persuasion,” *Journal of Economic Theory*, **64**, October 1994, 253–264.
- [36] Shin, H. S., “Disclosures and Asset Returns,” *Econometrica*, **71**, January 2003, 105–133.
- [37] Stein, J., “Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior,” *Quarterly Journal of Economics*, **101**, November 1989, 655–669.
- [38] Verrecchia, R. E., “Discretionary Disclosure,” *Journal of Accounting and Economics*, **5**, 1983, 179–194.
- [39] Wen, X., “Voluntary Disclosure and Investment,” *Contemporary Accounting Research*, **30**, Summer 2013, 677–696.