האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

DISCLOSURE AND CHOICE

By

**ELCHANAN BEN-PORATH, EDDIE DEKEL,
and BARTON L. LIPMAN**

מרכז פדרמן לחקר הרציונליות

THE FEDERMANN CENTER FOR
THE STUDY OF RATIONALITY

Disclosure and Choice[1]

Elchanan Ben-Porath [2]   Eddie Dekel [3]   Barton L. Lipman[4]

First Draft
December 2014

[2]Department of Economics and Center for Rationality, Hebrew University. Email: benporat@math.huji.ac.il.

[3]Economics Department, Northwestern University, and School of Economics, Tel Aviv University. Email: dekel@northwestern.edu.

[4]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 can disclose the true outcome to the observer. We show that choice is inefficient: the agent favors riskier projects even with lower expected returns. If information can 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.

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 of the choice 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 has an incentive to 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. 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 always has an incentive to try to choose a project that looks better than it is. In equilibrium, though, the observer cannot be fooled, so the agent simply hurts himself. In particular, he would be better off if he could commit to never disclosing anything or to any other "nonstrategic"[1] disclosure policy. We refer to the outcome that gives the best possible payoff to the agent as the *first best* and show that this is the outcome when the agent cannot affect disclosure. We also show that the agent's utility loss relative to the first best 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. Note that the manager's utility function can be identical to that of the stockholders in the firm, so the inefficiency we identify is not due to a standard moral hazard problem. Here the first–best project is the one which maximizes the expected value of the firm.

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

---

[1]By "nonstrategic," we mean any policy where the probability that information is disclosed is independent of the information being disclosed.

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 will 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.

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 may 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 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.

We will show that in the extreme case where all information comes from 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 minimize 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 information. 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.[2]

In all cases, we also characterize the worst possible equilibrium payoff for the agent relative to the first–best payoff. For example, we show that the agent's payoff can be as

---

[2]If information is "close" to balanced, then production decisions are "close" to the first best.

low as 50% of the first–best payoff but cannot be any lower.

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 model. 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. Section 7 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 consider a model where an agent wishes to persuade an observer, but only through disclosure — the agent does not 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 (2013) 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 use 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. 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 (2014), or Gao and Liang (2013)). While these do generate costs which can have effects on the firm's productive investments, they are very different sources of costs than the incentive effects we focus on.

A few other papers consider incentive effects. A number of these follow 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 the effect on the manager's diversion of cash flows or other investment distortions. See, for example, Kanodia and Mukherji (1996), Kanodia, Sapra, and Venugopalan (2004), Edmans, Heinle, and Huang (2013), or Gigler, Kanodia, Sapra, and Venugopalan (2013). The short–termism effect explored in these papers is similar to the inefficiency we consider in that both approaches consider how the manager of a firm can manipulate what the market observes. There are two key differences between the approaches. First, the inefficiencies we identify are driven by strategic disclosure, a consideration unrelated to the short–termism results. Second, our results are on inefficiencies in the riskiness of investment while the short–termism literature concerns inefficiencies in its timing.[3]

A different approach to incentive effects is taken by Beyer and Guttmann (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.

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

---

[3]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).

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, the one he would commit himself to if he could. 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 $(1-q_1)(4)+q_1[(1/2)(4)+(1/2)(6)] > 4$. 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.[4] 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 3 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).

---

[4]If q_1 is larger, the unique equilibrium is mixed.

3   Model

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

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.[5]

In the second stage, there is a random determination of whether the agent or challenger has evidence demonstrating the outcome of the project. 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.[6] If a player has evidence, then he can either present it, demonstrating conclusively what the outcome of the project is, 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.[7] 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.[8] 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

---

[5]The assumption that $\mathcal{F}$ is finite is a simple way to ensure equilibrium existence. It is not difficult to allow for unbounded supports as long as all relevant expectations exist.

[6]As shown in Section 6, our results do not rely on the first of these independence assumptions. We use it only for notational convenience.

[7]As will be clear from the analysis, the results also hold if the players move sequentially.

[8]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.

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 $\alpha \in [0, 1]$. 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 α, 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.[9]

In the remainder of this section, we do the following. First, we discuss the benchmark case where the information seen by the observer cannot be affected by the agent or challenger — where it is entirely exogenous. As we will show, this case generates the first–best outcome, which is the outcome which maximizes the agent's expected payoff over all feasible projects. Second, we discuss the structure of equilibria in this game more generally to set up our detailed discussion of the inefficiencies of equilibria in the following sections.

3.1 Benchmark

First, 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 $q \in [0, 1]$ and that the agent and challenger cannot affect whether the observer sees this information.

Except for the degenerate case where $\alpha = q = 0$, the optimal project choice by the agent is any project F which maximizes $\mathrm{E}_F(x)$ where E_F denotes the expectation with respect to the distribution F. We refer to such a project F as a *first–best project*.

To see why the agent chooses a first–best project, fix an equilibrium. 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 is

$$\alpha \mathrm{E}_F(x) + (1 - \alpha) \left[q \mathrm{E}_F(x) + (1 - q)\hat{x} \right].$$

Obviously, if $\alpha = q = 0$, then the agent's payoff is $\hat{x}$, regardless of the F he chooses, so he is indifferent over all projects. Otherwise, his payoff is maximized by maximizing $\mathrm{E}_F(x)$. To be more precise, choosing any such F strictly dominates choosing any project with a strictly lower expectation. (The degeneracy of the case where $\alpha = q = 0$ will appear again below.)

---

[9]Specifically, we use what Mas-Colell, Whinston, and Green (1995) refer to as weak perfect Bayesian equilibrium. Our results would not change if we used a stronger notion such as Kreps and Wilson's (1982) sequential equilibrium.

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. Hence he ends up choosing a project with a lower expected value and is worse off as a result.

3.2   Equilibrium

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 σ 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. First, suppose the agent obtains proof that the outcome is x where $x > \hat{x}$. In this case, the agent will disclose the outcome in any equilibrium, regardless of the strategy of the challenger. 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. Similar reasoning shows that if the challenger obtains proof that the outcome is x where $x < \hat{x}$, then the challenger discloses this with probability 1.

So suppose the agent obtains proof that the outcome is $x < \hat{x}$. Similar reasoning to the above shows that he hides this information in equilibrium except in the trivial case where $q_2 = 1$. When $q_2 = 1$, the challenger will necessarily also have this information. From the above, we know the challenger will disclose it. Hence in this case, the agent is indifferent between disclosing and not. In short, if the agent's disclosure decision matters, then he does not disclose in this situation. For simplicity, we simply focus on the equilibrium where the agent never discloses when $x < \hat{x}$. Similar reasoning shows that we can also assume without loss of generality that the challenger never discloses when $x > \hat{x}$.

One can show that the equilibrium is entirely unaffected by the disclosure choices when $x = \hat{x}$, so for simplicity we assume both the agent and challenger disclose in this

situation.[10]

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

$$V_A(F, \hat{x}) = \alpha E_F(x) + (1 - \alpha)\big[(1 - q_1)(1 - q_2)\hat{x} + q_1(1 - q_2)E_F \max\{x, \hat{x}\} \qquad (1)$$
$$+ q_2(1 - q_1)E_F \min\{x, \hat{x}\} + q_1 q_2 E_F(x)\big].$$

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 σ, $\hat{x}$ 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 σ. The most convenient way to state this is to use the law of iterated expectations to write it as

$$\sum_{F \in \mathcal{F}} \sigma(F) E_F(x) = \sum_{F \in \mathcal{F}} \sigma(F)\big[(1 - q_1)(1 - q_2)\hat{x} + q_1(1 - q_2)E_F \max\{x, \hat{x}\} \qquad (2)$$
$$+ q_2(1 - q_1)E_F \min\{x, \hat{x}\} + q_1 q_2 E_F(x)\big].$$

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.

Equation (2) implies that the agent's equilibrium expected payoff is $\sum_{F \in \mathcal{F}} \sigma(F) E_F(x)$. 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.

---

[10]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 player 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.

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 reduces 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

$$V_A(F, \hat{x}) = \alpha \mathrm{E}_F(x) + (1 - \alpha)[(1 - q_1)\hat{x} + q_1 \mathrm{E}_F \max\{x, \hat{x}\}]. \tag{3}$$

Thus the agent chooses the project F to maximize $\mathrm{E}_F[\alpha x + (1 - \alpha)q_1 \max\{x, \hat{x}\}]$ 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 domination, 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 $\mathrm{E}_F(x) = \mathrm{E}_G(x)$.

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. 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.

Our first result on risk taking uses a stronger notion of riskier.

Definition 2. *Given two distributions F and G over $[a, b]$, G strongly dominates F in the sense of second–order stochastic domination, 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 $\mathrm{E}_F(x) = \mathrm{E}_G(x)$.

10

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 u is convex and not linear, while G yields strictly higher expected utility than F if u is concave and not linear.

Theorem 1. *Suppose $q_2 = 0$. Suppose there are distributions $F, G \in \mathcal{F}$ such that F is strongly riskier than G. Then if $\alpha < 1$ and $q_1 \in (0,1)$, there is no pure strategy equilibrium in which the agent chooses G.*[11]

Proof. Suppose to the contrary that it is a pure equilibrium for the agent to choose G. Then the payoff to G must exceed the payoff to F. Using equation (3), this implies

$$\alpha \mathrm{E}_G(x) + (1 - \alpha)q_1 \mathrm{E}_G \max\{x, \hat{x}\} \geq \alpha \mathrm{E}_F(x) + (1 - \alpha)q_1 \mathrm{E}_F \max\{x, \hat{x}\}.$$

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

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

Note that

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

Integration by parts shows that

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

so

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

Hence we must have

$$\mathrm{E}_G(x) + \int_0^{\hat{x}} G(x) \, dx \geq \mathrm{E}_F(x) + \int_0^{\hat{x}} F(x) \, dx.$$

Again, since F is strongly riskier than G, we have $\mathrm{E}_G(x) = \mathrm{E}_F(x)$ implying

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

Since F is strongly riskier than G, this implies that $\hat{x} \notin (a, b)$. Hence, in particular, either $\hat{x}$ is strictly outside the support of G or is either the upper or lower bound of the support.

---

[11]It is worth noting that 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.

From equation (2), $\hat{x}$ must satisfy

$$\mathrm{E}_G(x) = (1 - q_1)\hat{x} + q_1 \mathrm{E}_G \max\{x, \hat{x}\}.$$

If $\hat{x}$ is less than or equal to the lower bound of the support of G, then this equation says $\mathrm{E}_G(x) = (1 - q_1)\hat{x} + q_1 \mathrm{E}_G(x)$. Hence, given the assumption that $q_1 < 1$, we obtain $\hat{x} = \mathrm{E}_G(x)$, a contradiction unless G is degenerate at $\hat{x}$. If $\hat{x}$ is greater than or equal to the upper bound of the support of G, then this equation says $\mathrm{E}_G(x) = \hat{x}$, again a contradiction unless G is degenerate at $\hat{x}$.

So suppose G is degenerate at $\hat{x}$. Since $F \neq G$, F cannot be degenerate at $\hat{x}$. Since $\mathrm{E}_F(x) = \mathrm{E}_G(x)$, $\hat{x}$ must be in the interior of the support of F. But then $\hat{x} \in (a, b)$, a contradiction. ∎

Our next result uses weaker hypotheses — comparing distributions using riskiness rather than strong riskiness, allowing mixed equilibria, and not imposing parameter restrictions other than non–degeneracy on α and q_1. Consequently, the conclusion is weaker as well. Specifically, we show that if there are two distributions in $\mathcal{F}$ which can be compared in terms of riskiness, the agent never chooses the less risky of the two if the difference is ever relevant. To understand this result, note that the agent's objective function is piecewise linear, not strictly convex. Hence there are certain comparisons of lotteries where the difference in risk is irrelevant to the agent. To make this last part of the statement precise requires another definition.

Definition 3. *F is equilibrium–indifferent to G if for every equilibrium in which G receives positive probability, there is another equilibrium in which the agent's mixed strategy is unchanged except the probability he played G previously is now moved to F, the observer's strategy is unchanged, and the agent's expected payoff is unchanged.*

It is not hard to show that if F is equilibrium–indifferent to G, then $\lambda F + (1 - \lambda)G$ is also equilibrium–indifferent to G for all $\lambda \in (0, 1)$. In other words, if F is equilibrium–indifferent to G, then the agent makes no distinction between F and G whatsoever and the observer's behavior makes no distinction.

Theorem 2. *Suppose $q_2 = 0$. Suppose there are distributions $F, G \in \mathcal{F}$ such that F is riskier than G. Then for any equilibrium of any nondegenerate game, either the agent puts zero probability on G or else F is equilibrium–indifferent to G.*

Proof. Fix distributions F and G with F riskier than G. Fix an equilibrium in which G is in the support of the agent's mixed strategy and define $\hat{x}$ to be the observer's response if no evidence is presented in the equilibrium. Since G is given positive probability, we must have

$$\alpha \mathrm{E}_G(x) + (1 - \alpha)q_1 \mathrm{E}_G \max\{x, \hat{x}\} \geq \alpha \mathrm{E}_F(x) + (1 - \alpha)q_1 \mathrm{E}_F \max\{x, \hat{x}\}.$$

Since the game is nondegenerate, either $\alpha > 0$ or $q_1 > 0$ or both. Hence if $(1 - \alpha)q_1 = 0$, we must have $\alpha > 0$ so the agent maximizes $\mathrm{E}_F(x)$ and the result follows. So for the remainder, assume $(1 - \alpha)q_1 > 0$. The same integration by parts and the same substitutions as used in the proof of Theorem 1 imply

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

F riskier than G implies the reverse weak inequality, so

$$\int_0^{\hat{x}} G(x)\,dx = \int_0^{\hat{x}} F(x)\,dx,$$

implying

$$\mathrm{E}_G \max\{x, \hat{x}\} = \mathrm{E}_F \max\{x, \hat{x}\}.$$

Change the agent's strategy by switching the probability he plays G to playing F. If $\hat{x}$ does not change, his new strategy is still a best reply. From equation (2), the appropriate $\hat{x}$ can be defined by

$$(1 - q_1)\hat{x} + q_1 \sum_{F' \in \mathcal{F}} \sigma(F')\mathrm{E}_{F'} \max\{x, \hat{x}\} = \sum_{F' \in \mathcal{F}} \sigma(F')\mathrm{E}_{F'}(x).$$

Since $\mathrm{E}_F(x) = \mathrm{E}_G(x)$ and $\mathrm{E}_F \max\{x, \hat{x}\} = \mathrm{E}_G \max\{x, \hat{x}\}$, we see that $\hat{x}$ does not change. Hence this is an equilibrium. Clearly, the agent obtains the same expected payoff. So F is equilibrium–indifferent to G. ∎

Theorems 1 and 2 compare distributions with the same means, but it is easy to see that, in general, the agent will accept a lower mean in order to obtain more risk.

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 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 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

13

arbitrarily large. Below, we give tight lower bounds on the ratio of the agent's equilibrium payoff to his best feasible payoff. 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.[12] More precisely, given a game $(\mathcal{F}, \alpha, q_1, q_2)$, let

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

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 will give 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.

Specifically, we have the following result.[13]

Theorem 3. *For any nondegenerate game, we have*

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

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

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

---

[12]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).

[13]The exact statements of the lower bounds in Theorems 3 and 5 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. We thank Bruno Strulovici for raising this issue.

We offer several comments on this result. First, there is a discontinuity in the function R at the degenerate case where $\alpha = q_1 = q_2 = 0$. To see this, note from the characterization of the minimum over q_1 that $R(\alpha, q_1, 0) \geq 1/2$ if $\alpha > 0$, but $R(0, 0, 0) = 0$. To understand this discontinuity, note that when $q_1 = q_2 = 0$, there is no information that will be revealed to the observer at stage 2. When $\alpha = 0$, the only thing the agent cares about is the observer's belief. Since no information will be revealed to the observer, the agent cannot do anything to affect the only thing he cares about. In particular, 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. Consequently, our remaining remarks focus on the nondegenerate case.

Second, it is easy to see that $R(\alpha, q_1, 0)$ is increasing in α 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 α, the value of q_1 which minimizes the bound is $q_1 = \sqrt{\alpha}/[1 + \sqrt{\alpha}]$, which is interior for any $\alpha \in (0, 1)$.

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 3. It shows $R(\alpha, q_1, 0)$ as a function of q_1 for various values of α.

The proof of Theorem 3 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 must be at least half the first–best in any nondegenerate game. That is, we prove the last statement of the theorem for $\alpha = 0$.

So fix any feasible set of projects $\mathcal{F}$ and any $q_1 \in (0, 1]$. Fix any equilibrium mixed strategy σ for the agent and any project F in the support of σ which has the lowest expected value of x across projects in the support. Fix the $\hat{x}$ of the equilibrium. Let G be any first–best project. As seen in the proof of Theorems 1 and 2, $q_1 > 0$, F in the support of σ, and the optimality of σ imply

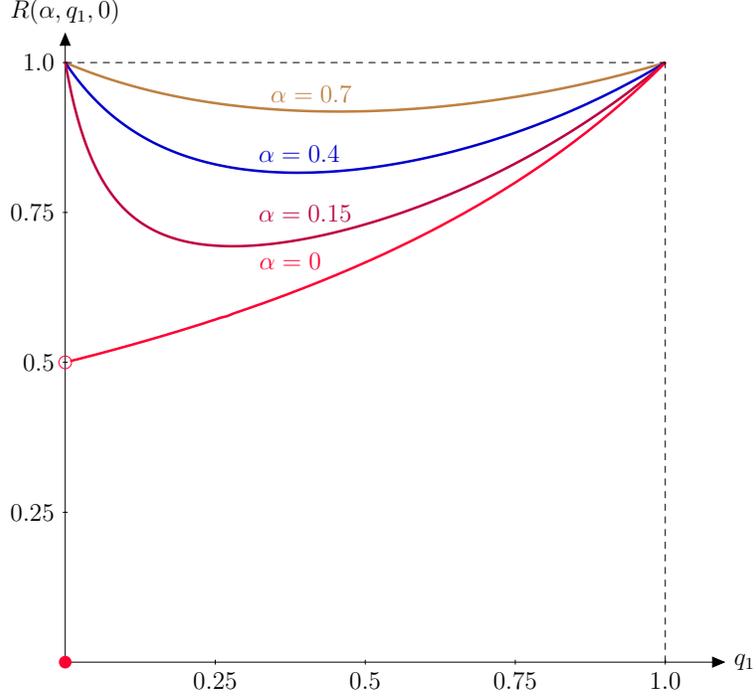$$\mathrm{E}_F \max\{x, \hat{x}\} \geq \mathrm{E}_G \max\{x, \hat{x}\}$$

Figure 1: "Worst Case" as a Function of q_1.

or

$$\mathrm{E}_F(x) + \int_0^{\hat{x}} F(x)\, dx \geq \mathrm{E}_G(x) + \int_0^{\hat{x}} G(x)\, dx.$$

Since $F(x) \leq 1$ and $G(x) \geq 0$, this requires

$$\mathrm{E}_F(x) + \hat{x} \geq \mathrm{E}_G(x). \tag{4}$$

From equation (2),

$$\sum_{F' \in \mathcal{F}} \sigma(F') \mathrm{E}_{F'}(x) = (1 - q_1)\hat{x} + q_1 \sum_{F' \in \mathcal{F}} \sigma(F') \mathrm{E}_{F'} \max\{x, \hat{x}\}.$$

Since $\mathrm{E}_{F'} \max\{x, \hat{x}\} \geq \mathrm{E}_{F'}(x)$, we see that

$$\sum_{F' \in \mathcal{F}} \sigma(F') \mathrm{E}_{F'}(x) \geq \hat{x}.$$

Also, by our assumption that F is one of the projects with the lowest mean in the support, we have

$$\sum_{F' \in \mathcal{F}} \sigma(F') \mathrm{E}_{F'}(x) \geq \mathrm{E}_F(x).$$

Hence equation (4) implies

$$2 \sum_{F' \in \mathcal{F}} \sigma(F') \mathrm{E}_{F'}(x) \geq \mathrm{E}_F(x) + \hat{x} \geq \mathrm{E}_G(x).$$

16

So the agent's payoff $\sum_{F' \in \mathcal{F}} \sigma(F') E_{F'}(x)$ 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 $E_F(x) = 1$. 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

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

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

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

It is easy to show 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.

The implication of Theorem 3 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 α increases for fixed $\mathcal{F}$. In the example, there is a mixed strategy equilibrium with payoffs that are decreasing in α. 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 α. Finally, the payoff in the worst pure strategy equilibrium is also decreasing in α for a certain range, showing that this result is not an artifact related to mixed equilibria.

17

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

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

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 $(1-\alpha)q_2 > 0$, this function is concave, so the agent's choices are effectively risk averse. This gives the following analog to Theorem 2.[14]

Theorem 4. *Suppose $q_1 = 0$. Suppose there are distributions $F, G \in \mathcal{F}$ such that F is riskier than G. Then for any equilibrium of any nondegenerate game, either the agent puts zero probability on F or else G is equilibrium–indifferent to F.*

Proof. Fix F and G as above and suppose we have an equilibrium where F is in the support of the agent's mixed strategy. Clearly, we must have

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

Since the game is nondegenerate, either $\alpha > 0$ or $q_2 > 0$ or both. If $(1-\alpha)q_2 = 0$, the agent chooses the project to maximize $E_F(x)$ and the result holds. So assume $(1 - \alpha)q_2 > 0$. Since F is riskier than G, we have $E_F(x) = E_G(x)$. Hence,

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

Note that
$$\min\{a, b\} + \max\{a, b\} = a + b,$$

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

Hence we must have

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

We showed earlier that $E_F \max\{x, \hat{x}\} = E_F(x) + \int_0^{\hat{x}} F(x)\, dx$. Substituting,

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

---

[14]It is straightforward to give an analog for Theorem 1 as well.

Since F is riskier than G, this must hold with equality. Change the agent's strategy by switching the probability he plays F to playing G. If $\hat{x}$ does not change, his new strategy is still a best reply. From equation (2), $\hat{x}$ can be defined by

$$(1 - q_2)\hat{x} + q_2 \sum_{F' \in \mathcal{F}} \sigma(F') \mathrm{E}_{F'} \min\{x, \hat{x}\} = \sum_{F' \in \mathrm{supp}(\sigma)} \sigma(F') \mathrm{E}_{F'}(x).$$

Since $\mathrm{E}_F(x) = \mathrm{E}_G(x)$ and $\mathrm{E}_F \min\{x, \hat{x}\} = \mathrm{E}_G \min\{x, \hat{x}\}$, we see that $\hat{x}$ does not change. Hence this is an equilibrium. Clearly, the agent obtains the same expected payoff. ∎

We can also characterize R for this case. More specifically, we have the following analog to Theorem 3:

Theorem 5. *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 α as used in Figure 1.

Theorem 5 has some features in common with Theorem 3. 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 α increases.

On the other hand, this result also shows several differences from Theorem 3. First, while there is a discontinuity, it is somewhat different from the discontinuity noted previously. To be specific, for $q_1 > 0$,

$$R(0, q_1, 0) = \frac{1}{2 - q_1} \geq \frac{1}{2},$$

while $R(0, 0, 0) = 0$. Here we see that

$$R(0, 0, q_2) = 0$$

for all q_2. Hence the function $R(0, q_1, 0)$ is discontinuous in q_1 at $q_1 = 0$ but the function $R(0, 0, q_2)$ is not discontinuous in q_2 at $q_2 = 0$. On the other hand, there is a discontinuity at $(0, 0, 0)$ since $R(\alpha, 0, 0) = 1$ for all $\alpha > 0$.
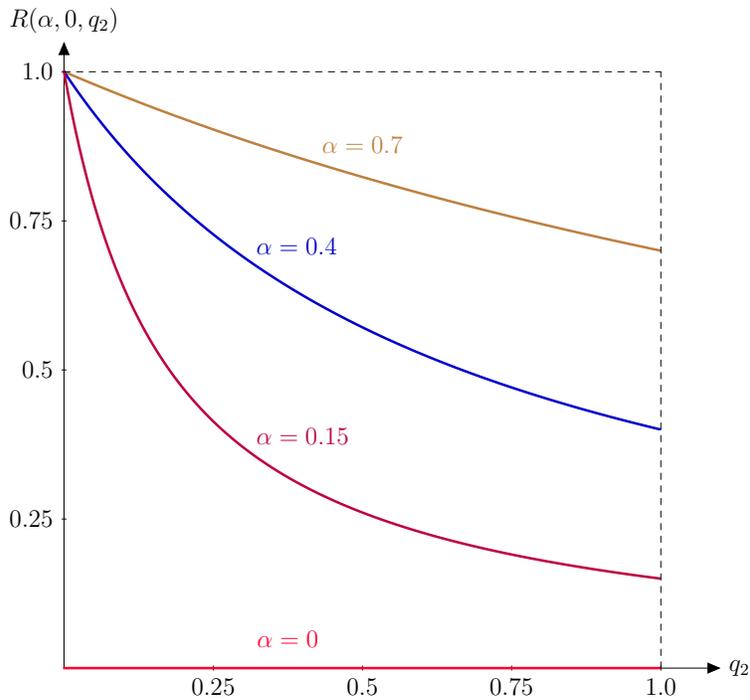
19

Figure 2: "Worst Case" as a Function of q_2.

Second, this result implies that the worst case over α when $q_1 > 0$ and $q_2 = 0$ is better than the worst case when $q_1 = 0$ and $q_2 > 0$. In the former case, we have $R = 1/2$, while in the latter, we have $R = 0$. Since the lower bound is zero and payoffs are non–negative, this implies that in the case where only the challenger speaks, the agent could be arbitrarily worse off than at the first–best.

Third, recall that for $\alpha \in (0,1)$, the worst case payoff in Theorem 3 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 expects, it is as if the observer always saw the true x and so the outcome would seem to necessarily be first–best.

To understand why we do not necessarily obtain the first best at $q_2 = 1$, consider the following example. Suppose $q_2 = 1$ and that $\mathcal{F} = \{F, G\}$ where F gives 1 with probability $1/2$ and 3 with probability $1/2$, 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 if α is not too large. To see this, consider the case where $\alpha = 0$. Suppose F is the project the observer expects the agent to choose. Then if the

20

challenger presents no evidence, the observer believes the outcome to have been 3 since this is the worst possible outcome for the challenger under F. Because of this, the agent has no incentive to deviate to G. If he does deviates 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 3. Thus the agent's expected payoff to the deviation is $(1/2)(0) + (1/2)(3) < 2$, so the agent prefers F. For small enough $\alpha > 0$, the same conclusion will follow.

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. We do not necessarily get unraveling if the agent deviates to an unexpected project and this creates the possibility of inefficient equilibria.

On the other hand, the efficient outcome is also an equilibrium.[15]

Theorem 6. *For any α, if $q_1 = 0$ and $q_2 = 1$, then there is an equilibrium in which the agent chooses the first–best project.*

The proof of this is straightforward. Suppose the agent is expected to choose F where $\mathrm{E}_F(x) \geq \mathrm{E}_G(x)$ for all $G \in \mathcal{F}$. 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 $\mathrm{E}_F(x)$. If he deviates to any other feasible project G, his expected payoff is

$$\alpha \mathrm{E}_G(x) + (1 - \alpha)\mathrm{E}_G \min\{x, x^*\} \leq \mathrm{E}_G(x) \leq \mathrm{E}_F(x).$$

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.

---

[15]It is also worth noting that the efficient outcome is the only equilibrium if all projects have the same support. We thank Georgy Egorov for pointing this out.

Theorem 7. *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 7 implies the corollary, suppose we have a nondegenerate game, so it is *not* the case that $\alpha = q_1 = q_2 = 0$. By Theorem 7, 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 3, the outcome must be first–best in this case.

Proof of Theorem 7. 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

$$\alpha \mathrm{E}_F(x) + (1 - \alpha)\big[(1 - q_1)(1 - q_2)\hat{x} + q_2(1 - q_1)\mathrm{E}_F \min\{x, \hat{x}\}$$
$$+ q_1(1 - q_2)\mathrm{E}_F \max\{x, \hat{x}\} + q_1 q_2 \mathrm{E}_F(x)\big].$$

Note that

$$\mathrm{E}_F \min\{x, \hat{x}\} + \mathrm{E}_F \max\{x, \hat{x}\} = \mathrm{E}_F\left[\min\{x, \hat{x}\} + \max\{x, \hat{x}\}\right] = \mathrm{E}_F(x) + \hat{x}.$$

Hence

$$\mathrm{E}_F \min\{x, \hat{x}\} = \mathrm{E}_F(x) + \hat{x} - \mathrm{E}_F \max\{x, \hat{x}\}. \tag{5}$$

Substituting, we can rewrite the agent's payoff as

$$[\alpha + (1 - \alpha)q_2]\mathrm{E}_F(x) + (1 - \alpha)\left[(1 - q_1)\hat{x} + (q_1 - q_2)\mathrm{E}_F \max\{x, \hat{x}\}\right]. \tag{6}$$

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

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

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

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

22

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

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

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

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

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

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

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, so the agent's payoff as a function of F and $\hat{x}$ is now

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

If we again substitute from equation (5), we obtain

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

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 (6) and complete the proof exactly as above.

We can use Theorem 7 to extend Theorems 3 and 5 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

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

23

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 (α, 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 7 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 α 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   Discussion

The simplicity of the model implies that many important factors are omitted, but also suggests that the force pointed to here is very basic. As stressed in the introduction, the key observation is that to the extent that the agent can control the flow of information, he has incentives to take excessive risks 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 concluding section, we briefly discuss some of the potentially interesting factors omitted here. First, we have omitted 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 α 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 Σ, 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 $\psi(\cdot \mid x)$ and this distribution has full support on Σ 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, then the discontinuity at zero noise disappears. To see this, think of the observer's belief about the project chosen by the agent as giving the observer's prior belief over x. 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."[16] 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.

A second simplification is our assumption that all projects are equally "transparent" in the sense that the probability that the agent or challenger receives evidence to disclose is independent of the project chosen by the agent. It is not difficult to give analogs of our main results for a model where the probability with which evidence is received varies with the project. For example, it is easy to show that 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 Theorems 1 and 2. One can also give worst case results analogous to Theorem 3. For example, it is not difficult to show that if q can vary across projects, then for any set of feasible projects and any α, the agent's

---

[16]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.

equilibrium payoff must be at least α times the first–best payoff and that this lower bound is approximately achievable. Both of these results are stated more formally and proved in Appendix D.

A different way to think about variation in the transparency of projects is to suppose 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. 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.

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,[17] there are good reasons for expecting managerial compensation to depend positively on both short and long run stock prices. First, if the long run is indeed long, the manager requires compensation in the short run too. Given 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.[18] 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

---

[17]See, for example, the discussion in Stein (1989) or Edmans, Heinle, and Huang (2013).

[18]This formulation is common in the literature. See, for example, Diamond and Verrecchia (1991) or Gigler, Kanodia, Sapra, and Venugopalan (2013).

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.

Appendix

A Proof of Theorem 3

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 $\mathrm{E}_F(x)$, 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

$$\alpha \mathrm{E}_F(x) + (1 - \alpha)[(1 - q_1)\hat{x} + q_1 \mathrm{E}_F \max\{x, \hat{x}\}]. \tag{7}$$

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

$$(\alpha + (1 - \alpha)q_1)\mathrm{E}_F(x) + (1 - \alpha)(1 - q_1)\hat{x} + (1 - \alpha)q_1 \int_0^{\hat{x}} F(x)\, dx.$$

Fix an equilibrium mixed strategy for the agent σ and the associated $\hat{x}$. Let $U = \sum_{F' \in \mathcal{F}} \sigma(F')\mathrm{E}_{F'}(x)$, 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 $\mathrm{E}_F(x) \leq U$ and let G be any other feasible project. Then we must have

$$(\alpha + (1 - \alpha)q_1)\mathrm{E}_G(x) + (1 - \alpha)q_1 \int_0^{\hat{x}} G(x)\, dx$$
$$\leq (\alpha + (1 - \alpha)q_1)\mathrm{E}_F(x) + (1 - \alpha)q_1 \int_0^{\hat{x}} F(x)\, dx.$$

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

$$(\alpha + (1 - \alpha)q_1)\mathrm{E}_G(x) \leq (\alpha + (1 - \alpha)q_1)\mathrm{E}_F(x) + (1 - \alpha)q_1 \int_0^{\hat{x}} F(x)\, dx.$$

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.[19] Since $F(x) \in [0,1]$, we must have $z \in [0,1]$. Then we can rewrite this equation as

$$(\alpha + (1 - \alpha)q_1)\mathrm{E}_G(x) \le (\alpha + (1 - \alpha)q_1)\mathrm{E}_F(x) + (1 - \alpha)q_1 z\hat{x}. \qquad (8)$$

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

$$(\alpha + (1 - \alpha)q_1)\mathrm{E}_F(x) + (1 - \alpha)(1 - q_1)\hat{x} + (1 - \alpha)q_1 z\hat{x} = U,$$

so

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

Substituting into equation (8) gives

$$(\alpha + (1 - \alpha)q_1)\mathrm{E}_G(x) \le (\alpha + (1 - \alpha)q_1)\mathrm{E}_F(x) + q_1 z \left[\frac{U - (\alpha + (1 - \alpha)q_1)\mathrm{E}_F(x)}{1 - q_1 + zq_1}\right]. \qquad (9)$$

Recall that $U \ge \mathrm{E}_F(x)$, so $U \ge (\alpha + (1 - \alpha)q_1)\mathrm{E}_F(x)$. Hence $q_1 \ge 0$ implies that the right–hand side is weakly increasing in z. Hence

$$(\alpha + (1 - \alpha)q_1)\mathrm{E}_G(x) \le (\alpha + (1 - \alpha)q_1)\mathrm{E}_F(x) + q_1 \left[U - (\alpha + (1 - \alpha)q_1)\mathrm{E}_F(x)\right]$$

or

$$[\alpha + (1 - \alpha)q_1]\mathrm{E}_G(x) \le Uq_1 + \mathrm{E}_F(x)(\alpha + (1 - \alpha)q_1)(1 - q_1).$$

Since the term multiplying $\mathrm{E}_F(x)$ is positive, the fact that $\mathrm{E}_F(x) \le U$ implies

$$(\alpha + (1 - \alpha)q_1)\mathrm{E}_G(x) \le U\left[q_1 + (\alpha + (1 - \alpha)q_1)(1 - q_1)\right].$$

Hence, taking G to be a first–best project,

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

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

---

[19]To see this, suppose $\hat{x} = 0$. Then equation (2) implies that either $q_1 = 1$ or $\mathrm{E}_F(x) = 0$ 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 (8) 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 (9), 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 5

Again, nondegeneracy implies that either $\alpha > 0$ or $q_2 > 0$ or both. Just as in the proof of Theorem 3, 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 σ 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.[20] As in the proof of Theorem 3, $F(x) \in [0, 1]$ implies $z \in [0, 1]$. Then we can rewrite this equation as

$$\alpha E_G(x) \leq \alpha E_F(x) + (1 - \alpha)q_2 z \hat{x}. \tag{10}$$

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

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

---

[20]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.

31

so

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

Substituting into equation (10) gives

$$\alpha \mathrm{E}_G(x) \leq \alpha \mathrm{E}_F(x) + q_2 z \left[\frac{U - \alpha \mathrm{E}_F(x)}{1 - q_2 + zq_2} \right]. \tag{11}$$

By assumption, $U \geq \mathrm{E}_F(x)$, so $U \geq \alpha \mathrm{E}_F(x)$. Hence the right–hand side is weakly increasing in z, so this implies

$$\alpha \mathrm{E}_G(x) \leq \alpha \mathrm{E}_F(x) + q_2 \left[U - \alpha \mathrm{E}_F(x) \right]$$

or

$$\alpha \mathrm{E}_G(x) \leq q_2 U + \alpha(1 - q_2)\mathrm{E}_F(x) \leq U[q_2 + \alpha(1 - q_2)] = U[\alpha + (1 - \alpha)q_2].$$

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 $\mathrm{E}_G(x) = y$. 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

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

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. ∎

32

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" ℓ_i otherwise. The specific values of h_i, ℓ_i, and p_i are given in the table below.

	h_i	ℓ_i	p_i	μ_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, $\mu_i = \mathrm{E}_{F_i}(x)$. Note that F_1 is the first–best project, F_2 is second best, and F_3 worst. Simple calculations show the range of α'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 α's where it is chosen in equilibrium. Similarly, for each pair of projects, there is a nonempty range of α'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 α. On the other hand, the equilibrium payoff when the agent randomizes between F_2 and F_3 is increasing in α.

To see the intuition, consider the case where the agent randomizes between F_1 and F_2. As α 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 α. Note that, as asserted, the equilibrium payoffs for two of the three mixed strategy equilibria are decreasing in α. Note also that the payoff to the worst equilibrium is decreasing in α for α between 1/4 and 1/3. Finally, note that if we focus only on pure strategy equilibria, the worst equilibrium payoff is decreasing in α as we move from the range where $\alpha \in [5/24, 1/3]$ to $\alpha > 1/3$.
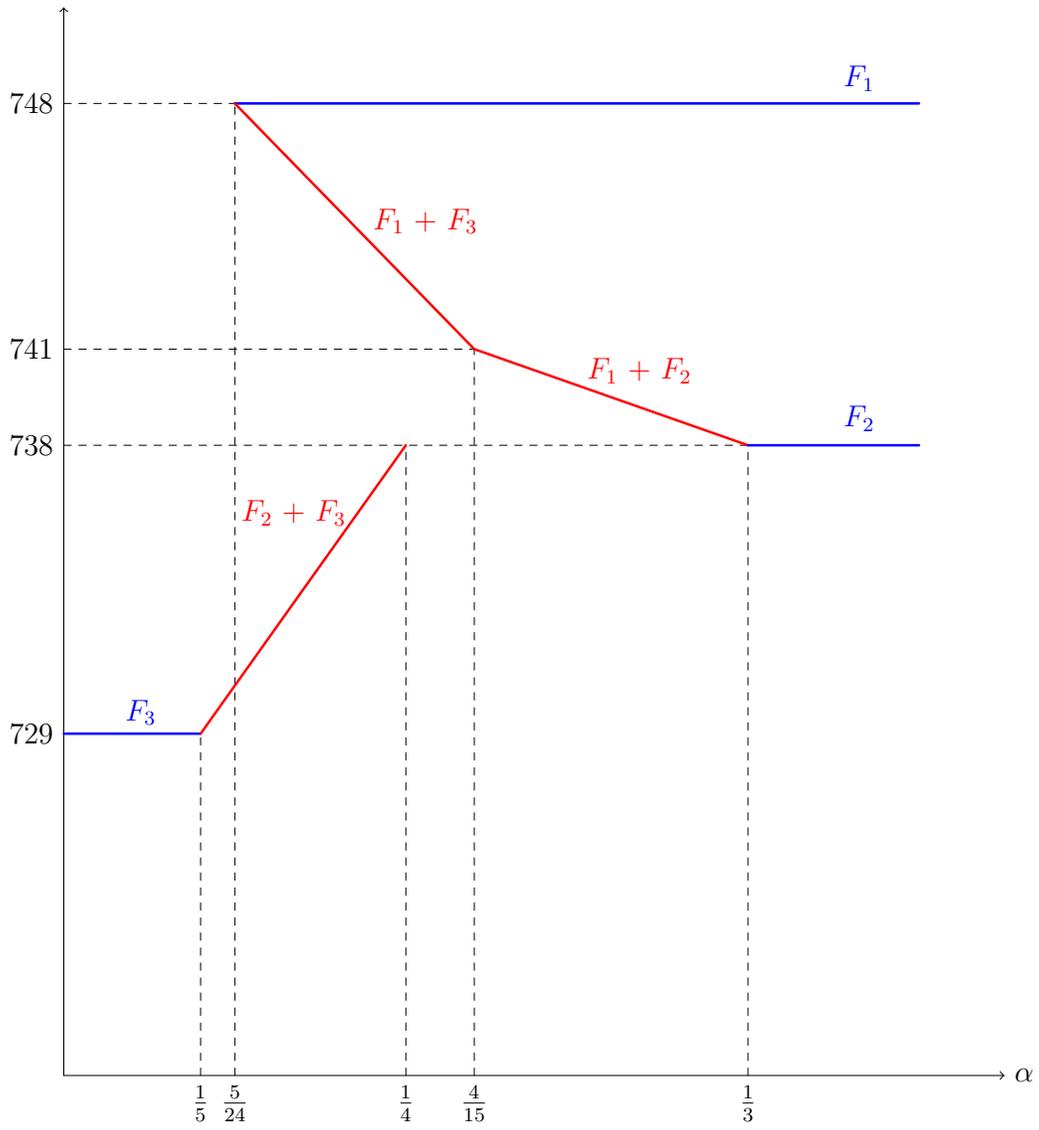
Figure 3: Equilibrium Payoffs

D q Varying 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. We show two results for this model which are analogs for Theorems 2 and 3.

Theorem 8. *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

$$\alpha \mathrm{E}_F(x) + (1 - \alpha)(1 - q_F)\hat{x} + (1 - \alpha)q_F \mathrm{E}_F \max\{x, \hat{x}\}$$
$$\geq \alpha \mathrm{E}_G(x) + (1 - \alpha)(1 - q_G)\hat{x} + (1 - \alpha)q_G \mathrm{E}_G \max\{x, \hat{x}\}$$

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

$$q_F[\mathrm{E}_F \max\{x, \hat{x}\} - \hat{x}] \geq q_G[\mathrm{E}_F \max\{x, \hat{x}\} - \hat{x}].$$

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 $\alpha \mathrm{E}_F(x) + (1 - \alpha)\hat{x} < \hat{x}$. 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

$$\sum_{F'} \sigma(F') \mathrm{E}_{F'}(x) \geq \hat{x},$$

where σ is the agent's mixed strategy. Hence the agent's equilibrium payoff strictly exceeds the payoff to project (F, q_F), a contradiction. ∎

Theorem 9. *For any set of feasible projects, any $\alpha \in [0, 1]$, and any equilibrium, the agent's payoff is at least α times the first–best payoff. Furthermore, there exists a set of feasible projects and an equilibrium such that the agent's payoff equals α times the first–best.*

Proof. To show the bound, fix any set of feasible projects, any α, 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

$$U \geq \alpha \mathrm{E}_F(x) + (1 - \alpha)q_F \mathrm{E}_F \max\{x, \hat{x}\} + (1 - \alpha)(1 - q_F)\hat{x}$$
$$\geq \alpha \mathrm{E}_F(x)$$

35

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

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," Stanford University working paper, 2012.

[3] Bond, P., and I. Goldstein, "Government Intervention and Information Aggregation by Prices," University of Pennsylvania working paper, 2014.

[4] Diamond, D., and R. Verrecchia, "Disclosure, Liquidity, and the Cost of Capital," *Journal of Finance*, **46**, September 1991, 1325–1359.

[5] Dye, R. A., "Disclosure of Nonproprietary Information," *Journal of Accounting Research*, **23**, 1985a, 123–145.

[6] Dye, R. A., "Strategic Accounting Choice and the Effect of Alternative Financial Reporting Requirements," *Journal of Accounting Research*, **23**, 1985b, 544–574.

[7] Edmans, A., M. Heinle, and C. Huang, "The Real Costs of Disclosure," Wharton working paper, May 2013.

[8] Fishman, M. and K. Hagerty, "The Optimal Amount of Discretion to Allow in Disclosure," *Quarterly Journal of Economics*, **105**, May 1990, 427–444.

[9] Forges, F. and F. Koessler, "Communication Equilibria with Partially Verifiable Types," *Journal of Mathematical Economics*, **41**, November 2005, 793–811.

[10] Forges, F. and F. Koessler, "Long Persuasion Games," *Journal of Economic Theory*, **143**, November 2008, 1–35.

[11] Gao, P., and P. Liang, "Informational Feedback Effect, Adverse Selection, and the Optimal Disclosure Policy," *Journal of Accounting Research*, **51**, December 2013, 1133–1158.

[12] Gigler, F., C. Kanodia, H. Sapra, 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.

[13] Glazer, J., and A. Rubinstein, "On Optimal Rules of Persuasion," *Econometrica*, **72**, November 2004, 1715–1736.

[14] Glazer, J., and A. Rubinstein, "A Study in the Pragmatics of Persuasion: A Game Theoretical Approach," *Theoretical Economics*, **1**, December 2006, 395–410.

[15] Grossman, S. J., "The Informational Role of Warranties and Private Disclosures about Product Quality," *Journal of Law and Economics*, **24**, 1981, 461–483.

[16] Guttman, I., I. Kremer, and A. Skrzypacz, "Not Only What but also When: A Theory of Dynamic Voluntary Disclosure," *American Economic Review*, forthcoming, 2013.

[17] 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.

[18] Jung, W., and Y. Kwon, "Disclosure When the Market is Unsure of Information Endowment of Managers," *Journal of Accounting Research*, **26**, 1988, 146–153.

[19] Kanodia, C., and A. Mukherji, "Real Effects of Separating Investment and Operating Cash Flows," *Review of Accounting Studies*, **1**, 1996, 51–71.

[20] Kanodia, C., H. Sapra, and R. Venugopalan, "Should Intangibles Be Measured: What Are the Economic Trade–Offs?" *Journal of Accounting Research*, **42**, March 2004, 89–120.

[21] 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.

[22] Kreps, D., and R. Wilson, "Sequential Equilibrium," *Econometrica*, **50**, July 1982, 863–894.

[23] Lipman, B., and D. Seppi, "Robust Inference in Communication Games with Partial Provability," *Journal of Economic Theory*, **66**, August 1995, 370–405.

[24] Milgrom, P., "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, **12**, 1981, 350–391.

[25] Mas-Colell, A., M. Whinston, and J. Green, *Microeconomic Theory*, New York: Oxford University Press, 1995.

[26] Okuno-Fujiwara, M., A. Postlewaite, and K. Suzumura, "Strategic Information Revelation," *Review of Economic Studies*, **57**, January 1990, 25–47.

[27] Roughgarden, T., *Selfish Routing and the Price of Anarchy*, MIT Press, 2005.

[28] Shin, H. S., "The Burden of Proof in a Game of Persuasion," *Journal of Economic Theory*, **64**, October 1994, 253–264.

[29] Shin, H. S., "Disclosures and Asset Returns," *Econometrica*, **71**, January 2003, 105–133.

[30] Stein, J., "Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior," *Quarterly Journal of Economics*, **101**, November 1989, 655–669.

[31] Verrecchia, R. E., "Discretionary Disclosure," *Journal of Accounting and Economics*, **5**, 1983, 179–194.