The Equilibrium Value of Employee Ethics[*]

Brendan Daley[†]   Simon Gervais[‡]

May 30, 2017

Preliminary and incomplete[§]

Abstract

We propose a model in which employees differ in both skills and ethics. Employees receive a private, non-verifiable signal about the quality of their project, and decide whether or not to undertake it. Skilled employees lead more projects to success than unskilled ones. Ethical employees internalize the overall costs and benefits of a project whereas strategic employees make their decision based solely on personal rewards (i.e., future wages). Hence, both skills and ethics are valuable to firms, which is reflected in how employee wages vary with firms' beliefs about an employee's type given his history. The model's main tension comes from the fact that a strategic employee keeps the option of showcasing his skills when he undertakes a project whereas, since ethical employees drop all negative-NPV projects, he improves his ethical capital when he drops a project. We characterize equilibrium employee decisions and wages and derive comparative statics, first in a one-period model with exogenous wages and then in a two-period model that endogenizes the value of skills and ethics. An inefficient equilibrium emerges when ethics are not prominent or skills are highly valuable to firms. In this equilibrium, skilled employees undertake more negative-NPV projects than their unskilled counterparts, as they rely on their skills to recoup the bad information they initially receive. In this way, the presence of ethics among some employees disciplines the behavior of strategic ones to make decisions more aligned with overall value. Because the disciplining effect is stronger on low-skill employees, the need for interventionist governance solutions is reduced.

1 Introduction

A critical issue in the study of organizations and corporate governance is that of preference alignment/misalignment between employees and the overall organization or shareholders (Shleifer and Vishny, 1997; Becht et al., 2003; Gibbons and Roberts, 2012). For concreteness, consider a manager who faces the decision of whether or not to undertake a risky project and who possesses private (non-verifiable) information regarding the likelihood of its success. A skilled manager is more likely to produce successes. Hence, if future wages are increasing in the market's perception of his skill, a strategic manager has an incentive to undertake projects to avoid being perceived as unskilled, even if the project's likelihood of success implies a negative net present value for the firm.

Problems such as this one have motivated the study of incentive instruments potentially available to the firm, such as performance contracts or costly technologies for monitoring/auditing. In this paper, we explore a different channel, based on two-dimensional heterogeneity in agents, with wages based only upon expected future productivity (as suggested by Fama, 1980, and formalized by Holmström, 1999).

It has been well documented empirically that workers vary not only in skill, but also in their commitment to the success of the firm overall.[1] Suppose then that an *ethical* agent always undertakes a project if and only if doing so creates positive expected (net) value for the firm. It is clear that such an attribute is valuable to employers.[2] However, if the main differentiating behavior of these agents is undertaking projects *less* often, then they can be very easily imitated by strategic agents.

Hence, the two dimensions of heterogeneity potentially create a tradeoff for strategic agents. Undertaking projects may increase the perception that they are skilled, whereas forgoing projects may increase the perception that they are ethical, both of which are valued by the market. We seek to understand how this tension resolves in equilibrium and the extent to which the presence (and value) of ethics among some agents decreases the desirability of the more interventionist instruments described above.

Notice that there are two elements endogenously determined in equilibrium: (i) given market values for skills and ethics, agent strategies must be optimal and market perceptions based on outcomes must be consistent with these strategies; and (ii) the market value for skills and ethics must be consistent with what types do in equilibrium. In the paper, we tackle these issues in sequence. We begin by characterized the equilibrium component in (i) with exogenous market

---

[1]Within the Management literature this documentation is categorized as observed "normative organizational commitment" (Meyer and Allen, 1997; Meyer et al., 2002; Jaros 2007) and "organizational citizenship behavior (OCBO)" (Bateman and Organ 1983; Williams and Anderson, 1991).

[2]Such an agent bears resemblance to the "friendly" agent in Sobel's (1985) more abstract model of "credibility."

values for skills and ethics. We then endogenize these values, as described in (ii).

We start our analysis with a one-period model in which a firm hires an agent to make an investment decision in a single project on its behalf. The agent's function within the firm involves the observation of a private signal about the profitability of the project and the subsequent decision as to whether or not to invest in this project. Unlike the classic principal-agent model of Ross (1973) and Holmström (1979) in which the agent's incentives are realigned using contracts that are contingent on publicly observable outcomes, we assume that the agent's motives are rooted in his reputation. That is, as in career-concern models (e.g., Harris and Holmström, 1982; Holmström, 1999), the agent cares about how the firm will value him in the future based on decisions that he makes now.

Our main innovation comes from the fact that the agent's type is two-dimensional: the agent can be skilled or unskilled, and he can be ethical or unethical (or strategic). The skilled agent is more likely to lead a project to success, while an ethical agent makes the investment decisions that the firm would like him to make even if it is not always personally advantageous to do so. We assume that these traits are drawn independently from a population, and that only the manager himself knows the realization of this draw at the outset. However, because both traits are valued by the firm, the agent knows that his actions during the period will allow the firm to form posterior beliefs about his skills. In this first part of the paper, we assume that the agent's payoff is exogenously specified based on these posteriors by the firm. In particular, his payoff is assumed to be linear in the posterior probability that he is skilled and the posterior probability that he is ethical.

Although the agent would like to be be perceived as skilled and ethical, convincing the firm that he is both is difficult. Indeed, an agent who passes on a project looks ethical, but he also foregoes an opportunity to showcase his skill which comes into effect when undertaking the project and leading it to success. Similarly, the investment into a project allows the agent to advertise his skill but, since ethical agents are on average more conservative in their investment decisions, such investment tend to be associated with lower ethics. This tension is captured in the model by the fact that investment decisions are based on profitability signals that are observed privately by the agent before he makes the investment decision.

We show that there are two equilibria to this game. In the efficient equilibrium, all agents, ethical or not, follow the investment policy that the firm would like them to. In this "pooling" equilibrium, the firm learns nothing about the agent's skill or ethics, and so compensation is outcome-independent at the end of the period. In the inefficient equilibrium, strategic agents undertake some negative-NPV projects, despite the fact that such projects are detrimental to the firm. Interestingly, skilled agents undertake such projects more often than unskilled agents do. This is due to the fact that, for them, the higher probability of success means that the ethical

2

component of their reputation is less likely to take the big hit that comes with failure. Intuitively, their information tells them that they start with a strike but, because of their skill, these agents bet on their ability to recoup this initial setback.

Unskilled agents, on the other hand, more often choose to side with ethical agents and improve their ethical capital by dropping negative-NPV projects. For them, the high risk of looking both unskilled and unethical when they fail pushes them to be more conservative than their skilled counterparts. This is not to say that they never undertake negative-NPV project, however. Indeed, all inefficient equilibria have the unskilled undertake a fraction of negative-NPV projects as, were it not the case, failure would come with an undue boost in one's skill reputation. Realizing this, unskilled agents play a mixed strategy that balances the skill reputation they get from failure with the ethical reputation they get from dropping projects.

When agents get private ("empire-building") benefits from undertaking projects, the equilibrium to the game is always unique: it is efficient when ethics are handsomely rewarded, but inefficient if ethics-related concerns are limited. As we show, the inefficient equilibrium becomes less inefficient (i.e., strategic agents undertake fewer negative-NPV projects) when skilled agents are difficult to imitate because they are vastly superior to unskilled agent, when the population is known to have few ethical agents, and when few projects have a positive net present value.

As mentioned above, the payoffs associated with high perceived skills and ethics are exogenous in this one-period model. The second part of the paper seeks to endogenize these payoffs by adding a second period to the model. That is, if skills and ethics are valuable to the firm, they should be associated with better investment decisions and results in the second period. Thus agents who can improve their two-dimensional reputation can presumably extract more surplus from the firm in the second period, and this is what drives their first-period incentives.

We also use the two-period model to capture the idea that some firms and industries have better prospects (i.e., higher project payoffs) and that investments in some firms and industries are more committed than in others (i.e., more of the capital must be invested up front, before anything is learned about the project). In this context, we show that the model in which the firm hires an agent to manage one project per period for two periods naturally maps into the one-period model. Specifically, we show that the value of ethical agents comes from the option value of the project's cost (i.e., it is incurred only when the project is worthwhile), while the value of skilled agents comes from the higher likelihood of project success. Thus ethical agents tend to be more valuable in firms whose investment is more gradual and contingent on intermediate information. Skilled agent, on the other hand, are more valuable in firms that cannot easily reverse their investments (something that unethical agents do not do well) but that have larger project payoffs.

We finish the paper with a version of the two-period model that allows the firm to allocate the

agent to one of two jobs in each period. In particular, after the firm learns about the skills and ethics of the agent through first-period outcomes, it has the option of redeploying his services in a job with firm payoffs that hinge largely on ethics or a job in which the payoffs are mainly driven by skills. This version of the model no longer naturally maps into the one-period model, as the option component that comes with the firm's job assignment implies that the compensation schedule of the agent depends on first-period outcomes. In a sense, the firm can undo some of the strategic actions that the agent takes in the first period in order to improve his perceived skills and ethics. Of course, in the first period, the agent anticipates these job allocation decisions and so his first-period strategy determines not only the firm's updating process (as before), but also the job and compensation scheme that the firm will subsequently adopt. One important result that emanates from this two-job model is the fact that agents tend to be more aggressive in their cheating (i.e., undertake more negative-NPV projects) as the two jobs become more dissimilar. That is, when the payoffs from the two jobs are very different for ethical and unethical agents, the unethical agent find it more optimal to game the system.

2 The Basic Model

2.1 The Economy

Consider a one-period model in which a principal (the *firm*) hires an *agent* (e.g., a CEO or, more generally, an *employee*) to make an investment decision on its behalf. The firm's only project requires an initial investment of $k > 0$. Its end-of-period payoff $\tilde{v}$ is either one or zero, which occur with probabilities $\tilde{p}$ and $1 - \tilde{p}$, respectively. The firm does not observe $\tilde{p}$; only the agent does. Once he observes $\tilde{p}$, the agent decides on whether or not to undertake the project. Without loss of generality, the discount rate is assumed to be equal to zero.

The probability of success is related to the quality $q \in (0,1)$ of the project and to the agent's skill $\tilde{s}$. Specifically, $\tilde{p}$ is equal to g with probability q, and is otherwise equal to $\tilde{s} \in \{h, \ell\}$, with $0 \leq \ell < h < k < g \leq 1$. Ex ante, the agent is *skilled* ($\tilde{s} = h$) with probability ϕ, and *unskilled* ($\tilde{s} = \ell$) with probability $1 - \phi$; this skill is known by the agent, but not by the firm. With this specification, the project has a positive net present value when it is inherently good (as $g - k > 0$). When the project is bad (i.e., when $\tilde{p} = \tilde{s}$), the agent's skill determines its success probability. Since $\tilde{s} < k$, the project then has a negative net present value and so it is always the case that the firm would prefer the agent not to undertake it in that event.

The agent follows his incentives in making this decision on behalf of the firm. These incentives come from two sources: ethics and payoffs. Like Etzioni (1988), Rabin (1995), and Sen (1997), we assume that *ethical* agents face moral constraints in this decision. Specifically we assume that a

fraction $\psi \in (0,1)$ of agents are *ethical* ($\tilde{\varepsilon} = e$) and that a fraction $1 - \psi$ of agents are *unethical* ($\tilde{\varepsilon} = u$); only the agent knows whether he is ethical or not (we also refer to unethical agents as *strategic*). Ethical agents behave exactly like the firm would like them to: they undertake the project if and only if they observe that $\tilde{p} = g$. Strategic agents, on the other hand, only consider their own payoffs in this decision.

For exposition purposes, we initially assume that these payoffs are specified exogenously, but we dedicate section 4 to endogenizing them. The payoffs are set to capture the idea that agents care about how their skills and ethics are perceived by the firm, or more generally in labor markets, at the end of the period. That is, as in the models of Holmström and Ricart i Costa (1986), and Holmström (1999), the career concerns of agents guide their decisions. However, we add the possibility that agents care not only about how their skills are perceived, but also their ethics as in Sobel's (1985) work on credibility. As we shall see, this dual concern creates unavoidable tensions: agents who are intent on looking skilled do so at the risk of losing their reputation for being ethical.

Let us denote by Ω the information set that is publicly available at the end of the period. We assume that the agent's end-of-period payoff is given by

$$W_\Omega \equiv w_{he} \Pr\{\tilde{s} = h, \tilde{\varepsilon} = e \mid \Omega\} + w_{\ell e} \Pr\{\tilde{s} = \ell, \tilde{\varepsilon} = e \mid \Omega\} + w_{hu} \Pr\{\tilde{s} = h, \tilde{\varepsilon} = u \mid \Omega\}, \quad (1)$$

where w_{he}, $w_{\ell e}$, and w_{hu} are positive constants. Because the agent's skills and ethics are privately known, the set Ω includes whether or not the project was undertaken and, if undertaken, whether or not it was successful. That is, from the public's perspective, three outcomes (or *states*) are possible: success, no investment, failure. The firm's net profit from the project in each of these states is $1 - k$, 0, and $-k$, respectively. Since $1 - k > 0 > -k$, we refer to these outcomes as good ($\Omega = \mathrm{G}$), medium ($\Omega = \mathrm{M}$), and bad ($\Omega = \mathrm{B}$), respectively, and use $P_\Omega^{s\varepsilon}$ to denote the probability $\Pr\{\tilde{s} = s, \tilde{\varepsilon} = \varepsilon \mid \Omega\}$ of each skill-ethics pair at the end of the period.

As we shall see, each of the three outcomes will be associated with a different public update about the agent's skill and ethics. In this sense, (1) implicitly assumes that different combinations of skill and ethics are worth a different amount to the agent, presumably because the opportunities of agents of each type, were it publicly known, differ. Indeed this is how we endogenize the agent's payoffs in section 4; that is, we calculate the market value of an agent with a given set of posterior probabilities, $P_\Omega^{s\varepsilon}$ for $\{s, e\} \in \{h, \ell\} \times \{e, u\}$, given outcome $\Omega \in \{\mathrm{G}, \mathrm{M}, \mathrm{B}\}$.

We impose the restrictions that $w_{he} \geq w_{hu}$ and $w_{he} \geq w_{\ell e}$, which are meant to capture the idea that agents who are both skilled and ethical are bound to be more valuable to firms than agents who exhibit only one of the two traits. This restriction is imposed exogenously in this one-period version of the model but, as we shall see in section 4, it endogenously comes about when a second period is added and agents are paid according to the value they create for the firm.

Finally, we assume that agents have private empire-building motives, in that they get a payoff boost of $b > 0$ when they undertake the project. As we shall see, this private benefit only serves to break ties in the equilibrium analysis. In fact, throughout the paper, we will assume that this benefit is arbitrarily close to zero. We also note that, although ethical agents receive this private benefit when they undertake the project, they do not factor it into the investment decision they make on behalf of the firm.

2.2 Strategies and Updating

An agent of skill $s \in \{h, \ell\}$ observes $\tilde{p} \in \{g, s\}$ and must decide whether or not to undertake the project. Since ethical agents always act in the best interest of the firm and $g > k > h > \ell$, they invest in the project if $\tilde{p} = 1$ and refrain from doing so otherwise. Strategic agents of skill s choose a mapping $a_s(p)$ from $\mathcal{P}_s \equiv \{g, s\}$ to $\mathcal{A} \equiv \{\text{I}, \text{N}\}$ (for Invest and Not invest, respectively). We allow for (and indeed find) mixed-strategy equilibria. To economize on notation, we denote the equilibrium (mixing) probability that an agent who observes $\tilde{p} = p \in \{g, h, \ell\}$ chooses to undertake the project by $\sigma_p \in [0, 1]$.[3]

The probability that an agent of skill-ethics type $\{s, \varepsilon\}$ undertakes a project that succeeds (or fails) depends on how often he undertakes the project and on his success rate when he does. Ethical agents of either skill undertake the project only when $\tilde{p} = g$, and thus

$$\Pr\{G \mid \tilde{s} = s, \tilde{\varepsilon} = e\} = \Pr\{\tilde{p} = g\}\Pr\{\tilde{v} = 1 \mid \tilde{p} = g\} = qg, \tag{2a}$$

$$\Pr\{B \mid \tilde{s} = s, \tilde{\varepsilon} = e\} = \Pr\{\tilde{p} = g\}\Pr\{\tilde{v} = 0 \mid \tilde{p} = g\} = q(1 - g), \quad \text{and} \tag{2b}$$

$$\Pr\{M \mid \tilde{s} = s, \tilde{\varepsilon} = e\} = 1 - q. \tag{2c}$$

Strategic agents, on the other hand, do not necessarily undertake the project only when $\tilde{p} = g$. For them, we have

$$\Pr\{G \mid \tilde{s} = s, \tilde{\varepsilon} = u\} = \Pr\{\tilde{p} = g\}\sigma_g\Pr\{\tilde{v} = 1 \mid \tilde{p} = g\} + \Pr\{\tilde{p} = s\}\sigma_s\Pr\{\tilde{v} = 1 \mid \tilde{p} = s\}$$
$$= qg\sigma_g + (1 - q)s\sigma_s, \tag{3a}$$

$$\Pr\{B \mid \tilde{s} = s, \tilde{\varepsilon} = u\} = \Pr\{\tilde{p} = g\}\sigma_g\Pr\{\tilde{v} = 0 \mid \tilde{p} = g\} + \Pr\{\tilde{p} = s\}\sigma_s\Pr\{\tilde{v} = 0 \mid \tilde{p} = s\}$$
$$= q(1 - g)\sigma_g + (1 - q)(1 - s)\sigma_s, \quad \text{and} \tag{3b}$$

$$\Pr\{M \mid \tilde{s} = s, \tilde{\varepsilon} = u\} = q(1 - \sigma_g) + (1 - q)(1 - \sigma_s) = 1 - q\sigma_g - (1 - q)\sigma_s. \tag{3c}$$

---

[3]This rules out the possibility that an agent of skill h who observes $\tilde{p} = g$ chooses a different strategy than an agent of skill ℓ who also observes $\tilde{p} = 1$. Because agents make one and only one decision before they receive their payoffs, this restriction is without loss of generality. This would not necessarily be the case if the decisions that agents make early anticipated future decisions.

The updated probability that an agent who experiences an outcome of $\Omega \in \{G, M, B\}$ is of skill-ethics type $\{s, \varepsilon\}$ is obtained via Bayes' rule,

$$P_\Omega^{s\varepsilon} = \frac{\Pr\{\Omega \mid \tilde{s} = s, \tilde{\varepsilon} = \varepsilon\} \Pr\{\tilde{s} = s\} \Pr\{\tilde{\varepsilon} = \varepsilon\}}{\sum_{s' \in \{h, \ell\}} \sum_{\varepsilon' \in \{e, u\}} \Pr\{\Omega \mid \tilde{s} = s', \tilde{\varepsilon} = \varepsilon'\} \Pr\{\tilde{s} = s'\} \Pr\{\tilde{\varepsilon} = \varepsilon'\}}, \tag{4}$$

where we have used the fact that $\tilde{s}$ and $\tilde{\varepsilon}$ are independent ex ante, something that is generally not true ex post, as we shall see.

3 Equilibrium Analysis

3.1 Definitions and Preliminaries

In this section, we proceed to derive and analyze the equilibrium to the model of section 2. We start with the following lemma, which shows that, because ethical agents of either skill make the same decisions with the same frequencies, the information that is publicly available at the end of the period never facilitates an update about the skill of ethical agents.

Lemma 1. *Conditional on an agent being ethical, the posterior probability that he is skilled is equal to the prior probability that he is skilled; that is*

$$\Pr\{\tilde{s} = h \mid \Omega, \tilde{\varepsilon} = e\} = \Pr\{\tilde{s} = h\} = \phi. \tag{5}$$

This result implies that (1) can always be written as

$$W_\Omega = w_{hu} \Pr\{\tilde{s} = h, \tilde{\varepsilon} = u \mid \Omega\} + \left[\phi w_{he} + (1 - \phi)w_{\ell e}\right] \Pr\{\tilde{\varepsilon} = e \mid \Omega\}. \tag{6}$$

For this reason, we define $w_{\text{E}} \equiv \phi w_{he} + (1 - \phi)w_{\ell e}$ to denote the compensation that an agent can expect to receive when he is known to be ethical. To keep the notation intuitive, we also let $w_{\text{S}} \equiv w_{hu}$ denote the compensation of an agent known to be skilled and strategic, and rewrite (6) as

$$W_\Omega = w_{\text{S}} \Pr\{\tilde{s} = h, \tilde{\varepsilon} = u \mid \Omega\} + w_{\text{E}} \Pr\{\tilde{\varepsilon} = e \mid \Omega\} = P_\Omega^{hu} w_{\text{S}} + \left(P_\Omega^{he} + P_\Omega^{\ell e}\right) w_{\text{E}}. \tag{7}$$

Note also that, with this specification, the restriction that $w_{he} \geq w_{hu}$ becomes $w_{\text{E}} \geq \phi w_{\text{S}}$.

When an agent observes $\tilde{p} = p \in \{g, h, \ell\}$ and chooses to undertake the project, the probability that the project succeeds is p and the probability it fails is $1 - p$. This means that the payoff that the agent can expect from undertaking the project is $pW_{\text{G}} + (1 - p)W_{\text{B}} + b$, where b represents the extra utility boost from empire-building. When the same agent chooses not to undertake the project, his end-of-period payoff is certain and equal to W_{M}. Thus the agent undertakes the project for sure ($\sigma_p = 1$) if $pW_{\text{G}} + (1 - p)W_{\text{B}} + b > W_{\text{M}}$, drops the project for sure ($\sigma_p = 0$) if $pW_{\text{G}} + (1 - p)W_{\text{B}} + b < W_{\text{M}}$, and mixes ($0 < \sigma_p < 1$) if $pW_{\text{G}} + (1 - p)W_{\text{B}} + b = W_{\text{M}}$. This leads to the following definition of an equilibrium

Definition 1 (EQUILIBRIUM). *An equilibrium consists of $\sigma_p \in [0,1]$, $p \in \{g, h, \ell\}$, satisfying the following conditions.*

(i) *For each outcome Ω, the posterior probability $P_\Omega^{s\varepsilon}$ of each skill-ethics type $\{s, \varepsilon\}$ is obtained via Bayes' rule, in (4).*

(ii) *For each outcome $\Omega \in \{G, M, B\}$, the wage to the agent is W_Ω, as specified in (7).*

(iii) *Conditional on observing $\tilde{p} = p \in \{g, h, \ell\}$, strategic agents*

- *undertake the project for sure ($\sigma_p = 1$) if $pW_G + (1-p)W_B + b > W_M$,*
- *drop the project for sure ($\sigma_p = 0$) if $pW_G + (1-p)W_B + b < W_M$, or*
- *employ a mixed strategy ($0 < \sigma_p < 1$) with $pW_G + (1-p)W_B + b = W_M$ otherwise.*

Clearly, if $W_G \neq W_B$, it cannot be the case that $pW_G + (1-p)W_B + b = W_M$ for more than one value of $p \in \{g, h, \ell\}$ and, in this case, at most one (and possibly none) of σ_g, σ_h, or σ_ℓ is ever strictly above zero and below one. Also, intuitively, since skilled agents are more likely than unskilled agents to experience a success when they undertake a project, it is always the case that $W_G \geq W_B$ and that $\sigma_g \geq \sigma_h \geq \sigma_\ell$ in equilibrium. The following lemma formalizes this result.

Lemma 2. *In equilibrium, it must always be the case that $W_G \geq W_B$.*

Because the private empire-building benefit that comes with undertaking a project affects the decisions of strategic agents, the comparison between W_M and either of W_G or W_B is affected by the size of b. As the following lemma shows, a weak ordering exists across the three state-dependent wages that the agent can receive at the end of the period, and this ordering depends on whether or not b is sufficiently close to zero.

Lemma 3. *In equilibrium, it must always be the case that $W_G + b \geq W_M \geq W_B + b$ or that $W_G + b \geq W_B + b > W_M$, with the former holding if and only $b \leq \bar{b}$ for some $\bar{b} > 0$ derived in the proof. In either case, we always have $\sigma_g = 1$.*

Because most of the paper's result are derived under the assumption that b is infinitesimally small, Lemma 3 tells us that we can safely rely on the fact that the agent's payoff is greatest (smallest) when he leads a project to success (to failure). Indeed, the case with $b > \bar{b}$ is not particularly interesting, as the strategic agents' private benefit from undertaking the project is so large that they never seriously consider the alternative.

The fact that project success always yields the greatest payoff for agents means that they are naturally attracted to undertaking the project when their initial information about it indicates

that it is of high quality. More specifically, as the following lemma shows, it is never strategically optimal for an agent to drop the project when $\tilde{p} = g$.

Lemma 4. *In equilibrium, it is always the case that $\sigma_g = 1$.*

This last result establishes that projects known to have a positive net present value are always undertaken, even by strategic agents. Given this, it is useful to restate the posterior probabilities that the agent is of a given skill-ethics type, as shown in (4), taking σ_h and σ_ℓ as fixed constants. Given (7), we are particularly interested in $\pi_S(\Omega) \equiv \Pr\{\tilde{s} = h, \tilde{\varepsilon} = u | \Omega\}$ and $\pi_E(\Omega) \equiv \Pr\{\tilde{\varepsilon} = e | \Omega\}$.

Corollary 1. *Assume that strategic agents undertake the project with probability $\sigma_h \in [0,1]$ when they observe $\tilde{p} = h$, and that they undertake it with probability $\sigma_\ell \in [0,1]$ when they observe $\tilde{p} = \ell$. Given a successful project, we have*

$$\pi_S(G) = \frac{\big[qg + (1-q)h\sigma_h\big]\phi(1-\psi)}{qg + (1-q)\big[\phi h\sigma_h + (1-\phi)\ell\sigma_\ell\big](1-\psi)}, \tag{8a}$$

$$\pi_E(G) = \frac{qg\psi}{qg + (1-q)\big[\phi h\sigma_h + (1-\phi)\ell\sigma_\ell\big](1-\psi)}. \tag{8b}$$

Given a failed project, we have

$$\pi_S(B) = \frac{\big[q(1-g) + (1-q)(1-h)\sigma_h\big]\phi(1-\psi)}{q(1-g) + (1-q)\big[\phi(1-h)\sigma_h + (1-\phi)(1-\ell)\sigma_\ell\big](1-\psi)}, \tag{9a}$$

$$\pi_E(B) = \frac{q(1-g)\psi}{q(1-g) + (1-q)\big[\phi(1-h)\sigma_h + (1-\phi)(1-\ell)\sigma_\ell\big](1-\psi)}. \tag{9b}$$

Given a dropped project, we have

$$\pi_S(M) = \frac{\phi(1-\psi)(1-\sigma_h)}{\psi + (1-\psi)\big[\phi(1-\sigma_h) + (1-\phi)(1-\sigma_\ell)\big]}, \tag{10a}$$

$$\pi_E(M) = \frac{\psi}{\psi + (1-\psi)\big[\phi(1-\sigma_h) + (1-\phi)(1-\sigma_\ell)\big]}. \tag{10b}$$

The comparative statics with respect to σ_h and σ_ℓ illustrate some of the tensions that exist in the model. For example, it is easy to verify that $\pi_S(G)$ increases in σ_h and decreases in σ_ℓ. Even though it is inefficient for skilled agents to undertake the project when $\tilde{p} = h < k$, doing so increases the probability that they will be perceived as skilled at the end of the period. The opposite is true for the unskilled agents. However, $\pi_E(G)$ is decreasing in both σ_h and σ_ℓ: the fact that strategic agents sometimes undertake negative-NPV projects reduces the fraction of successful projects that are led by ethical agents.

It can also be shown that, for similar reasons, $\pi_S(B)$ is increasing in σ_h and decreasing in σ_ℓ. That is, when more skilled agents undertake negative-NPV projects, failure comes with an increased

perception of their skill. Of course, since the average quality of the projects that unethical agents undertake decreases with both σ_h and σ_ℓ, relative to that of ethical agents (who never undertake the project with $\tilde{p} = h$ or $\tilde{p} = \ell$), it is also the case that failure is less likely to be associated with ethics as σ_h and σ_ℓ increase. Nevertheless, as we shall see below, this possibility of being perceived as skilled and unethical makes it worthwhile for unskilled agents to undertake projects that have a very low likelihood of succeeding. In fact, even when they are sure to fail (i.e., when $\tilde{p} = \ell = 0$), unskilled agents still choose to undertake the risky project with some probability in equilibrium (i.e., $\sigma_\ell > 0$).

The comparative statics for $\pi_S(M)$ and $\pi_E(M)$ are also intuitive. The former is decreasing in σ_h and increasing in σ_ℓ. Thus dropped projects are less (more) likely to be associated with skilled agents when skilled (unskilled) agents undertake more negative-NPV projects and drop fewer projects. Finally, because strategic cheaters tend to undertake more projects, they make it more likely that the decision to drop a project comes from an ethical agent; that is, $\pi_E(M)$ is increasing in both σ_h and σ_ℓ.

Because $\pi_E(\Omega)$ is monotonic in both σ_h and σ_ℓ for all $\Omega \in \{G, B, M\}$, we can readily conclude that $\pi_E(G) < \psi$, $\pi_E(B) < \psi$, and $\pi_E(M) > \psi$ in any equilibrium with $\sigma_g > 0$ or $\sigma_\ell > 0$. That is the decision of strategic agents to cheat implies that projects that are undertaken, wether they succeed or fail, are more likely to be managed by unethical agents. Although σ_h and σ_ℓ have opposite effects on $\pi_E(\Omega)$, $\Omega \in \{G, B, M\}$, it is still possible to derive conditions that allow a comparison of this quantity to $\phi(1 - \psi)$, the unconditional probability that the agent is skilled and unethical.

Lemma 5. *Suppose that at least one of σ_h or σ_ℓ is strictly positive. Then we have $\pi_E(G) < \psi$, $\pi_E(B) < \psi$, and $\pi_E(M) > \psi$. If $\sigma_h > \sigma_\ell$, then $\pi_S(G) > \phi(1 - \psi)$ and $\pi_S(M) < \phi(1 - \psi)$. Finally, $\pi_S(B) > \phi(1 - \psi)$ if and only if*

$$\frac{(1 - h)\sigma_h}{(1 - \ell)\sigma_\ell} > \frac{(1 - \phi)(1 - \psi)}{(1 - \phi)(1 - \psi) + \psi} = \frac{1}{1 + \frac{\psi}{(1 - \phi)(1 - \psi)}}. \tag{11}$$

Since $W_G \geq W_M \geq W_B$, it is always (and trivially) the case that $\sigma_h \geq \sigma_\ell$. Also, as we will see in section 3.2, this inequality is strict in any equilibrium that involves some cheating by unethical agents. In fact, more specifically, we will show that $1 = \sigma_h > \sigma_\ell > 0$. Finally, condition (11), when unskilled agents cheat sufficiently less than skilled ones (i.e., when σ_ℓ is small relative to σ_h), the expected skill that can be inferred from a failed project is greater than ex ante. As we show in section 3.2, this condition always holds in equilibrium. Essentially, unskilled agents are never so aggressive in their cheating in order to ensure that a failure, which is detrimental to their ethics reputation and which they are prone to experience when undertaking a project with $\tilde{p} = \ell$, still boosts their apparent skill.

3.2 Equilibrium

In equilibrium, the appeal of the risky project for strategic agents comes from the size of w_S relative to that of w_E. In particular, when $b = 0$, the critical comparison is whether or not

$$\frac{w_\mathrm{S}}{w_\mathrm{E}} > \frac{\psi}{1 - \phi(1 - \psi)}. \tag{12}$$

In what follows, we consider the generic cases in which either (12) holds or strictly fails to hold—that is, statements of (12) failing to hold mean that the strict inequality is reversed.[4]

Proposition 1. *Assume that the agent's private benefit for undertaking a project is $b = 0$. Then there always exists an equilibrium with $\sigma_g = 1$ and $\sigma_h = \sigma_\ell = 0$. If (12) holds, then there exists a second equilibrium with $\sigma_g = \sigma_h = 1$ and $\sigma_\ell = \sigma_\ell^* \in (0, 1)$. No other equilibrium exists.*

Recall that the firm always finds it optimal to undertake the project when $\tilde{p} = g$, but never does when $\tilde{p} \neq g$. The fact that the strategic agent always chooses to undertake the project when $\tilde{p} = g$ (i.e., the fact that $\sigma_g = 1$) shows that strategic agents are never more conservative than the firm would like them to be. However, the fact that σ_h and σ_ℓ can be strictly above zero implies that strategic agents will choose to undertake negative-NPV projects when w_S is large relative to w_E. Intuitively, since ethical agents systematically drop projects with $\tilde{p} = h$ or $\tilde{p} = 0$, the public interprets a dropped project as a sign that the agent is ethical. Strategic agents only value pooling with them when the payoff from looking ethical is large enough relative to that of looking skilled. In what follows, since σ_h and σ_ℓ measure the frequency with which negative-NPV projects are undertaken, we take these two quantities to measure the extent of *cheating* by the skilled and unskilled types, respectively.

Let us first analyze the equilibrium with $(\sigma_g, \sigma_h, \sigma_\ell) = (1, 0, 0)$. Since only the projects with $\tilde{p} = g$ are undertaken, this equilibrium achieves first-best. That is, the agents implement the investment policy that the firm would like them to implement. Under this policy, only the equality of the project matters; the agent's characteristics (skill and ethics) do not play a role. In fact, it is then trivially the case that the public posteriors about the agent's skill and ethics are identical to the priors; that is $\Pr\{\tilde{s} = h \mid \Omega\} = \phi$ and $\Pr\{\tilde{\varepsilon} = e \mid \Omega\} = \psi$ for any outcome $\Omega \in \{\mathrm{G}, \mathrm{M}, \mathrm{B}\}$. As Proposition 2 shows below, however, this equilibrium can be fragile.[5]

When (12) holds, an inefficient equilibrium emerges, one in which $\sigma_h = 1$ and $\sigma_\ell > 0$. In this equilibrium, strategic agents who are skilled never drop the project, even when they learn that its net present value is negative. Strategic unskilled agents undertake the project with positive

---

[4]With $b = 0$, if (12) is an equality, then $(\sigma_g, \sigma_h, \sigma_\ell)$ is an equilibrium if and only if $\sigma_g = 1, \sigma_h \in [0, 1]$, and $\sigma_\ell = 0$.

[5]Specifically, when (12) holds, any small empire-building benefit $b > 0$ renders the inefficient equilibrium (i.e., the one with $\sigma_h = 1$ and $\sigma_\ell > 0$) unique.

probability when they learn that its net present value is $\ell - k < 0$. The equilibrium probability σ_ℓ with which they do so is the subject of comparative static analysis in section 3.3 below. Suffice it to say for now that their tendency to cheat increase with $\frac{w_\mathrm{S}}{w_\mathrm{E}}$, the ratio of rewards they get from looking skilled and from looking ethical. In fact, given that (12) holds when the same ratio is sufficiently large, strategic skilled agents are also more prone to cheat when skills are better rewarded than ethics.

From here on, we assume that $\ell = 0$, so that an unskilled agent is sure to fail when $\tilde{p} \neq 1$ but he chooses to undertake the project anyway. This assumption is made purely for tractability purposes. Essentially, it turns the main equilibrium condition a quadratic equation into a linear equation, which greatly simplifies the exposition of the model without changing any of its properties or comparative statics.

Corollary 2. *If $b = 0$ (and $\ell = 0$), then σ_ℓ^* from Proposition 1 is given by*

$$\sigma_\ell^* = \frac{\phi\big[1 - qg - (1-q)h\big]\big[(1 - \phi + \phi\psi)w_\mathrm{S} - \psi w_\mathrm{E}\big]}{(1-\phi)\Big\{\big[1 - qg - (1-q)h\big]\phi(1-\psi)w_\mathrm{S} + (1-qg)\psi w_\mathrm{E}\Big\}} \in (0,1). \tag{13}$$

The results of Corollary 1 offer more specific insights about condition (12) leading to an equilibrium in which both skilled and unskilled agents cheat, and more generally why we can never have an equilibrium in which only skilled agents cheat. Recall from (7) that the compensation of the agent is $W_\Omega = w_\mathrm{S}\pi_\mathrm{S}(\Omega) + w_\mathrm{E}\pi_\mathrm{E}(\Omega)$ in state $\Omega \in \{\mathrm{G}, \mathrm{B}, \mathrm{M}\}$. Suppose that only skilled agents cheat, i.e., suppose that $\sigma_h > 0$ and $\sigma_\ell = 0$. Using (8a) and (8b) (with $\ell = 0$), we can write

$$W_\mathrm{G} = \frac{\big[qg + (1-q)h\sigma_h\big]\phi(1-\psi)w_\mathrm{S} + qg\psi w_\mathrm{E}}{qg + (1-q)h\sigma_h\phi(1-\psi)}$$

$$= \frac{qg\big[\phi(1-\psi)w_\mathrm{S} + \psi w_\mathrm{E}\big] + \sigma_h(1-q)h\phi(1-\psi)w_\mathrm{S}}{qg + \sigma_h(1-q)h\phi(1-\psi)}.$$

This last expression highlights the fact that, as skilled agents increase the frequency with which they cheat, a higher proportion of their wage from succeeding comes from w_S. Specifically, we have $W_\mathrm{G} = \phi(1-\psi)w_\mathrm{S} + \psi w_\mathrm{E}$ at $\sigma_h = 0$ and, as σ_h increases, W_G is a weighted average between $\phi(1-\psi)w_\mathrm{S} + \psi w_\mathrm{E}$ and w_S. This is because only skilled unethical agent ever undertake negative-NPV projects (i.e., projects with $\tilde{p} \neq g$): as σ_h, more of them undertake the project and more of them succeed, thereby increasing the likelihood that success comes from skilled unethical agents. This also makes it clear that, as σ_h increases, W_G increases only if $w_\mathrm{S} > \phi(1-\psi)w_\mathrm{S} + \psi w_\mathrm{E}$, which is equivalent to (12).

Of course, whether $\sigma_g > 0$ and $\sigma_\ell = 0$ can ever be sustained in equilibrium depends on what happens to W_B and W_M as σ_h increases. Let us first concentrate on W_B. Using (9a) and (9b) from

Corollary 1 (again with $\ell = 0$), we have

$$
\begin{aligned}
W_{\mathrm{B}} &= \frac{\big[q(1-g) + (1-q)(1-h)\sigma_h\big]\phi(1-\psi)w_{\mathrm{S}} + q(1-g)\psi w_{\mathrm{E}}}{q(1-g) + (1-q)\phi(1-h)\sigma_h(1-\psi)} \\
&= \frac{q(1-g)\big[\phi(1-\psi)w_{\mathrm{S}} + \psi w_{\mathrm{E}}\big] + \sigma_h(1-q)(1-h)\phi(1-\psi)w_{\mathrm{S}}}{q(1-g) + \sigma_h(1-q)(1-h)\phi(1-\psi)}.
\end{aligned}
$$

This illustrates that, like W_{G}, W_{B} is a weighted average of $\phi(1-\psi)w_{\mathrm{S}} + \psi w_{\mathrm{E}}$ and w_{S}, with more weight on w_{S} as σ_h increases. In particular, this means that both W_{G} and W_{B} increase (decrease) as σ_h increases when condition (12) holds (does not hold). This is intuitive. When no one cheats, the proportion of types in each state Ω is identical; in particular the proportion of skilled unethical agents is $\phi(1-\psi)$ and the proposition of ethical agents is ψ. As skilled unethical agents increase their cheating, more of them experience failures and success (since they succeed with probability $h < 1$ when they cheat), and this has a similar effect on W_{G} and W_{B}.

Now, why does this rule out the possibility of an equilibrium with $\sigma_h > 0$ and $\sigma_\ell = 0$? The reason is that the effect of an increase in σ_h has precisely the opposite effect on W_{M}. To see this, let us use (10a) and (10b) from Corollary 1 (along with $\ell = 0$) to calculate

$$
W_{\mathrm{M}} = \frac{\phi(1-\psi)(1-\sigma_h)w_{\mathrm{S}} + \psi w_{\mathrm{E}}}{\psi + (1-\psi)\big[\phi(1-\sigma_h) + (1-\phi)\big]} = \frac{1 \cdot \big[\phi(1-\psi)w_{\mathrm{S}} + \psi w_{\mathrm{E}}\big] - \sigma_h\phi(1-\psi)w_{\mathrm{S}}}{1 - \sigma_h\phi(1-\psi)}.
$$

Again, we can see that W_{M} is a weighted average of $\phi(1-\psi)w_{\mathrm{S}} + \psi w_{\mathrm{E}}$ and w_{S}, and is equal to $\phi(1-\psi)w_{\mathrm{S}} + \psi w_{\mathrm{E}}$ at $\sigma_h = 0$. In this case, we can also see that more weight gets put on $\phi(1-\psi)w_{\mathrm{S}} + \psi w_{\mathrm{E}}$ as σ_h increases. Thus, when an increase in σ_h leads to an increase (decrease) in W_{G} and W_{B}, it simultaneously leads to a decrease (increase) in W_{M}. In the former case (when (12) holds), the fact that W_{G} and W_{B} are both greater than W_{M} implies that $pW_{\mathrm{G}} + (1-p)W_{\mathrm{B}} > W_{\mathrm{M}}$ for any $p \in \{g, h, \ell\}$, and so all strategic agents would prefer to always undertake the project. In the latter case (when (12) does not hold), we have $pW_{\mathrm{G}} + (1-p)W_{\mathrm{B}} < W_{\mathrm{M}}$ for any $p \in \{g, h, \ell\}$ when $\sigma_h > 0$, and so it is then the case that no strategic agent would ever want to undertake the project.

Our next result shows that the multiplicity of equilibria that prevails in Proposition 1 when (12) holds is fragile and that, ultimately, only one equilibrium is robust to the presence of private empire-building benefits.

Definition 2. *An equilibrium of the $b = 0$ model, $(\sigma_g^0, \sigma_h^0, \sigma_\ell^0)$, is* **robust** *if for all $b > 0$ arbitrarily small there exists an equilibrium $(\sigma_g^b, \sigma_h^b, \sigma_\ell^b)$ arbitrarly close to $(\sigma_g^0, \sigma_h^0, \sigma_\ell^0)$.*[6]

---

[6]Precisely, σ^0 is robust if there exists a sequence $\{b_k, \sigma^{b_k}\}_{k=1}^{\infty}$ such that (i) for each k, $b_k > 0$ and σ^{b_k} is an equilibrium of the model with $b = b_k$, (ii) $\lim_{k\to\infty} b_k = 0$, and (iii) $\lim_{k\to\infty} \sigma^{b_k} = \sigma^0$.

Proposition 2. *If (12) holds, the unique robust equilibrium is $(1, 1, \sigma_\ell^*)$, where σ_ℓ^* is given by the expression in (13). If (12) does not hold, the unique robust equilibrium is $(1, 0, 0)$.*

Effectively, Proposition 2 shows that, although multiple equilibria can exist when $b = 0$ and $\frac{w_S}{w_E}$ is sufficiently large, only one equilibrium is robust to the presence of private empire-building benefits, even when such benefits are small. The fact that the $(\sigma_g, \sigma_h, \sigma_\ell) = (1, 0, 0)$ equilibrium is not robust in that case is intuitive. As discussed above, in this equilibrium, we have $\Pr\{\tilde{s} = h \mid \Omega\} = \phi$ and $\Pr\{\tilde{\varepsilon} = e \mid \Omega\} = \psi$ for any outcome $\Omega \in \{G, M, B\}$, and thus $W_G = W_M = W_B$. If strategic agents can, in addition to W_G or W_B, capture private benefits of $b > 0$ from undertaking the project, then they all prefer to do so, breaking the equilibrium in the process. The fact that the equilibrium survives when (12) does not hold comes from the fact that, when agents are generously rewarded for their ethics, they all seek to pool with the ethical agents.

3.3 Comparative Statics

As discussed above, Proposition 2 implies that, as $\frac{w_S}{w_E}$ increases from zero, we move from an equilibrium without cheating ($\sigma_h = \sigma_\ell = 0$) to one in which skilled agents always cheat while unskilled agents cheat probabilistically ($\sigma_h = 1$, $\sigma_\ell \in (0, 1)$). By showing that (13) is strictly increasing in $\frac{w_S}{w_E}$, the following result shows that cheating increases monotonically in $\frac{w_S}{w_E}$.

Corollary 3. *The extent of cheating by strategic agents is weakly increasing in $\frac{w_S}{w_E}$. It is strictly increasing in $\frac{w_S}{w_E}$ when (12) holds.*

It is particularly interesting that, despite the fact that unskilled agents have no chance of succeeding when they undertake a project, they still choose to do so when $\frac{w_S}{w_E}$ exceeds $\frac{\psi}{1-\phi+\phi\psi}$. As discussed in section 3.2, this is due to the fact that, if only skilled agents cheat, the compensation for both successful and failed projects increases when w_S is large relative to w_E. Intuitively, unskilled agents realize that they will surely fail when they undertake the project with $\tilde{p} = \ell = 0$, but the idea of pooling with skilled agents, even if they are unethical, is still valuable. Essentially, when w_S is large relative to w_E, looking skilled is attractive for unskilled agents, and vice versa when w_S is small relative to w_E.

The fact that $\sigma_\ell < 1$ in equilibrium comes from the balancing act that unskilled agents perform in equilibrium. As they increase their cheating, fewer of them remain in the pool of dropped projects, and so the average skill and ethics go up in that pool. At the same time, since all the projects in which they cheat end up failing, the pool of failed projects is associated with lower skill and ethics. In particular, it is easy to show that $\sigma_\ell = 1$ always makes $W_B < W_M$, and so unskilled agents always mix in equilibrium. This observation has important implications. Conditional on cheating, agents are more likely to be skilled, as their skill gives them a better chance to hide the

fact that they cheated. That is, skilled agents have a better chance to recoup. In this light, it is perhaps not surprising to see that employees or executives caught cheating are often in the upper portion of the skill distribution (e.g., executives at Enron, engineers at Volkswagen).

The following shows an even stronger result, namely that the equilibrium value of σ_ℓ is always such that (11) holds.

Corollary 4. *When (12) holds, the equilibrium strategies of the strategic agents, $\sigma_h = 1$ and σ_ℓ as in (13), are such that (11) holds.*

Through Lemma 5, this implies that, in equilibrium, we have $\pi_E(M) > \psi$, $\pi_S(M) < \phi(1 - \psi)$, $\pi_E(B) < \psi$, and $\pi_S(B) > \phi(1 - \psi)$. That is, while dropped projects improve the public perception of the agent's ethics, they also negatively affect the public perception of his skills. The opposite is true about failed projects.

The next result explores how various parameters of the model affect this tension between looking skilled and looking ethical for strategic agents.

Corollary 5. *Suppose that (12) holds. The equilibrium probability of cheating (σ_ℓ) by a strategic agent who observes $\tilde{p} = \ell = 0$ is decreasing in g, h, and ψ, and it is increasing in q. Also, σ_ℓ is increasing in ϕ if and only if $\frac{w_S}{w_E} > \bar{\omega}_\phi$ for some $\bar{\omega}_\phi \in \left(\frac{\psi}{1 - \phi(1 - \psi)}, 1 \right)$ derived in the proof.*

The fact that σ_ℓ decreases with ψ naturally complement our earlier result from Corollary 3 that σ_ℓ decreases with w_E. Intuitively, when combined, these two results mean that the presence of ethical agents reduces the unskilled agents' tendency to cheat, and does so more effectively when ethics are highly rewarded. The result that σ_ℓ is decreasing in h is related. It means that, when skilled agents become more difficult for unskilled agents to imitate (i.e., when $h - \ell = h$ increases), unskilled agents find it more beneficial to imitate ethical agents instead; to do so, they increase the frequency with which they drop projects. The effect from increasing ϕ is more subtle. For the unskilled strategic agents, it makes undertaking the project and failing (as they do with probability one) more attractive as a larger fraction of that pool will be skilled. At the same time, however, this reduces the number of unskilled agents in the pool of project droppers, increasing the posterior likelihood that such an outcome is associated with an ethical agent. The former effect dominates when w_S is large relative to w_E.

Ex ante, skills and ethics are uncorrelated but, as the following result shows, this is not the case ex post. Indeed, the fact that unethical agents do not adopt the same investment strategy as ethical agents and that their failure rate differs creates ex post correlation between skills and ethics.

Corollary 6. *The correlation between $\tilde{s}$ and $\tilde{\varepsilon}$ is strictly negative in the good and bad states ($\Omega = \mathrm{G}$ and $\Omega = \mathrm{B}$), and strictly positive in the medium state ($\Omega = \mathrm{M}$).*

This result is intuitive. The cheating of skilled strategic agents makes it more likely that undertaken projects, successful or not, are associated with skill, but less likely that they are associated with ethics. The fact that dropped projects result in a positive association between $\tilde{s}$ and $\tilde{\varepsilon}$ comes from the fact that skilled agents cheat (weakly) more than unskilled agents in equilibrium. Because ethical agents do not get sorted on skills, the proportion of skilled agents amongst the ethical agents stays at ϕ. In contrast, the cheating of strategic agents leaves more unskilled agents in the pool of dropped projects. Thus the lack of skills is more likely to be associated with a lack of ethics in this pool.

4 Two Periods and Endogenous Payoffs

So far we have assumed that the payoffs that the agent receives at the end of the period are exogenously specified. More specifically, we have assumed that his end-of-period payoff is linearly related to posterior beliefs that the public forms about his skills and ethics based on his observable performance in the first period. This, of course, was meant to capture the idea that firms value both the skills and ethics of their agents. We formalize this idea in this section by assuming that the agent can be hired for a second period. We proceed in two steps. First, we assume that the firm hiring the agent in the first period hires him again for an identical job in the second period. Second, we assume that the firm has the option to allocate the agent to a different job in the second period.

4.1 Same Job in Both Periods

Let us assume that the agent is initially hired by the firm to perform the same job in two periods. More specifically, the firm hires an agent to make investment decisions about two different and independent projects in two consecutive periods. At the beginning of each period, the agent is paid a competitive wage equal to the expected value that his presence and decisions bring to the firm.[7] In this calculation, we assume that the firm has easy access to unethical agents with low skills, so that the hired agent receives only the value that he adds relative to such an agent.

We also generalize the investment opportunity that the firm faces in each period to emphasize the separate roles of agent skills and ethics and the value these traits create for the firm. First, we increase the payoff of a successful project from one to $A \equiv 1 + a$, where $a \geq 0$. Second, we assume

---

[7]To micro-found this assumption, one could imagine that two similar firms compete for the services of the same agent in each of two periods.

that the project's investment cost of k consists of two components: one that is fixed at $f \in [0, k)$ and incurred at the very beginning of the period, before $\tilde{p}$ is observed and whether or not the project is undertaken; and a second portion of $k - f$ that is incurred, as before, only if the project is undertaken by the agent after he observes $\tilde{p}$. To ensure that the economic tradeoffs remain the same, we further assume that both a and f are sufficient close to zero (i.e., sufficiently close to the setup of previous sections) so that it is still optimal for the firm to undertake the project if and only if $\tilde{p} = g$. That is, we assume that $k - f \geq (1 + a)h$, which implies that projects with $\tilde{p} \neq g$ have a negative net present value, whether the agent is skilled or unskilled.

This more general specification for the firm's projects will allow us to better delineate the contribution of the agent's skills and ethics. Intuitively, an increase in f reduces the option value of a project and, since ethical agents are always perfectly aligned with the firm when they make option exercise decisions, decreases the value of ethics. An increase in a on the other hand makes every project that the firm undertakes more valuable. Since skilled (unethical) agents have a higher probability of leading a project to success than their unskilled counterparts, this means that an increase in a (potentially) corresponds to an increase in the value of skills. In fact, as we shall see, projects will be naturally indexed by the relative sizes of a and f.

Let $P_t^{s\varepsilon}$ denote the joint probability that, at the beginning of period $t \in \{1, 2\}$, the agent's skills and ethics are $\tilde{s} = s \in \{h, \ell\}$ and $\tilde{\varepsilon} = \varepsilon \in \{e, u\}$, respectively, based on the public information Ω_t that is available at that point:

$$P_t^{s\varepsilon} \equiv \Pr\{\tilde{s} = s, \tilde{\varepsilon} = \varepsilon \mid \Omega_t\}. \tag{14}$$

Since the only information that is publicly available at $t = 1$ is the prior distribution for $\tilde{s}$ and $\tilde{\varepsilon}$, which are independently distributed at the outset, we trivially have $P_1^{he} = \phi\psi$, $P_1^{hu} = \phi(1 - \psi)$, $P_1^{\ell e} = (1 - \phi)\psi$, and $P_1^{\ell u} = (1 - \phi)(1 - \psi)$. The values for $P_2^{s\varepsilon}$, of course, depend on the equilibrium decisions of the agent and investment outcomes in period 1, as before. Likewise, we use a t subscript to denote the random variables that apply to the project in each of the two periods (e.g., $\tilde{v}_t$, $\tilde{p}_t$).

Recall that three outcomes are possible for each project (in each period): successful investment, no investment, or failed investment. Let us denote the net cash flow from the project in each of these outcomes by $\tilde{x}_t \in \{A - k, -f, -k\}$. Let us focus on the second period first, as the expected wages in that period will determine the equilibrium strategy of the agent in the first period. Importantly, because wages get paid at the beginning of each period and because the game ends at the end of the second period, strategic agents cease no longer care about their skill and ethics reputation when they make investment decisions in the second period. This means that any private benefit $b > 0$, even if infinitesimally small, is sufficient to push the agent to invest in the second-period project, regardless of what he learns about the project's success probability, $\tilde{p}_t$. Ethical agents, on the other hand, invest in the project only when they learn that $\tilde{p}_2 = g$.

Given updated probabilities $P_2^{s\varepsilon}$ about the skill and ethics of the agent in the second period, the expected net cash flow from the second-period project is given by[8]

$$\mathbb{E}\big[\tilde{x}_2 \mid \Omega_2\big] = \big(P_2^{he} + P_2^{\ell e}\big)\big[q(Ag - k) - (1 - q)f\big] + P_2^{hu}\big[qAg + (1 - q)Ah - k\big] \\ + P_2^{\ell u}\big[qAg - k\big]. \quad (15)$$

Since $P_2^{\ell u} = 1 - P_2^{he} - P_2^{\ell e} - P_2^{hu}$ and since the project's expected net cash flow would be $\underline{x} \equiv qAg - k$ with an agent known to be unskilled and unethical, the expected value added by the agent is

$$\mathbb{E}\big[\tilde{w}_2 \mid \Omega_2\big] \equiv \mathbb{E}\big[\tilde{x}_2 - \underline{x} \mid \Omega_2\big] = \big(P_2^{he} + P_2^{\ell e}\big)(1 - q)(k - f) + P_2^{hu}(1 - q)Ah. \quad (16)$$

This makes intuitive sense. The first term comes from the fact that ethical agents save their firm the project's optional investment cost of $k - f$ when $\tilde{p}_2 \neq g$. The second term comes from the fact that skilled unethical agents improve the expected cash flows from the project from $A\ell = 0$ to Ah when $\tilde{p}_2 \neq g$.

A comparison of (16) with (7) shows that this two-period model is equivalent to the one-period model of sections 2 and 3 with $w_{\mathrm{E}} = (1 - q)(k - f)$ and $w_{\mathrm{S}} = (1 - q)Ah$. In particular, we can use Proposition 2 with these quantities to describe the unique equilibrium that prevails in the first period.

Proposition 3. *Assume that the agent's private benefits from undertaking a project is an arbitrarily small quantity $b > 0$. If*

$$\frac{Ah}{k - f} > \frac{\psi}{1 - \phi(1 - \psi)}, \quad (17)$$

then the unique equilibrium in the first period is given by $\sigma_g = \sigma_h = 1$ and

$$\sigma_\ell = \frac{\phi\big[1 - qg - (1 - q)h\big]\big[(1 - \phi + \phi\psi)Ah - \psi(k - f)\big]}{(1 - \phi)\Big\{\big[1 - qg - (1 - q)h\big]\phi(1 - \psi)Ah + (1 - qg)\psi(k - f)\Big\}} \in (0, 1). \quad (18)$$

If (17) does not hold, the unique equilibrium is $(\sigma_g, \sigma_h, \sigma_\ell) = (1, 0, 0)$.

As before, the cheating of strategic agents intensifies as $\frac{w_{\mathrm{S}}}{w_{\mathrm{E}}}$ increases. With endogenized wages, we have $\frac{w_{\mathrm{S}}}{w_{\mathrm{E}}} = \frac{Ah}{k-f}$, and so we can readily conclude that cheating is increasing in A and f, and decreasing in k. That is, as the payoff for a successful project increases, the fact that skilled unethical agents lead projects to success more frequently makes them more valuable, and strategic agents undertake more projects in period one in order to look more skilled. Recall that $k - f$ measures the optional portion of the project's cost. As k decreases and as f increases, this option value decreases, and so does the value of ethical agents whose contribution to value comes exclusively

---

[8]Note that we keep assuming that $\ell = 0$ for this and subsequent calculations.

from their optimal use of $\tilde{p}_2$ to make investment decisions on behalf of the firm. This in turns leads to a smaller value of w_{E} and a reduced appeal for strategic agents to drop projects in an effort to look ethical in the first period. Interestingly, although the quality q of the project affects both w_{S} and w_{E}, the fact that they are both quantities are proportional to $1 - q$ implies that the impact of q on the equilibrium of period is still felt only through its effect on first-period frequencies, and not on equilibrium wages in the second period. In particular, we know from Corollary 5 that the extent of cheating is increasing in q.

4.2 Two Types of Job in the Second Period

We now consider the possibility that the firm can allocate the agent to one of two jobs in each of the two periods. In particular, we assume that agents known to be ethical are better suited for (i.e., create more value in) job $\mathcal{E}$, while agents known to be skilled and unethical are better suited for job $\mathcal{S}$.

4.2.1 Setup: The Two Jobs

The differences between the two jobs revolve around the project characteristics introduced in section 4.1. Specifically, we assume that job $\mathcal{E}$ has a payoff of one, an investment cost of k, which is entirely optional (i.e., it only has to be incurred if the agent chooses to undertake the project). Job $\mathcal{S}$, on the other hand, has a payoff of $A = 1 + a$, a non-optional investment cost of f, and an optional investment cost of $k - f$. In essence, therefore, job $\mathcal{S}$ requires more ex ante commitment, but also has a higher potential payoff. As in section 4.1, we assume that f is sufficiently small so that the project's net present value is negative when $\tilde{p}_t = h$, for $t \in \{1, 2\}$. That is, we restrict f in such a way that $k - f \geq (1 + a)h$.

We assume that the firm assigns the agent to a second-period job based on the information about his traits that first-period outcomes reveal.[9] Of course, as we will see, this second-period assignment affects the incentives of the agent in the first period and, in turn, the information contained in each of the possible first-period outcomes. For this two-job model to have any bite, it must be the case that the firm actually finds it optimal to assign the agent to each of the two jobs with positive probability. For example, if A is so large and f so small that job $\mathcal{S}$ becomes so appealing, then the two-job model reduces to the two-period model of section 4.1. Indeed, the firm then always finds it optimal to assign the agent to job $\mathcal{S}$ regardless of first-period outcomes; that is, ethical and skilled agents then both generate more value in job $\mathcal{S}$ than in job $\mathcal{E}$. To create an

---

[9]Job assignment will be irrelevant in the first period, as none of the parameters that describe first-period jobs (i.e., A, f, and k) affect the investment strategy of agents in that period. That is, the agent's strategy in period 1 is affected by A, f, and k only to the extent that these parameters affect the wages that the agent can expect from the second period.

interesting economic tradeoff, therefore, we must restrict a and f in such a way that neither job is excessively appealing in the second period.[10]

As in section 4.1, we endogenize second-period wages by assuming that the firm has access to a large pool of unskilled unethical workers. In particular, we assume that the job that is not filled by the agent hired by the firm is automatically filled by an unskilled unethical worker (i.e., one with $\tilde{s} = \ell = 0$ and $\tilde{\varepsilon} = u$). The following lemma calculates the value contribution of the agent when he is assigned to each of the two jobs in period 2. For the time being, we use exogenous values of $P_2^{s\varepsilon}$ to denote the probability that the agent is of skill s and ethics ε entering the second period. These quantities, as before, will be endogenized through the agent's equilibrium strategy and firm outcomes in the first period.

Lemma 6. *The expected value added by the agent in job* $\mathcal{E}$ *is*

$$\mathbb{E}\big[\tilde{w}_2^{\mathcal{E}} \mid \Omega_2\big] = (1-q)\Big[hP_2^{hu} + k(P_2^{he} + P_2^{\ell e})\Big]. \tag{19}$$

The expected value added by the agent in job $\mathcal{S}$ *is*

$$\mathbb{E}\big[\tilde{w}_2^{\mathcal{S}} \mid \Omega_2\big] = (1-q)\Big[(1+a)hP_2^{hu} + (k-f)(P_2^{he} + P_2^{\ell e})\Big]. \tag{20}$$

As in section 4.1, the calculation of (19) and (20) makes use of the fact that unethical agents always find it optimal to undertake the second-period project for any small $b > 0$. Intuitively, (20) captures the idea that jobs with a bigger investment f committed to the project ex ante (and thus with a lower option option value) also comes with a larger potential payoff. That is, the reduction in option value comes with an increase in potential payoff. Also note that (19) and (20) are both similar to (16). By putting more weight on $P_2^{he} + P_2^{\ell e}$, job $\mathcal{E}$ rewards ethics more than job $\mathcal{S}$. In contrast, job $\mathcal{S}$ puts more weight on P_2^{hu} and thus rewards skills more than job $\mathcal{E}$ does. Because the firm can assign the agent to either job, however, it is no longer possible to map the agent's compensation in period 2 to that assumed in (7) for the one-period model. This is because the option (in job assignment) that is available to the firm prevents a direct mapping to the one-period model and, as such, we can no longer a linear specification for the agent's payoff.

The calibration that we adopt to ensure that the firm indeed assigns the agent to a different job depending on first-period outcomes imposes that the firm would be indifferent between the two jobs if its information about the agent's traits in the second period corresponded to the prior information. That is, we consider pairs of a and f that make (19) and (20) equal when $P_2^{he} = \phi\psi$, $P_2^{\ell e} = (1 - \phi)\psi$, and $P_2^{hu} = \phi(1 - \psi)$. This results in the following relationship between a and f.

---

[10]We also restrict a and f in such a way that the agent's actions are the same in the first period regardless of the job he gets assigned to at the outset.

Assumption 1. *We assume that the payoff a and fixed cost f for job $\mathcal{S}$ satisfy*

$$a = \frac{\psi f}{h\phi(1-\psi)}. \tag{21}$$

With this assumption in hand, deviations from the priors resulting from first-period outcomes will lead the firm to allocate the agent to the one job in which he is expected to create more value.[11] Recall from section 3 that the agent is perceived as being more ethical and less skilled when he drops a project, and less ethical and more skilled when he undertakes one. Of course, such results are not automatic here, as the investment decisions of the agent in the first period will strategically internalize future job assignments. That is, at this point, we can only conjecture that these results will hold in equilibrium, and verify them to hold later.

4.2.2 Equilibrium Analysis

Let us therefore start our analysis by conjecturing an equilibrium in which $\sigma_g = 1$, $\sigma_h = 1$, and $\sigma_\ell \in (0,1)$. The following lemma shows how the firm allocates the agent in the second period based on each of the three possible outcomes, $\Omega \in \{\mathrm{G}, \mathrm{M}, \mathrm{B}\}$, in the first period.

Lemma 7. *Suppose that an equilibrium exists with $\sigma_g = 1$, $\sigma_h = 1$, and $\sigma_\ell \in (0,1)$. In this equilibrium, it is optimal for the firm to assign the agent to job $\mathcal{E}$ if he drops the first-period project, and to job $\mathcal{S}$ if he undertakes the first-period project, whether or not this project succeeds.*

A comparison of (19) and (20) is the first step to understanding this result. When the probabilities in these two equations correspond to the prior probabilities of each skill-ethics combination, we know that they are equal, by Assumption 1 on a and f. Thus, whether (19) or (20) is greater with the actual posterior probabilities depends on how $P_2^{he} + P_2^{\ell e}$ changes from its prior value relative to $P_2 hu$. Specifically, if

$$\frac{P_2^{he} + P_2^{\ell e}}{P_2^{hu}} > \frac{\psi}{\phi(1-\psi)}, \tag{22}$$

then job $\mathcal{S}$ is relatively more appealing in the second period than job $\mathcal{E}$, and vice versa when the inequality (22) holds (strictly) with the opposite sign.[12] Lemma 7 then amounts to showing that (22) holds for any value of $\sigma_\ell \in (0,1)$ when the agent drops the project in period 1, and that the opposite inequality holds otherwise.

---

[11]Note that, given (21), our earlier restriction that $k - f \geq (1+a)h$ is equivalent to

$$f \leq \frac{\phi(1-\psi)}{\psi + \phi(1-\psi)}(k - h).$$

In essence, therefore, we are mostly interested in small increases of a and f from zero, so that the optimal investment policy of the firm is still to invest if and only if $\tilde{p}_t = g$.

[12]The firm remains indifferent between the two jobs if $\frac{P_2^{he} + P_2^{\ell e}}{P_2^{hu}}$ remains equal to $\frac{\psi}{\phi(1-\psi)}$. However, the proof of Lemma 7 shows that this never happens.

Lemma 7 pins down the form of compensation that the agent can expect when he chooses to drop or undertake the project in the first period. To find an equilibrium with $\{\sigma_g, \sigma_h, \sigma_\ell\} = \{1, 1, \sigma_\ell,$ we must verify that all agents with $\tilde{p}_1 = g$ and $\tilde{p}_1 = h$ always prefer to undertake the project, and that agents with $\tilde{p}_1 = \ell$ play a mixed strategy between undertaking and dropping. This is the subject of the next proposition.

Proposition 4. *Assume that the agent's private benefits from undertaking a project is an arbitrarily small quantity $b > 0$. If*

$$\frac{h}{k} > \frac{\psi}{1 - \phi(1 - \psi)}, \tag{23}$$

or if (23) holds with the opposite inequality sign and $f > \bar{f}$ for some $\bar{f}$ derived in the proof, then there exists a first-period equilibrium with $\sigma_g = \sigma_h = 1$ and $\sigma_\ell = \sigma_\ell^* \in (0, 1)$.

Condition (23) is the same as (17), but with $a = 1$ and $f = 0$. As before, it says that if the incentives to imitate the ethical agents are sufficiently weak (i.e., ψ and k are sufficiently small), then an inefficient equilibrium with cheating prevails. Interestingly, with two jobs, even if this condition is not satisfied, making the two jobs sufficiently distinct (by increasing f and a subject to (21)) also results in the same inefficient equilibrium.

As before the extent of cheating in this equilibrium is measure by σ_ℓ^*, the mixing probability of unskilled unethical types. However, although the expression for σ_ℓ^* is obtained in closed form,[13] this expression is somewhat messy and does not lend itself to easy interpretation. The following result establishes that σ_ℓ^* is increasing in f. That is, as f and a increase while still satisfying (21), it is the case that cheating by strategic agents increases.

Corollary 7. *The mixing probability σ_ℓ^* derived in Proposition 4 is strictly increasing in f.*

This result says that unskilled agents have more of a tendency to cheat and undertake negative-NPV projects when job $\mathcal{E}$ and job $\mathcal{S}$ greatly differ. This is due to the fact that as f and a increase, the firm's option to allocate the agent to one of two jobs effectively protects it from the presence of unethical agents. Indeed, because it can allocate such agents to a job where their empire-building motives do not have much of a negative effect on firm value, it is willing to pay them more for that job. Realizing this, the unskilled unethical agents have more of a tendency to shoot for such jobs by undertaking more projects, even if the probability of success is $\tilde{p}_1 = \ell = 0$.

---

[13]It is the unique solution to a linear equation.

Appendix

Proof of Lemma 1. Using Bayes' rule, we have

$$\Pr\{\tilde{s} = h \mid \Omega, \tilde{\varepsilon} = e\} = \frac{\Pr\{\Omega \mid \tilde{s} = h, \tilde{\varepsilon} = e\} \Pr\{\tilde{s} = h \mid \tilde{\varepsilon} = e\}}{\Pr\{\Omega \mid \tilde{\varepsilon} = e\}}$$

$$= \frac{\Pr\{\Omega \mid \tilde{\varepsilon} = e\} \Pr\{\tilde{s} = h\}}{\Pr\{\Omega \mid \tilde{\varepsilon} = e\}} = \Pr\{\tilde{s} = h\} = \phi. \qquad \blacksquare$$

Proof of Lemma 2. We prove this result by contradiction. Suppose that $W_{\mathrm{G}} < W_{\mathrm{B}}$. Then, since $g > h > \ell$, we have

$$gW_{\mathrm{G}} + (1-g)W_{\mathrm{B}} < hW_{\mathrm{G}} + (1-h)W_{\mathrm{B}} < \ell W_{\mathrm{G}} + (1-\ell)W_{\mathrm{B}},$$

which implies that undertaking the project is most (least) attractive when $\tilde{p} = \ell$ (when $\tilde{p} = g$), and in turn that $\sigma_\ell \geq \sigma_h \geq \sigma_g$. If we can show that this strategy ranking for strategic agents implies that $W_{\mathrm{G}} \geq W_{\mathrm{B}}$, we will have a contradiction. Using (7), we have

$$W_{\mathrm{G}} - W_{\mathrm{B}} = \left(P_{\mathrm{G}}^{hu} - P_{\mathrm{B}}^{hu}\right)w_{\mathrm{S}} + \left(P_{\mathrm{G}}^{he} + P_{\mathrm{G}}^{\ell e} - P_{\mathrm{B}}^{he} - P_{\mathrm{B}}^{\ell e}\right)w_{\mathrm{E}}. \qquad (A1)$$

We first show that the second term in this last equation is always positive; that is, with the above strategies, the agent is more likely to be ethical after a successful project than after a failed project. Using (4), we have

$$P_{\mathrm{G}}^{he} + P_{\mathrm{G}}^{\ell e} = \frac{qg\psi}{P_{\mathrm{G}}}$$

and

$$P_{\mathrm{B}}^{he} + P_{\mathrm{B}}^{\ell e} = \frac{q(1-g)\psi}{P_{\mathrm{B}}},$$

where

$$P_{\mathrm{G}} = qg\left[\psi + (1-\psi)\sigma_g\right] + (1-q)(1-\psi)\left[h\sigma_h\phi + \ell\sigma_\ell(1-\phi)\right]$$

and

$$P_{\mathrm{B}} = q(1-g)\left[\psi + (1-\psi)\sigma_g\right] + (1-q)(1-\psi)\left[h(1-\sigma_h)\phi + \ell(1-\sigma_\ell)(1-\phi)\right].$$

Straightforward algebra shows that

$$P_{\mathrm{G}}^{he} + P_{\mathrm{G}}^{\ell e} - P_{\mathrm{B}}^{he} - P_{\mathrm{B}}^{\ell e} = \frac{q(1-q)\psi(1-\psi)}{P_{\mathrm{G}} + P_{\mathrm{B}}}\left[(g-h)\phi\sigma_h + (g-\ell)(1-\phi)\sigma_\ell\right] \geq 0.$$

Using this result and the fact that $w_{\mathrm{E}} \geq \phi w_{\mathrm{S}}$ in (A1), we have

$$W_{\mathrm{G}} - W_{\mathrm{B}} \geq \left(P_{\mathrm{G}}^{hu} - P_{\mathrm{B}}^{hu}\right)w_{\mathrm{S}} + \left(P_{\mathrm{G}}^{he} + P_{\mathrm{G}}^{\ell e} - P_{\mathrm{B}}^{he} - P_{\mathrm{B}}^{\ell e}\right)\phi w_{\mathrm{S}}. \qquad (A2)$$

23

We can use (4) again to calculate

$$P_{\mathrm{G}}^{hu} = \frac{\big[qg\sigma_g + (1-q)h\sigma_h\big]\phi(1-\psi)}{P_{\mathrm{G}}}$$

and

$$P_{\mathrm{B}}^{hu} = \frac{\big[q(1-g)\sigma_g + (1-q)(1-h)\sigma_h\big]\phi(1-\psi)}{P_{\mathrm{B}}}.$$

Tedious but straightforward algebra shows that

$$\big(P_{\mathrm{G}}^{he} + P_{\mathrm{G}}^{\ell e}\big)\phi + P_{\mathrm{G}}^{hu} - \big(P_{\mathrm{B}}^{he} + P_{\mathrm{B}}^{\ell e}\big)\phi - P_{\mathrm{B}}^{hu} = \frac{(1-q)\phi(1-\phi)(1-\psi)}{P_{\mathrm{G}} + P_{\mathrm{B}}} \times$$
$$\Big(q\big[\psi + (1-\psi)\sigma_g\big]\big[(g-\ell)\sigma_\ell - (g-h)\sigma_h\big] + (1-q)(1-\psi)(h-\ell)\sigma_h\sigma_\ell\Big).$$

Because $g > h > \ell$ and $\sigma_\ell \geq \sigma_h$, this last expression is nonnegative which, through (A2), implies that $W_{\mathrm{G}} \geq W_{\mathrm{B}}$, the contradiction that we were looking for. ∎

Proof of Lemma 3. Let us first establish that W_{M} can never be strictly greater than both $W_{\mathrm{G}} + b$ and $W_{\mathrm{B}} + b$. Suppose it were the case. Then, since $pW_{\mathrm{G}} + (1-p)W_{\mathrm{B}} + b < W_{\mathrm{M}}$, strategic agents never undertake the project. This means that only ethical agents ever undertake the project, and so the probability that the agent is ethical given project success ($\Omega = \mathrm{G}$) or project failure ($\Omega = \mathrm{B}$) is one: $\Pr\{\tilde{\varepsilon} = e \,|\, \Omega = \mathrm{G}\} = \Pr\{\tilde{\varepsilon} = e \,|\, \Omega = \mathrm{B}\} = 1$. Also, since no strategic agents ever undertake the project, we have $\Pr\{\tilde{s} = h, \tilde{\varepsilon} = u \,|\, \Omega = \mathrm{G}\} = \Pr\{\tilde{s} = h, \tilde{\varepsilon} = u \,|\, \Omega = \mathrm{B}\} = 0$. Thus the compensation that the agent can expect to receive from project success or project failure is $W_{\mathrm{G}} = W_{\mathrm{B}} = w_{\mathrm{E}}$. It is then also the case that the skill of strategic agents never affects their decisions and outcomes; thus $\Pr\{\tilde{s} = h, \tilde{\varepsilon} = u \,|\, \Omega = \mathrm{M}\} = \phi(1-\psi)$. Finally, since some ethical agents undertake projects whereas no skilled agents do so, it must be the case that the fraction of ethical agents in the pool of dropped projects is lower than the prior probability that an agent is ethical: $\Pr\{\tilde{\varepsilon} = e \,|\, \Omega = \mathrm{M}\} < \psi$. Thus, $W_{\mathrm{M}} < \phi(1-\psi)w_{\mathrm{S}} + \psi w_{\mathrm{E}}$ and we have

$$W_{\mathrm{G}} - W_{\mathrm{M}} = W_{\mathrm{B}} - W_{\mathrm{M}} > w_{\mathrm{E}} - \phi(1-\psi)w_{\mathrm{S}} - \psi w_{\mathrm{E}} = (1-\psi)(w_{\mathrm{E}} - \phi w_{\mathrm{S}}) \geq 0.$$

This further implies that W_{M} is strictly smaller than both W_{G} and W_{B}, and so strictly smaller than both $W_{\mathrm{G}} + b$ and $W_{\mathrm{B}} + b$. This is a contradiction.

To see that W_{M} cannot be strictly smaller than both $W_{\mathrm{G}} + b$ and $W_{\mathrm{B}} + b$ when b is small, assume that b is close to zero and suppose that $W_{\mathrm{M}} < W_{\mathrm{B}} + b < W_{\mathrm{G}} + b$. Since $pW_{\mathrm{G}} + (1-p)W_{\mathrm{B}} + b > W_{\mathrm{M}}$, it is then the case that strategic agents always undertake the project. This also means that only ethical agents ever drop the project; thus $\Pr\{\tilde{\varepsilon} = e \,|\, \Omega = \mathrm{M}\} = 1$, $\Pr\{\tilde{s} = h, \tilde{\varepsilon} = e \,|\, \Omega = \mathrm{M}\} = 0$, and $W_{\mathrm{M}} = w_{\mathrm{E}}$. When strategic agents always undertake the project, we have $\sigma_g = \sigma_h = \sigma_\ell = 1$,

and so (3b) yields $\Pr\{B \mid \tilde{s} = h, \tilde{\varepsilon} = u\} = q(1-g) + (1-q)(1-h)$ and $\Pr\{B \mid \tilde{s} = \ell, \tilde{\varepsilon} = u\} = q(1-g) + (1-q)(1-\ell)$. We can use these expressions, along with (2b), in (4) and (7) to get, after simplifications,

$$W_{\mathrm{B}} = \frac{\big[q(1-g) + (1-q)(1-h)\big]\phi(1-\psi)w_{\mathrm{S}} + q(1-g)\psi w_{\mathrm{E}}}{q(1-g) + (1-q)(1-\psi)\big[(1-h)\phi + (1-\ell)(1-\phi)\big]}.$$

Therefore,

$$W_{\mathrm{M}} - W_{\mathrm{B}} = w_{\mathrm{E}} - \frac{\big[q(1-g) + (1-q)(1-h)\big]\phi(1-\psi)w_{\mathrm{S}} + q(1-g)\psi w_{\mathrm{E}}}{q(1-g) + (1-q)(1-\psi)\big[(1-h)\phi + (1-\ell)(1-\phi)\big]}$$

$$= \frac{q(1-g)(1-\psi)(w_{\mathrm{E}} - \phi w_{\mathrm{S}}) + (1-q)(1-\psi)\big[(1-h)\phi(w_{\mathrm{E}} - w_{\mathrm{S}}) + (1-\ell)(1-\phi)w_{\mathrm{E}}\big]}{q(1-g) + (1-q)(1-\psi)\big[(1-h)\phi + (1-\ell)(1-\phi)\big]}$$

which, since $w_{\mathrm{E}} \geq \phi w_{\mathrm{S}}$, is strictly greater than zero. Thus, as long as $b \leq \bar{b}$ where $\bar{b}$ is defined as the expression for $W_{\mathrm{M}} - W_{\mathrm{B}}$ above, we have $W_{\mathrm{M}} \geq W_{\mathrm{B}} + b$, a contradiction. In that case, $W_{\mathrm{G}} + b \geq W_{\mathrm{M}} \geq W_{\mathrm{B}} + b$. Otherwise, it is indeed the case that $W_{\mathrm{G}} + b \geq W_{\mathrm{B}} + b \geq W_{\mathrm{M}}$. ∎

Proof of Lemma 4. From Lemma 3, we know that $W_{\mathrm{G}} + b \geq W_{\mathrm{B}} + b > W_{\mathrm{M}}$ when $b > \bar{b}$. In that case, strategic agents clearly undertake the project when $\tilde{p} = g$ as $gW_{\mathrm{G}} + (1-g)W_{\mathrm{B}} + b > W_{\mathrm{M}}$. Thus, we concentrate on the case in which $b \leq \bar{b}$, which implies that $W_{\mathrm{G}} + b \geq W_{\mathrm{M}} \geq W_{\mathrm{B}} + b$.

We prove the result by contradiction. Suppose that $\sigma_g < 1$; then either $\sigma_g \in (0,1)$ or $\sigma_g = 0$. Suppose it is the latter. Then, because $W_{\mathrm{G}} + b \geq W_{\mathrm{M}} \geq W_{\mathrm{B}} + b$, we must also have $\sigma_h = \sigma_\ell = 0$. But then only ethical agents ever undertake the project. As we show in the first part of the proof of Lemma 3, this leads to $W_{\mathrm{G}} = W_{\mathrm{M}} = w_{\mathrm{E}}$, $W_{\mathrm{M}} < \phi(1-\psi)w_{\mathrm{S}} + \psi w_{\mathrm{E}}$, and to W_{M} being strictly smaller than both W_{G} and W_{B} (and so to $W_{\mathrm{G}} + b$ and $W_{\mathrm{B}} + b$ as well). This contradicts $\sigma_g = \sigma_h = \sigma_\ell = 0$ as strategic agents would then all prefer to always undertake the project.

Suppose instead that $\sigma_g \in (0,1)$. We must consider two scenarios: (i) W_{G} is strictly greater than W_{B}; (ii) W_{G} is equal to W_{B}. In (i), it must be the case that $\sigma_h = \sigma_\ell = 0$ as, if the agent is indifferent between undertaking or dropping the project when $\tilde{p} = g$, then he must prefer to drop it when $\tilde{p} = h < g$ or $\tilde{p} = \ell < g$. Again however, since fewer unethical agents undertake the project when $\tilde{p} = g$ than ethical agents (i.e., a fraction σ_g of unethical agents versus all ethical agents), the payoffs from undertaking the projects (W_{G} and W_{B}) are greater than that from dropping the project (W_{M}). This leads to $\sigma_h = \sigma_\ell = 1$, a contradiction.

Finally, in (ii), we have $W_{\mathrm{G}} + b = W_{\mathrm{M}} = W_{\mathrm{B}} + b$. In this case, σ_h and σ_ℓ can also take any value in $[0,1]$. However, tedious but straightforward calculations show that there is no combination of $\sigma_g \in (0,1)$, $\sigma_h \in [0,1]$, and $\sigma_\ell \in [0,1]$ leading to $W_{\mathrm{G}} + b = W_{\mathrm{M}} = W_{\mathrm{B}} + b$. ∎

Proof of Corollary 1. These results are obtained by using (2a)-(2c) and (3a)-(3c) with $\sigma_g = 1$ in (4), and simplifying. ∎

Proof of Proposition 1. We know from Lemma 4 that, in equilibrium, it must be the case that $\sigma_g = 1$. To show this result, we proceed in four steps. First, we verify that $(\sigma_g, \sigma_h, \sigma_\ell) = (1, 0, 0)$ is an equilibrium with $W_G = W_M = W_B$. Second, we show that there are no other equilibria with $\sigma_g = 1$ and $W_G = W_M = W_B$. Third, we show that for any equilibrium with $W_G > W_M > W_B$, it must be the case that $\sigma_h = 1$. Finally, we show that there is a unique $\sigma_\ell^* \in (0, 1)$ such that $(\sigma_g, \sigma_h, \sigma_\ell) = (1, 1, \sigma_\ell^*)$ is an equilibrium with $W_G > W_M > W_B$.

Suppose that strategic agents, like ethical agents, undertake the project if and only $\tilde{p} = g$; that is, suppose that $(\sigma_g, \sigma_h, \sigma_\ell) = (1, 0, 0)$. Since ethical and strategic agents make the exact same decisions and since decisions and outcomes do not depend on their skills, the ex post probability of each type given any outcome $\Omega \in \{G, M, B\}$ is equal to the prior probability of each type. In particular, $\pi_E(\Omega) = \psi$ and $\pi_S(\Omega) = \phi(1 - \psi)$ for all $\Omega \in \{G, M, B\}$. This means that $W_G = W_M = W_B = \psi w_E + \phi(1 - \psi)w_S$. Since the wages that strategic agents can expect from any outcome is the same, there is no benefit to deviating from the postulated strategies. Thus $(\sigma_g, \sigma_h, \sigma_\ell) = (1, 0, 0)$ is an equilibrium.

Similarly, $\sigma_g = 1$, $\sigma_h \in [0, 1]$, and $\sigma_\ell \in [0, 1]$ would be an equilibrium with one of σ_h or σ_ℓ strictly greater than zero if we can find values of σ_h and σ_ℓ that make W_G, W_M, and W_B equal. However, tedious but straightforward calculations shows that there is no combination of $\sigma_h \in [0, 1]$ and $\sigma_\ell \in [0, 1]$ other than $\sigma_h = \sigma_\ell = 0$ that leads to $W_G = W_M = W_B$.

If any other equilibrium exists, Lemma 3 tells us that it must be the case that $W_G > W_M > W_B$. This also means that $\sigma_g > \sigma_\ell$ and, since $\sigma_g = 1$, at most one of σ_h and σ_ℓ can be strictly between zero and one. Let us consider the possibility that $\sigma_g \in (0, 1)$ and $\sigma_\ell = 0$. This implies that a skilled strategic agent observing $\tilde{p} = h$ is indifferent between undertaking or dropping the project; that is, σ_h must satisfy

$$hW_G + (1 - h)W_B = W_M, \tag{A3}$$

where we can use Corollary 1 with $\sigma_\ell = 0$ to calculate $W_G = \pi_S(G)w_S + \pi_E(G)w_E$, $W_B = \pi_S(B)w_S + \pi_E(B)w_E$, and $W_M = \pi_S(M)w_S + \pi_E(M)w_E$. The solution to (A3) can, after some algebraic manipulations, be shown to be

$$\sigma_h = -\frac{q\big[(g - h)^2 + g\big]}{h(1 - h)\phi(1 - \psi)}.$$

Since this quantity is strictly negative, the postulated equilibrium fails to hold. Hence, we must have $\sigma_g = 1$ if an equilibrium other than $(\sigma_g, \sigma_h, \sigma_\ell) = (1, 0, 0)$ holds.

Finally, let us conjecture that $\sigma_g = \sigma_h = 1$. First, we rule out an equilibrium that would have $\sigma_\ell = 0$. Suppose that an equilibrium with such strategies holds. We can use $\sigma_g = \sigma_h = 1$ and $\sigma_\ell = 0$ in Corollary 1 to obtain

$$W_{\mathrm{G}} = \frac{\big[qg + (1-q)h\big]\phi(1-\psi)w_{\mathrm{S}} + qg\psi w_{\mathrm{E}}}{qg + (1-q)h\phi(1-\psi)},$$

$$W_{\mathrm{B}} = \frac{\big[q(1-g) + (1-q)(1-h)\big]\phi(1-\psi)w_{\mathrm{S}} + q(1-g)\psi w_{\mathrm{E}}}{q(1-g) + (1-q)(1-h)\phi(1-\psi)}, \quad \text{and}$$

$$W_{\mathrm{M}} = \frac{\psi w_{\mathrm{E}}}{1 - \phi(1-\psi)}.$$

Tedious but straightforward calculations show that $gW_{\mathrm{G}} + (1-g)W_{\mathrm{B}} > W_{\mathrm{M}}$, $hW_{\mathrm{G}} + (1-h)W_{\mathrm{B}} > W_{\mathrm{M}}$, and $\ell W_{\mathrm{G}} + (1-\ell)W_{\mathrm{B}} > W_{\mathrm{M}}$ if and only if (12) holds. This means that if a strategic agent finds it optimal to undertake (drop) the project with $\tilde{p} = \ell$, it is also the case that a strategic finds it optimal to undertake (drop) the project with $\tilde{p} = \ell$. This rules out an equilibrium with $(\sigma_g, \sigma_h, \sigma_\ell) = (1, 1, 0)$.

An equilibrium with $\sigma_g = \sigma_h = 1$ and $\sigma_\ell \in (0, 1)$ has the unskilled agents who observe $\tilde{p} = \ell$ mix between undertaking and dropping the project. In this case, the use of Corollary 1 leads to

$$W_{\mathrm{G}} = \frac{\big[qg + (1-q)h\big]\phi(1-\psi)w_{\mathrm{S}} + qg\psi w_{\mathrm{E}}}{qg + (1-q)(1-\psi)\big[h\phi + \sigma_\ell\ell(1-\phi)\big]},$$

$$W_{\mathrm{B}} = \frac{\big[q(1-g) + (1-q)(1-h)\big]\phi(1-\psi)w_{\mathrm{S}} + q(1-g)\psi w_{\mathrm{E}}}{q(1-g) + (1-q)(1-\psi)\big[(1-h)\phi + \sigma_\ell(1-\ell)(1-\phi)\big]}, \quad \text{and}$$

$$W_{\mathrm{M}} = \frac{\psi w_{\mathrm{E}}}{1 - \phi(1-\psi) - \sigma_\ell(1-\phi)(1-\psi)}.$$

The equilibrium σ_ℓ must satisfy

$$\ell W_{\mathrm{G}} + (1-\ell)W_{\mathrm{B}} = W_{\mathrm{M}}. \tag{A4}$$

Since W_{G} and W_{B} are both strictly decreasing in σ_ℓ, the left-hand side of (A4) is strictly decreasing in σ_ℓ. Since W_{M} is strictly increasing in σ_ℓ, there can be at most one value of σ_ℓ that makes (A4) hold. As shown above, when $\sigma_\ell = 0$, we have $\ell W_{\mathrm{G}} + (1-\ell)W_{\mathrm{B}} > W_{\mathrm{M}}$ if and only if (12) holds. It can also be easily shown that, when $\sigma_\ell = 1$, we have $\ell W_{\mathrm{G}} + (1-\ell)W_{\mathrm{B}} < W_{\mathrm{M}}$. Therefore, if (12) holds, there is a unique $\sigma_\ell^* \in (0, 1)$ that makes (A4) hold. This completes the proof. ∎

References

Akerlof, George A., 1982, "Labor Contracts as Partial Gift Exchange," *Quarterly Journal of Economics*, 97(4), 543-569.

Bateman, Thomas S., and Dennis W. Organ, 1983, "Job Satisfaction and the Good Soldier: The Relationship between Affect and Employee 'Citizenship'," *The Academy of Management Journal*, 26(4), 587-595.

Becht, Marco, Patrick Bolton, and Ailsa Röell, 2003, "Corporate Governance and Control," in *The Handbook of Economics and Finance*, Volume 1, Part A, George M. Constantinides, Milton Harris and René M. Stulz (Eds.), Elsevier, 1-109.

Bénabou, Roland, and Jean Tirole, 2016a, "Bonus Culture: Competitive Pay, Screening, and Multitasking," *Journal of Political Economy*, 124(2), 305-370.

Bénabou, Roland, and Jean Tirole, 2016b, "Mindful Economics: The Production, Consumption, and Value of Beliefs," *Journal of Economic Perspectives*, 30(3), 141-164.

Berle, Adolph, and Gardiner Means, 1932, *The Modern Corporation and Private Property*, Macmillan: New York, NY.

Carlin, Bruce, and Simon Gervais, 2009, "Work Ethic, Employment Contracts, and Firm Value," *Journal of Finance*, 64(2), 785-821.

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

Etzioni, Amitai, 1988, *The Moral Dimension: Toward a New Economics*, Macmillan: New York, NY.

Fama, Eugene F., 1980, "Agency Problems and the Theory of the Firm," *Journal of Political Economy*, 88(2), 288-307.

Gibbons, Robert, and John Roberts, 2012, "Economic Theories of Incentives in Organizations," in *The Handbook of Organizational Economics*, Robert Gibbons and John Roberts (Eds.), Princeton University Press, 56-99.

Harris, Milton, and Bengt Holmström, 1982, "A Theory of Wage Dynamics," *Review of Economic Studies*, 49(3), 315-333.

Holmström, Bengt, 1979, "Moral Hazard and Observability," *Bell Journal of Economics*, 10(1), 74-91.

Holmström, Bengt, 1999, "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66(1), 169-182.

Holmström, Bengt, and Joan Ricart i Costa, 1986, "Managerial Incentives and Capital Management," *Quarterly Journal of Economics*, 101(4), 835-860.

Jaros, Stephen, 2007, "Meyer and Allen Model of Organizational Commitment: Measurement Issues," *ICFAI Journal of Organizational Behavior*, 6, 7-25.

Jensen, Michael C., and William H. Meckling, 1976, "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics*, 3(4), 305-360.

Kaplan, Steven N., and Morten Sørensen, 2016, "Are CEOs Different? Characteristics of Top Managers," Working Paper, University of Chicago.

Meyer, John P., and Natalie J. Allen. 1997, *Commitment in the Workplace: Theory, Research, and Application*, Sage Publications.

Meyer, John P., David J. Stanley, Lynne Herscovitch, and Laryssa Topolnytsky, 2002, "Affective, Continuance, and Normative Commitment to the Organization: A Meta-analysis of Antecedents, Correlates, and Consequences", *Journal of Vocational Behavior*, 61(1), 20-52.

Noe, Thomas H., and Michael J. Rebello, 1994, "The Dynamics of Business Ethics and Economic Activity," *American Economic Review*, 84(3), 531-547.

Pavesi, Filippo, and Massimo Scotti, 2017, "Good Lies," Working Paper, University of Technology Sydney.

Rabin, Matthew, 1995, "Moral Preferences, Moral Constraints, and Self-Serving Biases," Working Paper, University of California at Berkeley.

Ross, Stephen A., 1973, "The Economic Theory of Agency: The Principal's Problem," *American Economic Review*, 63(2), 134-139.

Sen, Amartya, 1997, "Maximization and the Act of Choice," *Econometrica*, 65(4), 745-779.

Shleifer, Andrei and Robert W. Vishny, 1997, "A Survey of Corporate Governance," *Journal of Finance*, 52(2), 737-783.

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

Williams, Larry J., and Stella E. Anderson, 1991, "Job Satisfaction and Organizational Commitment as Predictors of Organizational Citizenship and In-Role Behaviors," *Journal of Management*, 17(3), 601-617.