Transparency and Innovation in Organizations

Osamu Hayashida,* Kimiyuki Morita,† and Takeharu Sogo‡

March 26, 2022

Abstract

We study the effects of organizational transparency on the incentives to innovate. In each of the two periods, an agent chooses between a new idea and a well-established idea and exerts effort to develop that idea. While exerting effort, the agent acquires knowledge, which is partially wasted if the idea is switched. We show that transparency regarding interim performance measures promotes idea exploration, but it is counter-productive if the output relies heavily on effort over idea quality. Perhaps surprisingly, transparency can be further counter-productive if the acquired knowledge becomes less idea-specific, or if the interim performance measure becomes more precise.
JEL-Classification: D80; D83; D86; M50; M54 Keywords: Transparency; Innovation; Employee Development

*Faculty of Business Administration, Osaka University of Economics, Japan, hayasida@osaka-ue.ac.jp

†School of Economics, Senshu University, Japan, kimiyuki1229@gmail.com

‡SKEMA Business School—Université Côte d'Azur, France, takeharu.sogo@skema.edu

1 Introduction

Although transparency in organizations has drawn much attention in both business and economics[1] and its benefits have been well documented,[2] it is not always beneficial. Eulogy, a London-based communications agency, increased transparency by inviting its clients to brainstorming sessions, where the creative team would stop pursuing early-stage ideas if they received negative reactions from the clients. This transparency hurt the motivation to innovate. The CEO Adrian Brady explains, "A client's immediate negative reaction to a potentially great idea can end a conversation before it takes flight, making it hard to do anything big or new."[3]

Transparency may weaken the incentive to implement new ideas if they are not expected to be adopted, as in the case of Eulogy. R&D teams may not try hard to translate a new idea from research findings into products if they anticipate switching to another idea. Employees may not work hard to learn a new method if they anticipate switching to another method.

Innovation is recognized in management literature to consist of two stages: idea generation and idea implementation (Anderson et al., 2014). The idea implementation is the process of converting new ideas into new and improved products, services, or ways of doing things (Baer, 2012). Since "ideas are useless unless used" (Levitt, 1963, p.79), the idea implementation is essential for innovation to create values. Despite the importance of the idea implementation and the widespread use of organizational transparency, few investigate how they interact.

We study the interaction between transparency and the incentive for idea im-

[1]The number of articles in major business journals mentioning transparency increased more than tenfold between 1990–93 and 2006–2009 (Schnackenberg and Tomlinson, 2016).

[2]For example, Reed Hasting, the CEO of Netflix, emphasizes the importance of transparency: "It's up to the leader to live the message of transparency by sharing as much as possible with everybody. Big thing, small thing, whether or good or bad—if your first instinct is to put most information out there, others will do the same. As Netflix, we call this 'sunshining,' and we make an effort to do a lot of it" (Hastings and Meyer, 2020, p.105).

[3] https://www.mckinsey.com/business-functions/organization/our-insights/the-dark-side-of-transparency.

plementation to examine when transparency is (un)desirable for organizations. In particular, we analyze the following two-period principal-agent model. The principal (she) first selects either a "transparent" or "opaque" organization.[4] In the transparent organization, she commits ex ante to make interim performance measures observable within the organization, but these measures are not contractible. In the opaque organization, she commits ex ante to make them unobservable. In each period, the agent (he) adopts either a "new" idea or a "known" idea, and privately exerts effort to implement the adopted idea. The outcome depends on the quality of the adopted idea and the effort. While the quality of the known idea is certain, that of the new idea is uncertain, but their expected qualities are the same.

Our model has two features that are common to innovation processes. First, while making more effort to implement an adopted idea in the first period, the agent acquires more knowledge that can be utilized in the second period. Second, the acquired knowledge is idea-specific, so the agent cannot fully utilize it if he uses an idea that is different from that in the first period. These features are consistent with the observation in Holmstrom (1989) that most innovations are labor-intensive and idiosyncratic. Since each innovation is idiosyncratic, the effort and the acquired knowledge are (partially) wasted once the idea is switched.

A key insight from the model is that transparency creates the tension between idea sorting and innovation incentives. When a new idea is explored in the first period, transparency reveals the idea quality, and may lead to the switching of an idea. The advantage of this is that it improves the expected idea quality through better sorting—the agent can continue pursuing the new idea if receiving a good signal, otherwise, he can switch to a known idea. We call this the *sorting effect* of transparency. The disadvantage is that upon receiving a bad signal, the agent can change his ideas, thereby wasting his efforts (acquired knowledge) to some degree. We call this the *wasting effect*. Furthermore, the possibility of a waste of efforts hurts his effort incentive to implement a new idea in the first period, as in the case of Eulogy. We call this the *demotivating effect*.

[4]We assume the principal is female and the agent is male. It is for identification only.

Contracts contingent on idea choice or (interim/ex-post) performance measures could induce the agent's proper behavior (e.g., overcoming the commitment problem regarding switching ideas). However, it is difficult for the principal to write such contingent contracts for the following reasons. First, most innovation activities cannot be planned and contracted upon ex ante. Second, employees are often the ones who generate ideas and implement them, whether they are drastic new ideas or well-known incremental ideas, and the principal does not possess relevant ideas and performance measures ex ante. Reflecting these, regardless of whether an adopted idea is new or well known, we adopt the framework of incomplete contracts and ex-post bargaining as in Grossman and Hart (1986); Aghion and Tirole (1994).

We first identify when transparency is desirable for the principal. While transparency promotes the exploration of new ideas, it reduces the expected payoff to the principal if the output relies heavily on implementation effort rather than idea quality—that is, when the wasting and demotivating effects outweigh the sorting effect. The intuition is simple: Transparency may waste first-period efforts, and this possibility weakens effort incentives in the first period, which is too costly when the implementation effort is sufficiently important.

One might conjecture that these drawbacks of transparency can be mitigated, and the principal can always enjoy higher output if the acquired knowledge becomes less idea-specific, or if the interim performance measures become more precise.

However, this conjecture is false. We find that when the acquired knowledge becomes *less idea-specific*, or when the interim performance signal becomes *more precise*, the final output *decreases* if the innovation relies heavily on effort.

Let us first understand the effect of idea specificity. Indeed, an increase in the generality of the acquired knowledge reduces the wasting effect. However, it increases the switching probability and, hence, the probability of wasting efforts. When the output depends heavily on the effort, that is, when the agent has less incentive to risk wasting efforts to improve idea quality, the switching probability is so small that the former effect becomes sufficiently small relative to the latter. Thus, if the output relies heavily on effort, the more general the knowledge acquired, the smaller the incentive to make an effort, and hence, the smaller the final output.

Moreover, the logic regarding the signal precision is that the availability of a precise signal increases the benefit of the sorting effect, but this also increases the switching probability, thereby reducing effort incentives. When the output relies more heavily on implementation effort relative to idea quality, this negative effect dominates; therefore, a precise signal can hurt the principal.[5]

While transparency promotes the exploration of new ideas, we find that the agent is too inclined to adopt a new idea, from both the principal's and the social perspectives. If the agent adopts a new idea, he can save effort costs, since the demotivating effect reduces the first-period effort. The agent privately bears all effort costs while receiving only a portion of the return on effort. Therefore, from both the principal's and the social perspectives, the agent is overly inclined to explore a new idea because he overestimates the cost savings when adopting a new idea.

Our analysis provides guidance on how successful organizational transparency design should vary as per the three factors: the importance of idea qualities rather than the idea implementation, idea-specificity of the knowledge acquired during the idea implementation, and precision of interim performance signals. For a business whose performances rely heavily on the idea quality rather than the idea implementation, organizational transparency will be effective. This finding provides a theoretical foundation for companies, such as Netflix, to create a transparent organizational culture. Moreover, the benefit of transparency increases when the acquired knowledge is not too idea-specific, and precise interim performance signals are available (e.g., opinions from professionals or detailed customer feedback).

By contrast, for a business whose performance relies heavily on the idea implementation, transparency will weaken the incentive to implement ideas, and

[5]Although the setting is quite different from ours, this logic is related to Crémer (1995). In his dynamic contracting model, the availability of precise signals makes it less credible for the principal to commit to severe punishments, i.e., firing an underperformer, weakening effort incentives.

may, thus, be harmful. The harm would be greater if the acquired knowledge is not too idea-specific, and interim performance signals are more accurate (which is precisely when transparency is more beneficial for businesses relying on idea qualities).

Our incomplete contract approach complements the findings of Manso (2011) and Hellmann and Thiele (2011, hereafter HT) on optimal contracts that provide proper incentives for exploration.[6] Manso (2011) considers an agent who chooses between exploitative and explorative actions, and finds that the optimal incentive contract combines tolerance for early failure and reward for long-term success.[7] However, as noted in Manso (2011), incentive contracts may not be available, especially to low-level employees, because it is often hard to find verifiable performance measures. Our analysis is closest to HT. They consider a multitasking model between planned and unplanned tasks, where performance measures are available only for the planned (non-innovative) tasks, and the unplanned (innovative) tasks require ex-post bargaining. HT study the optimal strength of the incentives, that is, the optimal bonus size for the non-innovative task that achieves an appropriate balance between non-innovative and innovative tasks. In contrast, in our model, the agent engages only in unplanned tasks that require ex-post bargaining.

Our work relates to research on transparency within organizations.[8] Prat (2005) studies the effects of transparency when the action of workers are observable to

[6]Furthermore, managers might want to know what they can do without incentive contracts for their organizations to facilitate innovation because the adoption of high-powered incentives may have negative effects on innovation (See, e.g., Deci 1972 for an argument of how external rewards harm intrinsic motivation, Amabile et al. 1996 for a survey on creativity in psychology, and Onishi, Owan and Nagaoka 2021 for empirical evidence from corporate inventors in Japan that shows that stronger financial incentives, based on the commercial success of an invention, can be counter-productive).

[7]This finding is supported by the experimental evidence in Ederer and Manso (2013) and the empirical evidence from the academic life sciences (Azoulay, Graff Zivin and Manso, 2011).

[8]Research on transparency to outsiders includes Zhong (2018) and Brown and Martinsson (2019). Zhong (2018) empirically studies how transparency affects the relationship between R&D and managerial career concerns, while Brown and Martinsson (2019) empirically study how transparency to capital markets affects R&D activities. In these studies, transparency is defined by measures such as the use of international accounting standards, financial disclosures, and auditing activities.

their principal, while Cato and Ishihara (2017) study when the actions of workers are observable to other workers.[9] Jehiel (2015) finds that full transparency is suboptimal because the agent makes less effort upon receiving a bad signal about his productivity. In our model, although transparency hurts effort incentives as well, it is because the agent anticipates that he may switch his idea in the future—an effect not studied in Jehiel's static model.

Our analysis also provides similar guidance on feedback for employee development; our model of transparency and innovation can be interpreted as a model of employee development feedback. A subordinate chooses either the exploration of a new work method or the exploitation of a well-known method. The match quality between the employee and the work method (corresponding to the quality of an adopted idea) is uncertain, and the supervisor, who receives the interim performance signals, commits ex ante to giving either full feedback (corresponding to the transparent organization) or no feedback (corresponding to the opaque organization). With this reinterpretation, the role of feedback is to improve future performance of employees through performance management (e.g., Murphy et al., 2018, p.24).[10] While performance management has been popular worldwide (Aguinis and Pierce, 2008), it has reportedly not been very successful.[11]

An implication of our results for employee development feedback is that if employees' tasks rely heavily on effort rather than the match quality between the employee and the work method, employee development feedback can be harmful. Further, it can be more counter-productive when the acquired knowledge is more general and when the feedback is provided by a person who can receive more precise interim performance signals (e.g., specialists, not generalists).

Recent papers study feedback in relation to the purpose of development. Wirtz

[9]Bernstein (2012) conducts a field experiment in a mobile phone factory in China and finds that maintaining observability of factory workers induces unproductive hiding and undermines their performance.

[10]By contrast, traditional performance appraisal is usually an annual evaluation linked to salary administration and has a retrospective focus (Murphy et al., 2018, p.24).

[11]95% of managers are dissatisfied with their performance management, and 59% of employees feel performance management reviews are not worth the time invested (Pulakos et al., 2015).

(2016) studies the effects of such feedback in a tournament model with strategic experimentation, where each agent can make a technology choice. She finds that partial feedback can be superior to full feedback because full feedback enhances competition and leads to inefficient technology switching; that is, the leader (resp. the follower) of the tournament has excessive incentive to retain (resp. discard) a previously adopted technology. By contrast, the source of inefficient switching in our study is the waste of acquired knowledge, and the resulting reduced effort incentives—effects not considered in Wirtz (2016). Gross (2017) also studies feedback for product development in commercial logo design tournaments.

The remainder of the paper is organized as follows. Section 2 presents the model. Section 3 provides our main results when interim signals are perfect, and most of their proofs are delegated to Appendix A. Section 4 considers two extensions, showing that our main results continue to hold when interim signals are imperfect. All the proofs for Section 4 are delegated to Appendix B. Section 5 concludes.

2 Model

A principal (she) hires an agent (he) to execute a project over two periods. Both are risk-neutral and there is no discounting. The project outcome depends both on the quality of the idea and the effort. In each period $t \in \{1, 2\}$, the agent privately chooses an idea i_t from a *new* idea N and a *known* idea K, and makes effort e_t to implement the idea (e.g., converting an idea into products or serves) at the cost $c(e_t) = (1/2)(e_t)^2$. We assume that the agent chooses K if he is indifferent between $\{N, K\}$ or initiating N incurs an arbitrarily small setup cost.

Let $v(i_t)$ be the quality of idea i_t. The quality of the new idea $v(N)$ is θ, which is initially unknown to both the parties, but is commonly known to be independently distributed according to a cumulative distribution F_θ with p.d.f. f_θ. The quality of the known idea, $v(K)$, is deterministic and is commonly known to be $\mu > 0$. For simplicity, we assume that the expected quality of each idea is identical $E[\theta] = \mu$.

In period 1, the agent chooses $i_1 \in \{N, K\}$, and then exerts e_1, producing the *interim* output

$$x_1 = \gamma v(i_1) + (1-\gamma)e_1. \tag{1}$$

The parameter $\gamma \in (0,1)$ captures the relative importance of the idea quality over the effort for the project. When γ is low, the output is determined more by effort, which makes it less valuable to explore the new idea.

At the end of period 1, if the organization is transparent, both parties observe a signal s regarding the interim output x_1:

$$s = x_1 + \varepsilon,$$

where ε is a measurement error, which is independently distributed according to a distribution F_ε with p.d.f. f_ε. We can interpret the measurement error as the degree of transparency because in a more transparent organization, one can obtain a more accurate signal.

In period 2, given the available information, the agent chooses $i_2 \in \{N, K\}$ and then exerts effort e_2, yielding the *final* output

$$x_2 = \begin{cases} \gamma v(i_2) + (1-\gamma)(e_1 + e_2) & \text{if } i_1 = i_2 \text{ or the agent keeps the same idea,} \\ \gamma v(i_2) + (1-\gamma)(\rho e_1 + e_2) & \text{if } i_1 \neq i_2 \text{ or the agent switches ideas.} \end{cases} \tag{2}$$

If the agent switches ideas between periods 1 and 2, he can only partially utilize his past effort e_1. The parameter $\rho \in (0,1)$ measures how much e_1 contributes to the final output once the idea is switched. Upon switching, the contribution of e_1 to x_2 declines by $(1-\gamma)(1-\rho)e_1$, which we call the *wasting effect*. We interpret ρ as the *generality* of the knowledge or know-how acquired while exerting effort to develop an idea in period 1. When ρ is high, the knowledge and know-how acquired in period 1 is general, such that the wasting effect is smaller.

In (2), the marginal benefit of effort is independent from the quality of an idea; there is no complementarity between effort and idea quality. The idea choice and the agent's belief about the quality of the new idea θ do not affect the effort choice.

As detailed in the introduction, we use the incomplete contract framework and assume that the principal cannot write any contract contingent on the interim signal s or the final output x_2. After the final output x_2 is realized, both parties observe x_2 and bargain over their share. We assume that the agent has the bargaining power of $\lambda \in (0, 1)$ so that the agent receives λx_2 while the principal receives $(1 - \lambda)x_2$.

The principal designs the organizational form—she decides whether to make the organization *transparent* or *opaque*. In the transparent organization, the principal commits ex ante to make the interim signal s observable to both parties. In the opaque organization, the principal commits ex ante not to make it observable.

The timing of the game is summarized as follows:

1. The principal designs the organization to be either "transparent" or "opaque."
2. In period 1, the agent privately chooses $i_1 \in \{N, K\}$ and then exerts e_1.
3. The interim output x_1 is realized. Both parties observe signal $s = x_1 + \varepsilon$ only in the transparent organization.
4. In period 2, the agent privately chooses $i_2 \in \{N, K\}$ and then exerts e_2.
5. The final output x_2 is realized. The agent receives λx_2 and the principal receives $(1 - \lambda)x_2$.

3 Main Analysis

In this section, we establish our benchmark results in a simple setting where the quality of a new idea θ is uniformly distributed over $(0, 2\mu)$ so that $E[\theta] = \mu$ and the interim signal $s = x_1$ is perfect (i.e., $\varepsilon = 0$).

Section 4 considers two extensions to demonstrate that our key results continue to hold when (1) the new idea quality is uniformly distributed, as in the main analysis, and the measurement error ε is uniformly distributed and (2) both the new idea quality θ and the measurement error ε are normally and independently distributed.

3.1 Opaque Organization

We first consider the opaque organization. The agent does not update the information about the quality θ of the new idea N. Thus, the agent has no incentive to switch ideas due to the wasting effect and he is indifferent between N and K because both the ideas have the same expected quality. Hence, by our assumption, the agent keeps working with the known idea ($i_1 = i_2 = K$) and his problem is reduced to choose (e_1, e_2) jointly to maximize his expected payoff given $i_1 = i_2 = K$:

$$\pi(K) = \lambda[\gamma\mu + (1-\gamma)(e_1 + e_2)] - \frac{1}{2}(e_1)^2 - \frac{1}{2}(e_2)^2,$$

where we denote $\pi(i_1)$ as the agent's expected payoff when adopting $i_1 \in \{N, K\}$. The optimal effort in the opaque organization is given by

$$e^{\text{OP}} \equiv e_1^{\text{OP}} = e_2^{\text{OP}} = \lambda(1-\gamma),$$

which decreases in γ because the output is determined less by effort as γ increases. Let π^{OP} be the agent's expected equilibrium payoff in the opaque organization, i.e., $\pi^{\text{OP}} = \lambda[\gamma\mu + (1-\gamma)(2e^{\text{OP}})] - 2c(e^{\text{OP}})$.

3.2 Transparent Organization

We study the transparent organization in the following order: second-period effort and idea choice, and first-period effort and idea choice.

Second-period Effort Since there is no complementarity between the idea quality and the effort, regardless of the idea choice, the second-period effort is

$$e_2 = e^{\text{OP}} = \lambda(1-\gamma).$$

Second-period Idea Choice Given the first-period choices (i_1, e_1) and the available information, the agent chooses $i_2 \in \{N, K\}$. First, suppose the known idea is adopted in the first period ($i_1 = K$). Then, as in the opaque organization, the agent does not update information about the quality of the new idea θ; hence, it is optimal to keep working with the known idea ($i_2 = K$).

Next, suppose that the new idea is adopted in the first period ($i_1 = N$). Under the perfect signal assumption, after observing a performance signal $s = \gamma\theta + (1-\gamma)e_1$, knowing his own choice e_1, the agent is certain about the quality of the new idea θ. With the wasting effect, the agent switches to K if and only if θ is low enough such that sufficient improvement in idea quality is expected:

$$\lambda[\gamma\theta + (1-\gamma)(e_1 + e^{\text{OP}})] - c(e^{\text{OP}}) \leq \lambda[\gamma\mu + (1-\gamma)(\rho e_1 + e^{\text{OP}})] - c(e^{\text{OP}}).$$

Rearranging this, together with the above argument, yields the following lemma:

Lemma 1. *Suppose $\varepsilon = 0$. The optimal idea choice in $t = 2$ is given as follows:*

- *After adopting the known idea in $t = 1$ ($i_1 = K$), the agent keeps working with it in $t = 2$ ($i_2 = K$).*
- *After adopting the new idea in $t = 1$ ($i_1 = N$), the agent switches to the known idea ($i_2 = K$) if and only if $\theta \leq \hat{\theta}(e_1)$, where $\hat{\theta}(e_1)$ is given by*

$$\hat{\theta}(e_1) \equiv \mu - \frac{(1-\gamma)}{\gamma}(1-\rho)e_1. \tag{3}$$

Observe that $\hat{\theta}(e_1) < \mu$ for $e_1 > 0$. Since the agent can fully utilize his acquired knowledge by not switching ideas, he may not switch even after learning that the quality of the new idea is lower than that of the known idea.

Given first-period choices $(i_1 = N, e_1)$, the switching probability is higher when e_1 is smaller, and ρ and γ are greater ($\partial\hat{\theta}(e_1)/\partial\rho = (1-\gamma)/\gamma e_1 > 0$ and $\partial\hat{\theta}(e_1)/\partial\gamma = (1/\gamma^2)(1-\rho)e_1 > 0$). The reasoning is simple: As e_1 decreases or ρ increases, the wasting effect becomes smaller and, thus, raises the net benefit of switching (i.e., the improvement in idea quality minus the wasting effect); as

γ rises, the outcome is determined more by idea quality and less by effort, which raises the net benefit of switching.

Moreover, $\hat{\theta}(e_1)$ can be negative. In that case, the agent never switches. This occurs when γ is so small that the wasting effect outweighs the benefit of improving the idea quality by switching (as detailed in Lemma 2 below).

Note that the agent's idea choice in the second period is efficient in the sense that it maximizes the expected final output given the first-period choices (i_1, e_1) (because there is no complementarity between idea quality and effort). Therefore, our analysis remains the same even if the principal makes the idea choice in the second period.

First-period Effort Let $e_1^{\mathrm{TR}}(i_1)$ be the first-period effort after adopting the first-period idea $i_1 \in \{N, K\}$. First suppose the known idea is adopted in the first period ($i_1 = K$). Then, as described above, regardless of the first-period effort choice, the agent keeps working with it ($i_2 = K$). Thus, the optimal first-period effort is the same as that in the opaque organization: $e_1^{\mathrm{TR}}(K) = e^{\mathrm{OP}}$.

Next, suppose that the new idea is adopted in the first period ($i_1 = N$). Then, the agent's expected payoff $\pi(N)$ can be written as

$$\begin{aligned}\pi(N) &= \lambda E[x_2 \mid i_1 = N, e_1, \hat{\theta}(e_1), e^{\mathrm{OP}}] - c(e_1) - c(e^{\mathrm{OP}}) \\ &= \lambda F_\theta(\hat{\theta}(e_1))[\gamma\mu + (1-\gamma)(\rho e_1 + e^{\mathrm{OP}})] \\ &\quad + \lambda \int_{\hat{\theta}(e_1)}^{2\mu} [\gamma\theta + (1-\gamma)(e_1 + e^{\mathrm{OP}})] dF_\theta - c(e_1) - c(e^{\mathrm{OP}}). \qquad (4)\end{aligned}$$

The first-order condition yields the optimal first-period effort $e_1^{\mathrm{TR}}(N)$ upon $i_1 = N$:

$$e_1^{\mathrm{TR}}(N) = \lambda(1-\gamma) - (1-\rho)F_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N)))\lambda(1-\gamma), \qquad (5)$$

where the left-hand side is the marginal cost from exerting $e_1^{\mathrm{TR}}(N)$, and the right-hand side is the corresponding marginal benefit. The second term is the *demotivating effect* of transparency, capturing a reduction in the marginal benefit of effort due to the wasting effect. Without this, (5) reduces to the corresponding condition for the opaque organization. Thus, $e_1^{\mathrm{TR}}(N) < e^{\mathrm{OP}}$ if the demotivating effect is positive, i.e., $F_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N))) > 0$. Moreover, an increase in the switching probability $F_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N)))$ increases the demotivating effect and thus reduces the first-period effort. We now establish:

Lemma 2. *If the known idea K is chosen in $t = 1$, the optimal first-period effort in the transparent organization is $e_1^{\mathrm{TR}}(K) = e^{\mathrm{OP}}$. If the new idea N is chosen in $t = 1$, the optimal first-period effort and the induced switching behavior are as follows:*

- *For $\gamma \leq \gamma_{\mathrm{A}}$, $e_1^{\mathrm{TR}}(N) = e^{\mathrm{OP}}$ and the agent never switches to the known idea, i.e., $\hat{\theta}(e_1^{\mathrm{TR}}(N)) \leq 0$, where $\gamma_{\mathrm{A}} \in (0, 1)$ solves*

$$\gamma_{\mathrm{A}}\mu - \lambda(1 - \gamma_{\mathrm{A}})^2(1 - \rho) = 0. \tag{6}$$

- *For $\gamma > \gamma_{\mathrm{A}}$, $e_1^{\mathrm{TR}}(N) < e^{\mathrm{OP}}$ and switching occurs with a positive probability, i.e., $\hat{\theta}(e_1^{\mathrm{TR}}(N)) > 0$. In particular, for $\gamma > \gamma_{\mathrm{A}}$,*

$$e_1^{\mathrm{TR}}(N) = \frac{\gamma\mu\lambda(1 - \gamma)(1 + \rho)}{2\gamma\mu - \lambda(1 - \gamma)^2(1 - \rho)^2} \in (0, e^{\mathrm{OP}}), \tag{7}$$

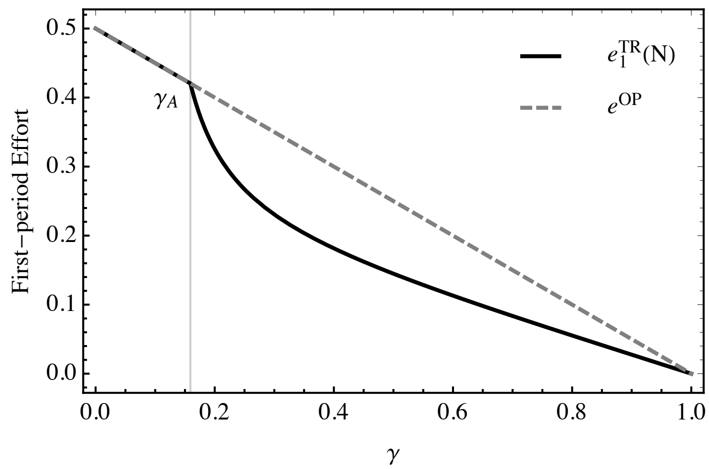
$$\hat{\theta}(e_1^{\mathrm{TR}}(N)) = \frac{2\mu\{\gamma\mu - \lambda(1 - \gamma)^2(1 - \rho)\}}{2\gamma\mu - \lambda(1 - \gamma)^2(1 - \rho)^2} \in (0, \mu), \tag{8}$$

  and $e_1^{\mathrm{TR}}(N)$ is decreasing in γ, while $\hat{\theta}(e_1^{\mathrm{TR}}(N))$ is increasing in γ.

Figure 1 illustrates Lemma 2. For $\gamma > \gamma_{\mathrm{A}}$, $e_1^{\mathrm{TR}}(N)$ lies below e^{OP}.

First-period Idea Choice We now analyze the idea choice in the first period. As seen above, when $\gamma \leq \gamma_{\mathrm{A}}$, regardless of the first-period idea choice $i_1 \in \{K, N\}$, no switching occurs and the effort level in each period is e^{OP}. When $\gamma > \gamma_{\mathrm{A}}$, switching occurs if and only if the agent adopts the new idea in the first period ($i_1 = N$) and the quality of the new idea θ is worse than $\hat{\theta}(e_1^{\mathrm{TR}}(N))$. Thus, the

Figure 1: Comparison of first-period effort



In Figure 1 , we set $\mu = 2$, $\rho = 0.1$, and $\lambda = 0.5$ so that $\gamma_{\mathrm{A}} \approx 0.1591$

agent's expected payoff $\pi(i_1)$ from adopting the first-period idea $i_1 \in \{N, K\}$ is given by

$$\pi(K) = \pi^{\mathrm{OP}} = \lambda[\gamma\mu + (1-\gamma)(2e^{\mathrm{OP}})] - 2c(e^{\mathrm{OP}}); \tag{9}$$

$$\begin{aligned}\pi(N) &= \lambda E[x_2 \mid i_1 = N, e_1^{\mathrm{TR}}(N), \hat{\theta}(e_1^{\mathrm{TR}}(N)), e^{\mathrm{OP}}] - c(e_1^{\mathrm{TR}}(N)) - c(e^{\mathrm{OP}}) \\ &= \lambda F_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N)))[\gamma\mu + (1-\gamma)(\rho e_1^{\mathrm{TR}}(N) + e^{\mathrm{OP}})] \\ &\quad + \lambda \int_{\hat{\theta}(e_1^{\mathrm{TR}}(N))}^{2\mu} [\gamma\theta + (1-\gamma)(e_1^{\mathrm{TR}}(N) + e^{\mathrm{OP}})] dF_\theta - c(e_1^{\mathrm{TR}}(N)) - c(e^{\mathrm{OP}}).\end{aligned} \tag{10}$$

for $\gamma > \gamma_{\mathrm{A}}$, and $\pi(N) = \pi^{\mathrm{OP}}$ for $\gamma \leq \gamma_{\mathrm{A}}$. Comparing these yields the following:

Proposition 1. *In the transparent organization, the agent adopts the new idea in $t = 1$ ($i_1 = N$) if and only if $\gamma > \gamma_{\mathrm{A}}$.*

Transparency promotes the exploration of the new idea. The benefit of transparency is the sorting effect. That is, after exploring the new idea, the agent can improve the expected idea quality by keeping the idea if its quality is better than $\hat{\theta}(e_1^{\mathrm{TR}}(N))$, or replacing it if it is worse. Moreover, if adopting the new idea in the first period, the agent receives at least $\pi(K) = \pi^{\mathrm{OP}}$ because he can obtain π^{OP} by keeping the new idea in the second period ($i_2 = N$) and choosing $e_1 = e_2 = e^{\mathrm{OP}}$. Therefore, the agent explores the new idea if the sorting effect exists—that is, if the switching probability $F_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N)))$ is positive.

If the switching probability is zero, the sorting effect does not exist, and the agent's expected payoff is π^{OP} regardless of the first-period idea choice, and thus, he adopts the known idea in the first period (because the agent adopts the known idea if he is indifferent).

By Lemma 2, the switching probability is positive if and only if $\gamma > \gamma_{\mathrm{A}}$. When $\gamma > \gamma_{\mathrm{A}}$, at the interim stage after learning the quality of the new idea θ, the benefit of improving the idea quality by switching outweighs the wasting effect if the agent receives sufficiently bad news, i.e., $\theta < \hat{\theta}(e_1^{\mathrm{TR}}(N))$. Therefore, switching occurs with a positive probability. As γ declines to $\gamma \leq \gamma_{\mathrm{A}}$, the wasting effect begins to outweigh the benefit of improving the idea quality even when the agent receives the worst news (i.e., $\theta = 0$); switching never occurs, making the sorting effect zero.

In Section 4.2, we consider a case where θ is normally distributed over $(-\infty, \infty)$. In that case, since the support of θ is not bounded from below, there always exists sufficiently bad news that induces him to switch; the switching probability is always positive. Hence, the cutoff corresponding to γ_{A} in the uniform distribution case does not exist in the normal distribution case. Section 4.2 shows that our key results continue to hold when θ is normally distributed, implying that it is not crucial whether the equilibrium is characterized by the cutoff γ_{A}.

3.3 Transparent vs. Opaque Organizations

Since the principal's expected payoff is $(1-\lambda)E[x_2]$, we simply compare the expected final output $E[x_2]$ under the two organizational forms.

In the opaque organization, since $i_1 = i_2 = K$ and $e_1 = e_2 = e^{\mathrm{OP}}$, the expected final output is given by $\Pi^{\mathrm{OP}} = \gamma\mu + (1-\gamma)(2e^{\mathrm{OP}})$.

Next, we consider the transparent organization. By Proposition 1, if $\gamma \leq \gamma_{\mathrm{A}}$, the agent works with the known idea in both the periods ($i_1 = i_2 = K$) and exerts

e^{OP} in each period. Hence, the expected final output is the same as that of the opaque organization: $\Pi^{\mathrm{TR}} = \Pi^{\mathrm{OP}}$ or $\Delta\Pi = 0$, where we denote the expected gain from transparency by $\Delta\Pi \equiv \Pi^{\mathrm{TR}} - \Pi^{\mathrm{OP}}$.

If $\gamma > \gamma_{\mathrm{A}}$, the agent chooses $i_1 = N$ and $(e_1, e_2) = (e_1^{\mathrm{TR}}(N), e^{\mathrm{OP}})$ while following the switching strategy $\hat{\theta}(e_1^{\mathrm{TR}}(N))$. Thus, $\Delta\Pi$ can be written as

$$\Delta\Pi = \underbrace{\gamma \int_{\hat{\theta}(e_1^{\mathrm{TR}}(N))}^{2\mu} (\theta_1 - \mu)\, dF_\theta}_{\text{Sorting effect}} - \underbrace{(1-\gamma)\left\{e^{\mathrm{OP}} - [1 - (1-\rho)F_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N)))]e_1^{\mathrm{TR}}(N)\right\}}_{\text{Wasting and demotivating effects}}. \tag{11}$$

The first term is positive and captures the sorting effect, thereby increasing the expected quality of the idea. The second term is negative and captures the wasting and demotivating effects. Transparency not only wastes the first-period effort due to switching of ideas, which is captured by $(1-\rho)F_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N)))e_1^{\mathrm{TR}}(N)$ in (11), but also reduces first-period effort $e_1^{\mathrm{TR}}(N) < e^{\mathrm{OP}}$ through the demotivating effect.

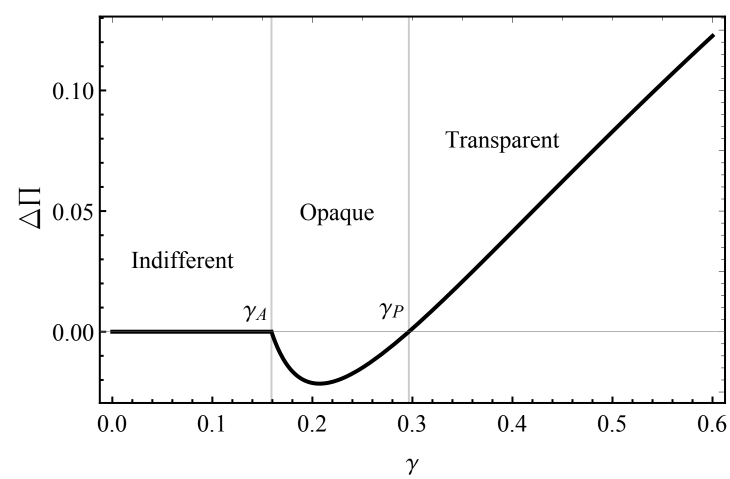
For $\gamma > \gamma_{\mathrm{A}}$, since $\Delta\Pi$ is the weighted sum of the positive and negative terms, and $(1-\gamma)$ is the weight for the negative term, for sufficiently small γ, the wasting and demotivating effects may dominate the sorting effect so that $\Delta\Pi < 0$. The following proposition verifies this:

Proposition 2. *Transparency may be counter-productive. In particular, the expected gain from transparency* $\Delta\Pi \equiv \Pi^{\mathrm{TR}} - \Pi^{\mathrm{OP}}$ *satisfies the following:* $\Delta\Pi = 0$ *if* $\gamma \le \gamma_{\mathrm{A}}$*;* $\Delta\Pi < 0$ *if* $\gamma \in (\gamma_{\mathrm{A}}, \gamma_{\mathrm{P}})$*; and* $\Delta\Pi > 0$ *if* $\gamma > \gamma_{\mathrm{P}}$*, where* $\gamma_{\mathrm{P}} > \gamma_{\mathrm{A}}$.

Figure 2 illustrates Proposition 2, showing how the expected gain from transparency as γ varies. As explained above in Proposition 1, if $\gamma \le \gamma_{\mathrm{A}}$, effort is sufficiently important such that no switching occurs, thus, organizational forms do not matter to the principal. For $\gamma > \gamma_{\mathrm{A}}$, the logic behind the result is straightforward. Transparency wastes the first-period effort, and this possibility weakens effort incentives, which is too costly when effort is sufficiently important (i.e., $\gamma < \gamma_{\mathrm{P}}$). Thus, if $\gamma \in (\gamma_{\mathrm{A}}, \gamma_{\mathrm{P}})$, transparency is counter-productive; that is, the principal strictly prefers the opaque organization.

Note also that the result reveals that the agent is too inclined to explore the new idea from the perspective of the principal. Indeed, for $\gamma \in (\gamma_{\text{A}}, \gamma_{\text{P}})$, the agent adopts the new idea although the principal wants him to adopt the known one. This is because the agent privately bears the cost of implementing the ideas $(c(e_1), c(e_2))$ and the first-period effort is lower for the new idea; hence, $c(e_1^{\text{TR}}(N)) < c(e_1^{\text{TR}}(K))$, as seen in Lemma 2. However, the principal does not consider this cost savings associated with the new idea. Thus, the agent is too inclined to explore the new idea from the principal's perspective $\gamma_{\text{A}} < \gamma_{\text{P}}$. We elaborate this further in Section 3.4.

Figure 2: Expected gain from transparency: $\Delta\Pi$



In Figure 2, we use the same parameter values as in Figure 1 so that $\gamma_{\text{A}} \approx 0.1591$ and $\gamma_{\text{P}} \approx 0.2964$.

The Role of Knowledge Generality The preceding analysis indicates that in the transparent organization, when the agent changes ideas, it wastes past effort and undermines the effort incentive in the first period. These drawbacks of transparency can outweigh its benefit. We now address how the generality ρ of knowledge or know-how affects the effort incentive and the final output.

When the acquired knowledge becomes more general (ρ increases), the wastage of past efforts is reduced. Thus, one might conjecture that in the transparent organization, greater ρ should always increase the effort incentive and hence the final output. However, we establish that $e_1^{\mathrm{TR}}(N)$, Π^{TR}, and $\Delta\Pi$ may decrease in ρ:

Proposition 3. *There exist γ_1 and γ_2 such that (i) $\gamma_{\mathrm{A}} < \gamma_1 < \gamma_2 < \gamma_{\mathrm{P}}$, (ii) $de_1^{\mathrm{TR}}(N)/d\rho$ is negative for $\gamma \in (\gamma_{\mathrm{A}}, \gamma_2)$ and positive for $\gamma > \gamma_2$, and (iii) $d\Delta\Pi/d\rho = d\Pi^{\mathrm{TR}}/d\rho$ is negative for $\gamma \in (\gamma_{\mathrm{A}}, \gamma_1)$ and positive for $\gamma > \gamma_1$. Moreover, $\hat{\theta}(e_1^{\mathrm{TR}}(N))$ is increasing in ρ, while both γ_{A} and γ_{P} are decreasing in ρ.*

Figure 3: The effect of knowledge generality ρ on $e_1^{\mathrm{TR}}(N)$ and $\Delta\Pi$.

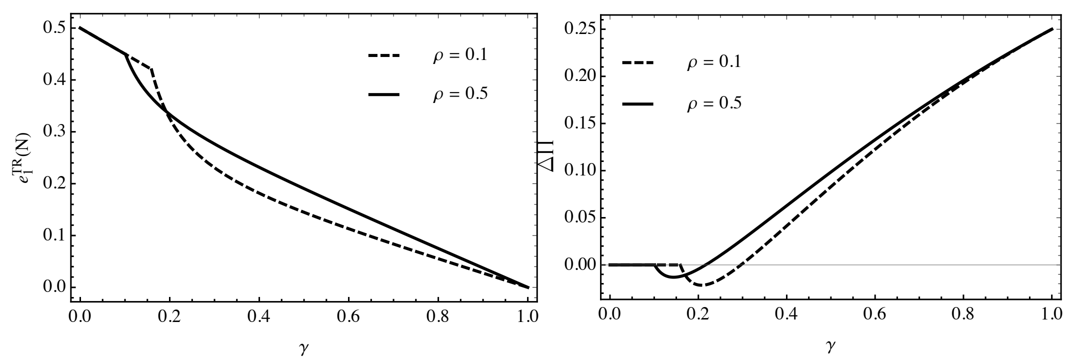


Figure 3 illustrates Proposition 3. In both the figures, the dashed curve is for $\rho = 0.1$ and the solid curve is for $\rho = 0.5$. The left-hand figure shows that $e_1^{\mathrm{TR}}(N)$ is smaller when ρ is greater for $\gamma \in (\gamma_{\mathrm{A}}, \gamma_2)$. The right-hand figure shows that $\Delta\Pi$ is smaller when ρ is greater for $\gamma \in (\gamma_{\mathrm{A}}, \gamma_1)$. Moreover, γ_{P}, γ at which $\Delta\Pi = 0$, is smaller for the solid curve. In both the figures, $\mu = 2$ and $\lambda = 0.5$, as in Figure 1. For $\rho = 0.1$, $\gamma_{\mathrm{A}} \approx 0.1591$, while $\gamma_{\mathrm{A}} \approx 0.1010$ for $\rho = 0.5$.

Let us first understand the effect of ρ on $e_1^{\mathrm{TR}}(N)$. Differentiating the marginal benefit from exerting $e_1^{\mathrm{TR}}(N)$ (i.e., the right-hand side of (5)) with respect to ρ yields

$$\lambda(1-\gamma)\Big\{F_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N))) - (1-\rho) f_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N))) \underbrace{\frac{d\hat{\theta}(e_1^{\mathrm{TR}}(N))}{d\rho}}_{>0}\Big\}. \tag{12}$$

The knowledge generality ρ has opposing impacts on the demotivating effect. When ρ is greater, the agent can utilize the first-period effort to a greater degree after switching ideas, which reduces the demotivating effect, captured by $\lambda(1-\gamma)F_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N)))$ in (12). However, a greater ρ also increases the switching probability $F_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N)))$, thereby, increasing the demotivating effect, captured by $\lambda(1-\gamma)(1-\rho)f_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N)))\frac{d\hat{\theta}(e_1^{\mathrm{TR}}(N))}{d\rho} > 0$ in (12) because $\hat{\theta}(e_1^{\mathrm{TR}}(N))$ is increasing in ρ. Collectively, when γ is sufficiently small or $\gamma \in (\gamma_{\mathrm{A}}, \gamma_2)$, i.e., when the effort is sufficiently important, the switching probability $F_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N)))$ is so small that the former effect is dominated.[12]

To understand the reason why a greater ρ can reduce the principal's payoff Π^{TR}, let us suppose $\gamma > \gamma_{\mathrm{A}}$. Differentiating Π^{TR} with respect to ρ yields

$$\begin{aligned}\frac{d\Pi^{\mathrm{TR}}}{d\rho} &= \frac{\partial \Pi^{\mathrm{TR}}}{\partial \rho} + \frac{\partial \Pi^{\mathrm{TR}}}{\partial e_1}\frac{de_1^{\mathrm{TR}}(N)}{d\rho} + \frac{\partial \Pi^{\mathrm{TR}}}{\partial \hat{\theta}}\frac{d\hat{\theta}(e_1^{\mathrm{TR}}(N))}{\partial \rho} \\ &= (1-\gamma)F_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N)))e_1^{\mathrm{TR}}(N) + (1-\gamma)[1-(1-\rho)F_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N)))]\frac{de_1^{\mathrm{TR}}(N)}{d\rho},\end{aligned} \tag{13}$$

where $\partial \Pi^{\mathrm{TR}}/\partial \hat{\theta} = 0$ holds by the envelop theorem, and the last equality holds by (10). The first term in (13) captures the positive impact of the reduced wasting effect. The expression $(1-\gamma)[1-(1-\rho)F_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N)))] > 0$ captures the change in the final output caused by a marginal change in e_1, $\partial \Pi^{\mathrm{TR}}/\partial e_1$. Thus, (13) can be negative only when $de_1^{\mathrm{TR}}(N)/d\rho < 0$, i.e., when $\gamma < \gamma_2$. However, γ must be even smaller ($\gamma < \gamma_1$), so that the switching probability $F_\theta(\hat{\theta}(e_1^{\mathrm{TR}}(N))$ is sufficiently small, and the second term dominates in (13). Thus, $d\Pi^{\mathrm{TR}}/d\rho < 0$ if $\gamma \in (\gamma_{\mathrm{A}}, \gamma_1)$. Simply put, when the output depends heavily on the effort, the switching rarely occurs, and the benefit of the reduced wasting effect due to a greater ρ is small, while the reduction in $e_1^{\mathrm{TR}}(N)$ due to a greater ρ is costly for production, thus, Π^{TR} is reduced.

This result highlights the importance of the organizational design. When the

[12]Differentiating the marginal cost $e_1^{\mathrm{TR}}(N)$ with respect to ρ yields $\partial e_1^{\mathrm{TR}}(N)/\partial \rho$. Equating this with (12) will reduce to (A5).

output depends more on the idea implementation rather than the idea quality, the more general knowledge and know-how the agent acquires in the process, the more counter-productive it is to make the organization transparent. Therefore, when managers design organizational transparency, they need to consider not only the importance of the effort of idea implementation in relation with idea quality during the innovation process, but also the generality of the knowledge and know-how gained during idea implementation.

3.4 Inefficiency of Idea Choice

We have shown that the agent is too inclined to adopt a new idea from the principal's perspective, i.e., $\gamma_{\mathrm{A}} < \gamma_{\mathrm{P}}$. In this subsection, we consider two different benchmarks to provide social perspectives, as in HT. One is the first-best benchmark in which an active social planner controls all the choices to maximize the total welfare $E[x_2] - c(e_1) - c(e_2)$. The other is the second-best benchmark in which a passive social planner makes idea choices to maximize the total welfare but takes the equilibrium effort choices $(e_1^{\mathrm{TR}}(N), e^{\mathrm{OP}})$ as given.

We first consider the first-best benchmark. The first-best outcome corresponds to the equilibrium choices as $\lambda \to 1$ because the agent's objective function approaches the total welfare as $\lambda \to 1$. The social planner chooses greater effort in each period ($e_1^{\mathrm{TR}}(N)$ and e^{OP} increase with λ). The first-best idea choice in period 2 is identical to the equilibrium choice because $\hat{\theta}(e_1)$ is independent from λ.[13] However, the first-best switching probability is smaller because of the greater first-period effort.

Recall that the agent explores the new idea if and only if $\gamma > \gamma_{\mathrm{A}}$, and at the cutoff γ_{A}, the wasting effect is just covered by the benefit of improving the idea quality from the worst quality (i.e., $\theta = 0$) to the mean quality μ. Similarly, the social planner chooses the new idea in the first period ($i_1 = N$) if and only if

[13]It is identical in terms of strategies.

$\gamma > \gamma_{\text{FB}} = \lim_{\lambda \to 1} \gamma_{\text{A}}$, where γ_{FB} solves

$$\gamma_{\text{FB}}\mu - (1 - \gamma_{\text{FB}})^2(1 - \rho) = 0. \tag{14}$$

Since the effort is smaller, reducing the wasting effect is smaller in relation to the first best, the net benefit of adopting a new idea is greater for the agent than for the social planner. Thus, $\gamma_{\text{A}} < \gamma_{\text{FB}}$ must follow; that is, the agent is too inclined to adopt the new idea in relation to the first best. One can also verify $\gamma_{\text{A}} < \gamma_{\text{FB}}$ by observing that the left-hand side of (6) is increasing in γ and decreasing in λ.

Furthermore, if the principal was able to choose the idea, she would be less likely to adopt the new idea in relation to the first best for sufficiently low λ:

Proposition 4. *There exist* $\lambda^* \in (0, 1/2)$ *such that* $\gamma_{\text{P}} < \gamma_{\text{FB}}$ *if and only if* $\lambda \in (0, \lambda^*)$.

First, γ_{FB} is independent from λ by definition. Moreover, γ_{P} increases with λ, or as λ rises, the principal becomes less willing to let the agent adopt the new idea. Intuitively, as λ rises, the agent exerts more effort and consequently more effort is wasted when adopting the new idea, i.e., the effort differential $e_1^{\text{TR}}(K) - e_1^{\text{TR}}(N)$ increases. This makes the new idea less attractive to the principal.

Next, we consider the second-best benchmark. The social planner chooses the idea in each period to maximize the total welfare given the equilibrium effort choices. Recall that $\hat{\theta}(e_1)$, the agent's idea choice in the second period, is efficient, given the first-period effort e_1. Thus, the second-best idea choice in the second period is the same as the equilibrium choice. Next, consider the idea choice in the first period. The difference in the total welfare between $i_1 = N$ and $i_1 = K$ is given by

$$\Delta W \equiv \pi(N) - \pi(K) + (1 - \lambda)\{E[x_2 \mid i_1 = N] - E[x_2 \mid i_1 = K]\}.$$

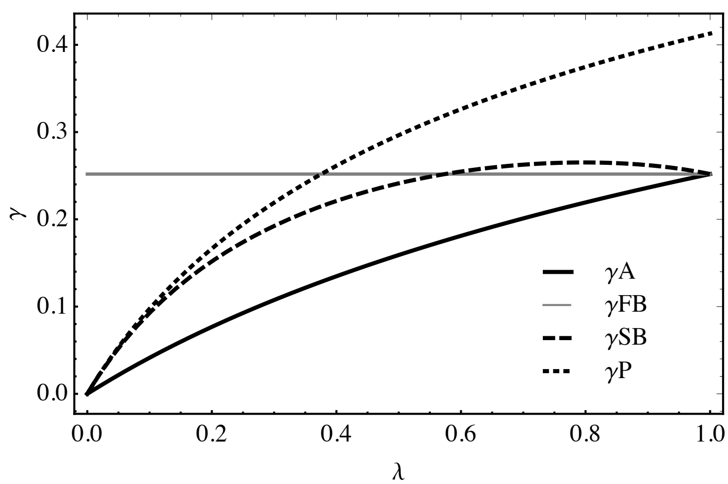
For $\gamma \leq \gamma_{\text{A}}$, $\Delta W = 0$ because no switching should occur according to $\hat{\theta}(e_1)$. For $\gamma > \gamma_{\text{A}}$, $\pi(N) > \pi(K)$ while $E[x_2 \mid i_1 = N] < E[x_2 \mid i_1 = K]$ only if $\gamma < \gamma_{\text{P}}$. The following proposition establishes that the threshold value of γ for $\Delta W > 0$ and $i_1 = N$ lies between $(\gamma_{\text{A}}, \gamma_{\text{P}})$:

Proposition 5. *There exists* $\gamma_{SB} \in (\gamma_A, \gamma_P)$ *such that adopting the new idea in period* 1 *(*$i_1 = N$*) is the second-best idea choice if and only if* $\gamma > \gamma_{SB}$.

Thus, the agent is too inclined to explore the new idea, in relation to not only the first best ($\gamma_A < \gamma_{FB}$), but also to the second best ($\gamma_A < \gamma_{SB}$). This is because the agent only obtains a fraction λ of the final output and thus places too much emphasis on the cost savings of choosing a new idea.

Figure 4 illustrates how all the thresholds of γ vary as λ changes. Let us first compare γ_{FB} with the others. While γ_{FB} is independent from λ, $\gamma_A < \gamma_{FB}$ with $\gamma_A \to \gamma_{FB}$ as $\lambda \to 1$. Additionally, $\gamma_P > \gamma_{FB}$ if and only if λ is large enough as per Proposition 4. Moreover, by Proposition 5, $\gamma_{SB} \in (\gamma_A, \gamma_P)$.[14]

Figure 4: Efficiency of the Agent's Idea Choice



In Figure 4, we set $\mu = 2$ and $\rho = 0.1$ as in Figure 1.

It is worth noting the difference between our model and HT. In our model, the agent is too inclined toward the new idea both from the perspective of the principal and the social planner, whereas in HT, the desirability of the idea choice depends on the relative bargaining power between the agent and the principal. This

[14] At $\lambda = 0$, $\gamma_A = \gamma_{SB} = \gamma_P = 0 < \gamma_{FB}$ because γ_{FB} does not depend on λ and $\hat{\theta} = \mu$ at $\lambda = 0$. One can also verify that $\gamma_{SB} = \gamma_A = \gamma_{FB} < \gamma_P$ at $\lambda = 1$.

difference arises from the difference in the settings. In HT's model, the innovative task requires ex-post bargaining while the non-innovative task does not, whereas in our model, both the known and new ideas require ex-post bargaining.

4 Extensions

In this section, we demonstrate that our comparative static result regarding ρ continues to hold (1) when the new idea quality is uniformly distributed as before, but the measurement error ε is uniformly distributed over $(-\delta, \delta)$ in Section 4.1 and (2) when both the new idea quality θ and the measurement error ε are normally (and independently) distributed in Section 4.2. Furthermore, we establish a key result—the availability of a *more precise* interim performance signal may *reduce* final output. We delegate all the proofs to Appendix B.

4.1 Imperfect Uniform Signal

Suppose that ε is uniformly distributed on $(-\delta, \delta)$. The noise is relevant only when the new idea is adopted and thus we consider the transparent organization. We first derive the conditional expectation of θ given the signal $s = \gamma\theta + (1-\gamma)\, e_1 + \varepsilon$ and $i_1 = N$.

Preliminaries: Conditional Expectation Denote the lowest and highest possible signals by

$$\underline{s}(e_1) \equiv (1-\gamma)e_1 - \delta \text{ and } \overline{s}(e_1) \equiv \gamma(2\mu) + (1-\gamma)e_1 + \delta$$

To simplify our analysis, we assume that the signal noise is not too large:

Assumption A1. $0 < \delta < \gamma\mu$.

Lemma 3. *Suppose **A1**. After $i_1 = N$ and e_1 are chosen in $t = 1$, the conditional expectation of θ given $s \in [\underline{s}(e_1), \overline{s}(e_1)]$ is given by*

$$E[\theta \mid s] = \begin{cases} \frac{s-(1-\gamma)e_1+2\gamma\mu-\delta}{2\gamma} & \text{if } s \in [\overline{s}(e_1) - 2\delta, \overline{s}(e_1)] \\ \frac{s-(1-\gamma)e_1}{\gamma} & \text{if } s \in [\underline{s}(e_1) + 2\delta, \overline{s}(e_1) - 2\delta] \\ \frac{s-(1-\gamma)e_1+\delta}{2\gamma} & \text{if } s \in [\underline{s}(e_1), \underline{s}(e_1) + 2\delta]. \end{cases}$$

Moreover, $E[\theta \mid s]$ is continuous and strictly increasing in s.

Second-period Choices As in the perfect signal case, the marginal benefit of effort is independent from the idea quality. Thus, the second-period effort is $e_2 = e^{\mathrm{OP}} = \lambda(1-\gamma)$ and it is optimal to adopt the known idea ($i_2 = K$) after $i_1 = K$.

Now consider the second-period idea choice after $i_1 = N$ and e_1 are chosen. As in the perfect signal case (Lemmas 1 and 2), if e_1 is sufficiently large, the wasting effect becomes too large to switch to $i_2 = K$, regardless of the signal s. However, if e_1 is sufficiently small, the agent switches to K if and only if the signal s is low enough such that sufficient improvement in the idea quality is expected:

Lemma 4. *Suppose **A1**. If $i_1 = K$, then it is optimal to keep working with $i_2 = K$. If $i_1 = N$ and $e_1 \geq \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, the agent always keeps working with $i_2 = N$. If $i_1 = N$ and $e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, then the agent switches to $i_2 = K$ if and only if he receives the signal $s \leq \hat{s}(e_1)$, where*

$$\hat{s}(e_1) = \begin{cases} 2\gamma\mu - (1-\gamma)(1-2\rho)e_1 - \delta & \text{if } \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)} \leq e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)} \\ \gamma\mu + (1-\gamma)\rho e_1 & \text{if } e_1 < \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}. \end{cases}$$

The cutoff signal $\hat{s}(e_1)$ takes different functional forms depending on the level of e_1 because $E[\theta \mid s]$ takes different functional forms, depending on the value of the signal s, as seen in Lemma 3. When $e_1 < \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}$, the cutoff signal $\hat{s}(e_1)$ lies in $(\underline{s}(e_1) + 2\delta, \overline{s}(e_1) - 2\delta)$. As the first-period effort e_1 rises, the cutoff signal $\hat{s}(e_1)$ and the switching probability continuously decrease. Eventually, e_1 reaches the smaller threshold $\frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}$, at which point $\hat{s}(e_1)$ falls onto $[\underline{s}(e_1), \underline{s}(e_1) + 2\delta]$. Once e_1 reaches the greater threshold $\frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, the wasting effect becomes so large that the switching probability becomes zero.

First-period Choices Upon adopting $i_1 = K$ in the first period, as in the perfect signal case, the optimal first-period effort is $e_1^{\mathrm{TR}}(K) = e^{\mathrm{OP}}$.

Now consider the first-period effort choice after adopting the new idea ($i_1 = N$). By Lemma 4, the agent switches to $i_2 = K$ if and only if $s \leq \hat{s}(e_1)$ and $e_1 < \frac{\mu\gamma}{(1-\gamma)(1-\rho)}$. We now establish that as in the perfect signal case, the possibility of switching ideas reduces the marginal benefit of the first-period effort, resulting in $e_1^{\mathrm{TR}}(N) < e^{\mathrm{OP}}$ (if γ is large enough to make the switching probability positive):

Lemma 5. *Suppose **A1** and the new idea is adopted in $t = 1$ ($i_1 = N$). If $\gamma \leq \gamma_{\mathrm{A}}$, then the agent never switches to the known idea and $e_1^{\mathrm{TR}}(N) = e^{\mathrm{OP}}$, and if $\gamma > \gamma_{\mathrm{A}}$, switching occurs with a positive probability and $e_1^{\mathrm{TR}}(N) < e^{\mathrm{OP}}$.*

More specifically, (i) if $\gamma \in (\gamma_{\mathrm{A}}, \gamma^*)$, the optimal first-period effort is $e_1^{\mathrm{TR}}(N) \in \left[\frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}, \frac{\gamma\mu}{(1-\gamma)(1-\rho)}\right)$, where $e_1^{\mathrm{TR}}(N)$ equates (OA11) in Appendix B to zero; (ii) if $\gamma \in [\gamma^*, 1]$, $e_1^{\mathrm{TR}}(N) < \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}$, where $e_1^{\mathrm{TR}}(N)$ equates (OA9) in Appendix B to zero; (iii) γ^* satisfies $\gamma^*\mu = \left\{\frac{2\gamma^*\mu-(1-\rho)\delta}{2\gamma^*\mu-2\delta}\right\} \lambda(1-\gamma^*)^2 (1-\rho)$ and is increasing in δ, and $\gamma^* = \gamma_{\mathrm{A}}$ at $\delta = 0$.

As in the perfect signal case, when $\gamma \leq \gamma_{\mathrm{A}}$, effort is sufficiently important so that no switching occurs because the wasting effect is too costly even after the agent learns that the quality of the new idea is zero.

We now analyze the first-period idea choice. As in Proposition 1, the agent explores the new idea as long as the sorting effect exists—that is, when the switching probability is positive, i.e., $\gamma > \gamma_{\mathrm{A}}$ by Lemma 5. This is because, as before, if adopting the new idea in the first period, the agent receives at least $\pi(K) = \pi^{\mathrm{OP}}$ by keeping the new idea in the second period ($i_2 = N$) and choosing $e_1 = e_2 = e^{\mathrm{OP}}$; moreover, the agent can improve the expected idea quality by keeping the idea if and only if he receives a sufficiently good signal (even if the signal is imperfect).

Meanwhile, if $\gamma \le \gamma_{\mathrm{A}}$, the switching never occurs and, therefore, the sorting effect is absent, which leaves the agent indifferent between N and K; thus, the agent adopts K in the first period. The following lemma summarizes these:

Lemma 6. *Suppose **A1**. In $t = 1$, in the transparent organization, the agent chooses the known idea ($i_1 = K$) if $\gamma \le \gamma_{\mathrm{A}}$ and the new idea ($i_1 = N$) if $\gamma > \gamma_{\mathrm{A}}$.*

The Roles of Knowledge Generality and Signal Precision We establish our key results regarding the effects of knowledge generality ρ and signal precision δ on the effort incentive and the final output. Let $F_s(s)$ denote the distribution function of the signal s. To ease our exposition, we restrict our attention to $\gamma \in (\gamma_{\mathrm{A}}, \gamma^*)$.[15]

In the opaque organization, the agent chooses the known idea in every period and the expected final output is the same as before: $\Pi^{\mathrm{OP}} = \gamma\mu + (1-\gamma)2e^{\mathrm{OP}}$. In the transparent organization, for $\gamma > \gamma_{\mathrm{A}}$, the switching occurs if and only if $s \le \hat{s}(e_1^{\mathrm{TR}}(N))$. To simplify the exposition, we simply write e_1^{TR}, $\hat{s}$, and $\overline{s}$ instead of $e_1^{\mathrm{TR}}(N)$, $\hat{s}(e_1^{\mathrm{TR}})$, and $\overline{s}(e_1^{\mathrm{TR}})$. The expected final output in the transparent organization is written as

$$\Pi^{\mathrm{TR}} = \gamma\Big(\mu + \int_{\hat{s}}^{\overline{s}} (E[\theta \mid s] - \mu) dF_s\Big) + (1-\gamma)\left\{[1-(1-\rho)F_s(\hat{s})]e_1^{\mathrm{TR}} + e^{\mathrm{OP}}\right\}. \quad (15)$$

For $\gamma > \gamma_{\mathrm{A}}$, the difference in the expected final output is given by

$$\Delta\Pi = \gamma \int_{\hat{s}}^{\overline{s}} (E[\theta \mid s] - \mu) dF_s - (1-\gamma)\left\{e^{\mathrm{OP}} - [1-(1-\rho)F_s(\hat{s})]e_1^{\mathrm{TR}})\right\},$$

where the first term captures the sorting effect, and the second term captures the wasting and demotivating effects, as in the perfect signal case (11).

[15] By Lemmas 4 and 5, for $\gamma \ge \gamma^*$, $e_1^{\mathrm{TR}}(N)$ is so small that $\hat{s}(e_1^{\mathrm{TR}}(N))$ falls in $[\underline{s}(e_1)+2\delta, \overline{s}(e_1)-2\delta]$, where $E[\theta \mid \hat{s}(e_1^{\mathrm{TR}}(N))] = \frac{\hat{s}(e_1^{\mathrm{TR}}(N))-(1-\gamma)e_1}{\gamma}$ is independent from signal precision δ by Lemma 3 (which keeps $e_1^{\mathrm{TR}}(N)$ and the switching probability the same as in the perfect signal case). This independence of $E[\theta \mid s]$ from δ is a property that is specific to the uniform-uniform assumption. Thus, we focus on the case of $\gamma < \gamma^*$.

We first establish that, as in the perfect signal case (Proposition 3), when the acquired knowledge becomes more general, the expected final output in the transparent organization is reduced for sufficiently small γ.

Proposition 6. *Suppose **A1** and $\gamma \in (\gamma_A, \gamma^*)$. There exist $\gamma_\rho \in (\gamma_A, \gamma^*)$ and $\delta_\rho > 0$ (specified in Appendix B) such that (i) $d\Delta\Pi/d\rho = d\Pi^{TR}/d\rho$ is negative if $\delta \leq \delta_\rho$ and (ii) $d\Delta\Pi/d\rho = d\Pi^{TR}/d\rho$ is negative for $\gamma \in (\gamma_A, \gamma_\rho)$ and positive for $\gamma \in [\gamma_\rho, \gamma^*)$ if $\delta > \delta_\rho$.*

The logic is the same as in the perfect signal case. With greater ρ, on the one hand, the first-period effort can be utilized to a greater degree after switching the ideas; on the other hand, greater ρ also increases the switching probability, which increases the demotivating effect and thus reduces e_1^{TR} for sufficiently small γ. When the effort is sufficiently important (i.e., γ is small), the latter effect dominates, yielding $d\Pi^{TR}/d\rho < 0$.

Note that by Lemma 5, γ^* is increasing in δ. Thus, when $\delta \geq \delta_\rho$, γ^* is sufficiently large so that $d\Pi^{TR}/d\rho > 0$ at γ close enough to γ^*, i.e., $\gamma \in (\gamma_\rho, \gamma^*)$.

Next, we consider how the accuracy of the interim performance signal affects the final output. One might expect that the final output would always decrease when the signal becomes less accurate, and indeed, the agent will be less able to assess the quality of the new idea and benefit less from idea sorting, thereby reducing the benefit of transparency. Perhaps surprisingly, however, we establish that the final output may *increase* as the signal becomes *less accurate*:

Proposition 7. *Suppose **A1** and $\gamma \in (\gamma_A, \gamma^*)$. There exist $\gamma_\delta \in (\gamma_A, \gamma^*)$ and $\delta_\delta > 0$ (specified in Appendix B) such that (i) $d\Delta\Pi/d\delta = d\Pi^{TR}/d\delta$ is positive if $\delta \leq \delta_\delta$ and (ii) $d\Delta\Pi/d\delta = d\Pi^{TR}/d\delta$ is positive for $\gamma \in (\gamma_A, \gamma_\delta)$ and negative for $\gamma \in [\gamma_\delta, \gamma^*)$ if $\delta > \delta_\delta$.*

To see why less accurate signal, i.e., greater δ may increase the final output, differentiating Π^{TR} with respect to δ and applying the envelop theorem yield:

$$\frac{d\Pi^{TR}}{d\delta} = \frac{\partial\Pi^{TR}}{\partial\delta} + \frac{\partial\Pi^{TR}}{\partial e_1}\frac{de_1^{TR}(N)}{d\delta}. \tag{16}$$

The first term $\frac{\partial \Pi^{\mathrm{TR}}}{\partial \delta}$ is negative because the lower the accuracy of the signal, the smaller the sorting effect (i.e., the less improvement in the idea quality).

However, less accurate interim performance signals can also have a positive effect. If the signal is less accurate, and the sorting effect is smaller, the probability of the switch is smaller, which reduces the demotivating effect and thus increases the first-period effort, i.e., $\frac{de_1^{\mathrm{TR}}(N)}{d\delta} > 0$. This increase in the first-period effort will increase the final output $\frac{\partial \Pi^{\mathrm{TR}}}{\partial e_1} > 0$.

If the output relies more heavily on effort rather than the idea quality (i.e., if γ is small), the benefit from greater effort $\frac{\partial \Pi^{\mathrm{TR}}}{\partial e_1} > 0$ is large while the reduced sorting effect $\frac{\partial \Pi^{\mathrm{TR}}}{\partial \delta} < 0$ is small; this is because the switching probability is small when effort is important. Thus, the positive effect dominates the negative effect in (16).

4.2 Normal Normal

This section demonstrates that the key insights we obtained under the uniform assumption extend to the case of normal distributions. Specifically, we assume that the quality of the new idea θ independently follows $N(\mu, \sigma_\theta^2)$ and the measurement error ε independently follows $N(0, \sigma_\varepsilon^2)$.

In the opaque organization, all analyses remain the same as in the uniform-uniform case: $e_1 = e_2 = e^{\mathrm{OP}}$ and $i_1 = i_2 = K$. Before analyzing the transparent organization, we first derive the conditional expectation of θ given the signal s.

Preliminaries: Conditional Expectation Observe first that receiving the signal $s = \gamma\theta + (1-\gamma)e_1 + \varepsilon$ is equivalent to receiving the signal $\omega \equiv s - (1-\gamma)e_1 = \gamma\theta + \varepsilon$ because the agent knows his own effort choice e_1. Since ω and θ are jointly normal, by a well-known formula, the expectation of θ conditional on ω is given by

$$E[\theta \mid \omega] = \mu + \frac{\gamma\sigma_\theta^2}{\gamma^2\sigma_\theta^2 + \sigma_\varepsilon^2}(\omega - \gamma\mu) = \frac{\gamma\sigma_\theta^2\omega + \sigma_\varepsilon^2\mu}{\gamma^2\sigma_\theta^2 + \sigma_\varepsilon^2}, \tag{17}$$

where the second equality holds because $\omega \sim N(\gamma\mu, \gamma^2\sigma_\theta^2 + \sigma_\varepsilon^2)$ by the reproductive property of normal distributions and $Cov(\theta, \omega) = \gamma\sigma_\theta^2$. Note that $E[\theta \mid \omega = \gamma\mu] = \mu$ and $E[\theta \mid \omega]$ is strictly increasing in ω.

Second-period Choices The optimal second-period effort and idea choices after $i_1 = K$ are the same as those in the uniform-uniform case.

Now, consider after $i_1 = N$ and e_1 are chosen. As in the uniform-uniform case, the agent switches to K if and only if the signal ω is low enough such that a sufficient improvement in the idea quality can be expected:

$$E[\theta \mid \omega] \leq \mu - \frac{(1-\gamma)}{\gamma}(1-\rho)e_1. \tag{18}$$

Define $\hat{\omega}(e_1)$ at which (18) holds with equality. Solving this for $\hat{\omega}(e_1)$ yields:

Lemma 7. *Given that the new idea is chosen in $t = 1$ ($i_1 = N$) and any $e_1 > 0$, the agent switches to the known idea in $t = 2$ ($i_2 = K$) if and only if he receives the signal $\omega \leq \hat{\omega}(e_1)$, where*

$$\hat{\omega}(e_1) = \gamma\mu - \frac{\gamma^2\sigma_\theta^2 + \sigma_\varepsilon^2}{\gamma\sigma_\theta^2}\frac{(1-\gamma)}{\gamma}(1-\rho)e_1. \tag{19}$$

Observe that unlike in the uniform-uniform case, for any γ and e_1, the switch occurs with a positive probability because, when θ takes any real number, the agent receives sufficiently bad signals with a positive probability, so that the benefit of switching outweighs the wasting effect. As in the uniform-uniform case, the cutoff $\hat{\omega}(e_1)$ is decreasing in e_1 and increasing in ρ.

First-period Choices After choosing the known idea in period 1 ($i_1 = K$), as in the uniform-uniform case, the optimal first-period effort is $e_1^{\mathrm{TR}}(K) = e^{\mathrm{OP}}$.

After adopting the new idea ($i_1 = N$), let F_ω be the distribution function of ω.

Thus, the expected agent payoff from choosing e_1 is

$$\pi(N) = \lambda F_\omega(\hat{\omega}(e_1))[\gamma\mu + (1-\gamma)(\rho e_1 + e^{\mathrm{OP}})] \\ + \lambda \int_{\hat{\omega}(e_1)}^{\infty} \{\gamma E[\theta \mid \omega] + (1-\gamma)(e_1 + e^{\mathrm{OP}})\} dF_\omega - c(e_1) - c(e^{\mathrm{OP}}).$$

The first-order condition for the optimal first-period effort is[16]

$$e_1^{\mathrm{TR}}(N) = \lambda(1-\gamma) - \lambda(1-\gamma)(1-\rho)F_\omega(\hat{\omega}(e_1^{\mathrm{TR}}(N)), \tag{20}$$

and thus, $e_1^{\mathrm{TR}}(N) < e^{\mathrm{OP}}$.

We now establish that the agent always adopts the new idea in the first period.

Lemma 8. *Suppose $\theta \sim N(\mu, \sigma_\theta^2)$ and $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. Then, in the transparent organization, the agent chooses the new idea in $t = 1$ ($i_1 = N$).*

In contrast to the uniform-uniform case in which the agent adopts a new idea in the first period only if $\gamma > \gamma_{\mathrm{A}}$, the agent now adopts a new idea in the first period ($i_1 = N$) for all γ, since the switch occurs with a positive probability for all γ as per Lemma 7.

The Roles of Knowledge Generality and Signal Precision We focus on demonstrating that our main results regarding the roles of knowledge generality and signal precision extend to the normal-normal case, instead of fully comparing the opaque and transparent organizations because it is too complex to solve explicitly.

In the opaque organization, the agent chooses $i_1 = i_2 = K$ and $e_1 = e_2 = e^{\mathrm{OP}}$, yielding the same expected final output as before: $\Pi^{\mathrm{OP}} = \gamma\mu + (1-\gamma)(2e^{\mathrm{OP}})$. In the transparent organization, by Lemma 8, the agent always adopts the new idea in the first period. Thus, we simply write $\hat{\omega}$ and e_1^{TR} instead of $\hat{\omega}(e_1^{\mathrm{TR}}(N))$ and

[16] We assume the second-order condition is met. A sufficient condition is given in Appendix B.

$e_1^{\mathrm{TR}}(N)$, respectively. Then the expected final output is written as

$$\begin{aligned}\Pi^{\mathrm{TR}} &= \gamma\left(\mu + \int_{\hat{\omega}}^{\infty}(E[\theta_1 \mid \omega] - \mu)dF_\omega\right) + (1-\gamma)\{[1-(1-\rho)F_\omega(\hat{\omega})]e_1^{\mathrm{TR}} + e^{\mathrm{OP}}\} \\ &= \gamma\left(\mu + \frac{\gamma\sigma_\theta^2}{\gamma^2\sigma_\theta^2 + \sigma_\varepsilon^2}\int_{\hat{\omega}}^{\infty}(\omega - \gamma\mu)dF_\omega\right) + (1-\gamma)\{[1-(1-\rho)F_\omega(\hat{\omega})]e_1^{\mathrm{TR}} + e^{\mathrm{OP}}\},\end{aligned}$$

where the second equality holds by (17) and $\omega \sim N(\gamma\mu, \gamma^2\sigma_\theta^2 + \sigma_\varepsilon^2)$. The difference in the expected final output is written as

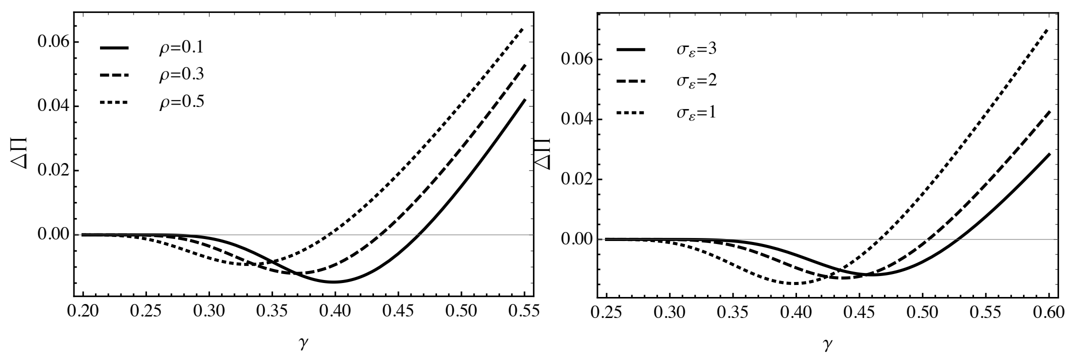
$$\Delta\Pi = \gamma\left(\frac{\gamma\sigma_\theta^2}{\gamma^2\sigma_\theta^2 + \sigma_\varepsilon^2}\int_{\hat{\omega}}^{\infty}(\omega - \gamma\mu)dF_\omega\right) - (1-\gamma)\{e^{\mathrm{OP}} - [1-(1-\rho)F_\omega(\hat{\omega})]e_1^{\mathrm{TR}}\},$$

where the first term captures the sorting effect, and the second term captures the wasting and demotivating effects. The coefficient in the first term $\frac{\gamma\sigma_\theta^2}{\gamma^2\sigma_\theta^2+\sigma_\varepsilon^2}$ is different from (11). It represents the degree to which the sorting effect contributes to $\Delta\Pi$, which is increasing in σ_θ and decreasing in σ_ε.

Figure 5 illustrates that our key results extend to the normal-normal case. The left-hand figure shows the effect of the knowledge generality ρ on $\Delta\Pi$. The solid curve depicts $\Delta\Pi$ as a function of γ for $\rho = 0.1$, the dashed curve is for $\rho = 0.3$, and the dotted curve for $\rho = 0.5$. As in the uniform-uniform case, as ρ rises, the region of γ, in which $\Delta\Pi > 0$ becomes greater, but $\Delta\Pi$ decreases for sufficiently small γ.

The right-hand figure shows the effect of the signal precision σ_ε on $\Delta\Pi$. The solid curve depicts $\Delta\Pi$ as a function of γ for $\sigma_\varepsilon = 3$, the dashed curve is for $\sigma_\varepsilon = 2$, and the dotted curve for $\sigma_\varepsilon = 1$. As the interim performance measure becomes more precise (as σ_ε becomes smaller), the region of γ in which $\Delta\Pi > 0$ becomes greater, but for sufficiently small γ, $\Delta\Pi$ decreases. We set $\mu = 2$ and $\lambda = 0.5$, as in Figure 1 for both the figures, $\sigma_\varepsilon^2 = \sigma_\theta^2 = 1$ for the left-hand figure, and $\rho = 0.1$ and $\sigma_\theta^2 = 1$ for the right-hand figure.

Figure 5: The effects of knowledge generality ρ and signal precision σ_ε on $\Delta\Pi$.



5 Conclusion

One of the challenges that organizations face in promoting innovation is that innovation activities often involve the selection of ideas and projects as well as effort exertion, which are hard to verify and write a contract upon. Organizations are therefore left with few tools to properly manage innovation activities.

We investigated the relationship between transparency and the incentive to implement ideas. Transparency increases incentives to explore new drastic ideas through efficient idea sorting, that is, retaining good ideas and replacing bad ideas. However, the incentive to implement new ideas is weakened if employees expect their ideas to be replaced. We found that transparency benefits the organization if and only if the output relies heavily on idea quality rather than the effort to implement the ideas.

Our analysis provides guidance as to how the optimal organizational design for transparency should vary according to the nature of the innovation process. If innovation performance relies heavily on the idea quality (resp. the idea implementation), transparency is beneficial (resp. harmful). This finding provides a theoretical foundation for why it is important for companies, like Netflix, to create a transparent organizational culture. Furthermore, if innovation output depends heavily on the idea implementation, transparency becomes more harmful when the knowledge and know-how gained during idea development is less idea-specific and when more precise interim performance signals are available.

This model can also be interpreted as a model for employee development feedback, in which a subordinate chooses to either explore new methods or exploit known methods, the match quality between the employee and the work method is uncertain, and the supervisor, who receives interim performance measures, commits ex ante to providing either full feedback or none at all.

Our analysis suggests that giving full feedback is generally helpful for employee development. However, full feedback hurts effort incentives, and may be counter-productive if employees' tasks require relatively more effort than the match quality between the employee and the work method (e.g., non-managerial or routine tasks). Furthermore, full feedback can be more counter-productive when the acquired knowledge is more general or when it is provided by a person who can receive more precise interim performance measures (e.g., specialists, not generalists).

Appendix A: Proofs

Proof of Lemma 2 It follows from (3) and (4) that

$$\frac{\partial \pi(N)}{\partial e_1} = \lambda(1-\gamma) - (1-\rho)F_\theta(\hat{\theta}(e_1))\lambda(1-\gamma) - e_1$$
$$\frac{\partial^2 \pi(N)}{\partial e_1^2} = -(1-\rho)f_\theta(\hat{\theta}(e_1))\lambda(1-\gamma)\frac{\partial \hat{\theta}(e_1)}{\partial e_1} - 1 = (1-\rho)^2 f_\theta(\hat{\theta}(e_1))\lambda(1-\gamma)^2\frac{1}{\gamma} - 1.$$

If $\hat{\theta}(e_1) \leq 0$, then $F_\theta(\hat{\theta}(e_1)) = f_\theta(\hat{\theta}(e_1)) = 0$ and the first-and second-order conditions (FOC and SOC) imply that $e_1^{\text{TR}}(N) = e^{\text{OP}}$.

Suppose $\hat{\theta}(e_1) > 0$. The SOC reduces to $\frac{\partial^2 \pi(N)}{\partial e_1^2} = \frac{(1-\rho)^2\lambda(1-\gamma)^2}{2\gamma\mu} - 1 < 0$, or equivalently, $2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)^2 > 0$. Let γ^{SOC} and γ_{A} be the solution to $2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)^2 = 0$ and $\gamma\mu - \lambda(1-\gamma)^2(1-\rho) = 0$, respectively. There are three cases to consider: (i) $\gamma > \gamma_{\text{A}}$, (ii) $\gamma \in (\gamma^{\text{SOC}}, \gamma_{\text{A}}]$, and (iii) $\gamma \leq \gamma^{\text{SOC}}$. To observe that they are mutually exclusive and collectively exhaustive, note that

$$\gamma\mu - \lambda(1-\gamma)^2\,(1-\rho) \leq 2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)^2$$

holds and both sides are negative at $\gamma = 0$, positive at $\gamma = 1$, and increasing in γ. Therefore, $0 < \gamma^{\text{SOC}} < \gamma_{\text{A}} < 1$, the SOC is satisfied if and only if $\gamma > \gamma^{\text{SOC}}$, and $\gamma\mu - \lambda(1-\gamma)^2\,(1-\rho) > 0$ if and only if $\gamma > \gamma_{\text{A}}$.

Consider case (i): $\gamma > \gamma_{\text{A}}$. The SOC is satisfied; hence, the FOC (5) and (3) yield (7) and (8). Moreover, $e_1^{\text{TR}}(N) > 0$ and $\hat{\theta}(e_1^{\text{TR}}(N)) > 0$ are followed by $\gamma > \gamma_{\text{A}} > \gamma^{\text{SOC}}$. Furthermore, the argument preceding the lemma yields $e_1^{\text{TR}}(N) \leq e^{\text{OP}}$. For case (ii): $\gamma \in (\gamma^{\text{SOC}}, \gamma_{\text{A}}]$, the FOC (5) and (3) yield (7), and $\hat{\theta}(e_1) \leq 0$, followed by $\gamma \in (\gamma^{\text{SOC}}, \gamma_{\text{A}}]$, which contradicts $\hat{\theta}(e_1) > 0$. For case (iii): $\gamma \leq \gamma^{\text{SOC}}$, the SOC is violated, hence, increasing e_1 will increase $\pi(N)$. However, this will make $\hat{\theta}(e_1) \leq 0$ by (3), a contradiction. Therefore, for cases (ii) and (iii), i.e., $\gamma \leq \gamma_{\text{A}}$, $\hat{\theta}(e_1) \leq 0$ holds, implying $e_1^{\text{TR}}(N) = e^{\text{OP}}$.

Finally, for $\gamma > \gamma_{\text{A}}$, it follows from (7) and (3) that

$$\begin{aligned}\frac{de_1^{\text{TR}}(N)}{d\gamma} &= -\frac{\mu\lambda(1+\rho)[2\gamma^2\mu + \lambda(1-\gamma)^2(1-\rho)^2]}{[2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)^2]^2} < 0,\\ \frac{d\hat{\theta}(e_1^{\text{TR}}(N))}{d\gamma} &= \frac{1}{\gamma^2}(1-\rho)e_1^{\text{TR}}(N) - \frac{1-\gamma}{\gamma}(1-\rho)\frac{de_1^{\text{TR}}(N)}{d\gamma} > 0. \qquad \square\end{aligned}$$

Proof of Proposition 1 If $\gamma \leq \gamma_{\text{A}}$, by Lemma 2, regardless of $i_1 \in \{N, K\}$, $e_1 = e^{\text{OP}}$ and no switching occurs. Thus, $\pi(N) = \pi(K)$. When the agent is indifferent, by our assumption, he chooses $i_1 = K$.

Suppose $\gamma > \gamma_{\text{A}}$. Let $\pi(N; e_1, \hat{\theta}(e_1))$ be the agent's expected payoff from choosing e_1 and following the switching strategy $\hat{\theta}(e_1)$ while adopting $i_1 = N$.

Thus, recalling that $\hat{\theta} = 0$ means that the switch never occurs, we have

$$\begin{aligned}\pi(N) &= \pi(N; e_1^{\mathrm{TR}}(N), \hat{\theta}(e_1^{\mathrm{TR}}(N))) \\ &> \pi(N; e^{\mathrm{OP}}, \hat{\theta}(e^{\mathrm{OP}})) \\ &\geq \pi(N; e^{\mathrm{OP}}, 0) = \pi(K),\end{aligned}$$

where the first inequality holds because $e_1^{\mathrm{TR}}(N)$ is the unique idea choice in period 1, the second inequality holds because $\hat{\theta}(e^{\mathrm{OP}})$ is the optimal second-period idea choice after $(i_1 = N, e_1 = e^{\mathrm{OP}})$, and the last equality holds because the agent's payoff is identical regardless of i_1 when no switching occurs and the effort levels are equal. □

Proof of Proposition 2 For $\gamma \leq \gamma_{\mathrm{A}}$, $i_1 = K$, hence, $\Delta\Pi = 0$.

In what follows, suppose $\gamma > \gamma_{\mathrm{A}}$. To ease our notation, whenever it is clear, let us simply write e_1^{TR} and $\hat{\theta}$ instead of $e_1^{\mathrm{TR}}(N)$ and $\hat{\theta}(e_1^{\mathrm{TR}}(N))$, respectively. First observe that by (7) we have

$$\begin{aligned}e^{\mathrm{OP}} - e_1^{\mathrm{TR}} &= \lambda(1-\gamma) - \frac{\gamma\mu\lambda(1-\gamma)(1+\rho)}{2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)^2} \\ &= \lambda(1-\gamma)\frac{[2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)^2] - \gamma\mu(1+\rho)}{2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)^2} \\ &= \lambda(1-\gamma)(1-\rho)\frac{\gamma\mu - \lambda(1-\gamma)^2(1-\rho)}{2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)^2} \\ &= \frac{(1-\rho)}{2\mu}\hat{\theta}e^{\mathrm{OP}},\end{aligned} \tag{A1}$$

where the last equality holds by (8) and $e^{\mathrm{OP}} = \lambda(1-\gamma)$. This yields

$$e_1^{\mathrm{TR}}(N) = \left(1 - \frac{1-\rho}{2\mu}\hat{\theta}(e_1^{\mathrm{TR}}(N))\right)e^{\mathrm{OP}}. \tag{A2}$$

Under the uniform assumption, (11) is given by

$$\begin{aligned}\Delta\Pi &= \frac{\gamma}{2\mu}\int_{\hat{\theta}}^{2\mu}(\theta-\mu)d\theta-(1-\gamma)\Big\{e^{\mathrm{OP}}-\Big(1-\frac{1-\rho}{2\mu}\hat{\theta}\Big)e_1^{\mathrm{TR}}\Big\}\\ &= \frac{\gamma(2\mu-\hat{\theta})}{4\mu}\hat{\theta}-(1-\gamma)(e^{\mathrm{OP}}-e_1^{\mathrm{TR}})-(1-\gamma)(1-\rho)e_1^{\mathrm{TR}}\frac{\hat{\theta}}{2\mu}\\ &= \frac{\gamma(2\mu-\hat{\theta})}{4\mu}\hat{\theta}-\lambda(1-\gamma)^2\frac{(1-\rho)}{2\mu}\hat{\theta}-\gamma(\mu-\hat{\theta})\frac{\hat{\theta}}{2\mu}\\ &= \frac{\hat{\theta}}{4\mu}[\gamma\hat{\theta}-2\lambda(1-\gamma)^2(1-\rho)], \qquad \text{(A3)}\end{aligned}$$

where the third equality holds by (A1) and (3). Since $\hat{\theta}>0$ by $\gamma>\gamma_{\mathrm{A}}$, the sign of $\Delta\Pi$ is determined by the expression in the square bracket, denoted by $\psi_{\mathrm{P}}(\gamma)$, i.e.,

$$\psi_{\mathrm{P}}(\gamma)\equiv\gamma\hat{\theta}(e_1^{\mathrm{TR}}(N))-2\lambda(1-\gamma)^2(1-\rho). \qquad \text{(A4)}$$

Then $\psi_{\mathrm{P}}(\gamma)$ is increasing, $\psi_{\mathrm{P}}(\gamma_{\mathrm{A}})<0$, and $\psi_{\mathrm{P}}(1)>0$. Thus, there exists $\gamma_{\mathrm{P}}\in(\gamma_{\mathrm{A}},1)$ with $\psi_{\mathrm{P}}(\gamma_{\mathrm{P}})=0$ such that $\Delta\Pi<0$ if $\gamma\in(\gamma_{\mathrm{A}},\gamma_{\mathrm{P}})$ and $\Delta\Pi>0$ if $\gamma>\gamma_{\mathrm{P}}$. □

Proof of Proposition 3 Differentiating (7) with respect to ρ yields

$$\frac{de_1^{\mathrm{TR}}(N)}{d\rho}=\frac{\gamma\mu\lambda(1-\gamma)\left\{2\gamma\mu-\lambda(1-\gamma)^2(1-\rho)(3+\rho)\right\}}{\left\{2\gamma\mu-\lambda(1-\gamma)^2(1-\rho)^2\right\}^2}, \qquad \text{(A5)}$$

which is positive if and only if $2\gamma\mu-\lambda(1-\gamma)^2(1-\rho)(3+\rho)>0$. The left-hand side is increasing in γ, negative at $\gamma=\gamma_{\mathrm{A}}$, and positive at $\gamma=1$. Hence, there exists $\gamma_2\in(\gamma_{\mathrm{A}},1)$ such that $de_1^{\mathrm{TR}}/d\rho>0$ if and only if $\gamma>\gamma_2$, where γ_2 is the solution to $2\gamma\mu-\lambda(1-\gamma)^2(1-\rho)(3+\rho)=0$.

To prove $\gamma_{\mathrm{P}}>\gamma_2$, it suffices to show that $de_1^{\mathrm{TR}}/d\rho>0$ at γ_{P}. Plugging $\psi_{\mathrm{P}}(\gamma_{\mathrm{P}})=0$ into (3) yields $\frac{2\lambda(1-\gamma_{\mathrm{P}})^2}{\gamma_{\mathrm{P}}}(1-\rho)=\mu-\frac{(1-\gamma_{\mathrm{P}})}{\gamma_{\mathrm{P}}}(1-\rho)e_1^{\mathrm{TR}}$, which reduces to $e_1^{\mathrm{TR}}=\frac{\gamma_{\mathrm{P}}}{1-\gamma_{\mathrm{P}}}\Big(\frac{\mu}{1-\rho}-\frac{2\lambda(1-\gamma_{\mathrm{P}})^2}{\gamma_{\mathrm{P}}}\Big)$, which is increasing in ρ.

We now consider $d\Delta\Pi/d\rho$, which equals $d\Pi^{\mathrm{TR}}/d\rho$ because ρ has an impact only after $i_1 = N$. Plugging (7) and (A5) into (13) under the uniform assumption yields

$$\begin{aligned}\frac{d\Pi^{\mathrm{TR}}}{d\rho} &= \frac{1-\gamma}{\{2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)^2\}^2}\left\{\frac{\hat{\theta}}{2\mu}\lambda\gamma\mu(1-\gamma)(1+\rho)[2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)^2]\right.\\ &\quad \left. + \left(1 - \frac{(1-\rho)\hat{\theta}}{2\mu}\right)\lambda\gamma\mu(1-\gamma)[2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)(3+\rho)]\right\}\\ &= \frac{\lambda\gamma\mu(1-\gamma)^2}{\{2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)^2\}^2}\\ &\quad \times \Big\{\underbrace{2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)(3+\rho) + \left(2\rho\gamma + \frac{\lambda}{\mu}(1-\gamma)^2(1-\rho)^2\right)\hat{\theta}}_{\equiv D(\gamma)}\Big\}.\end{aligned}$$

Thus, the sign of $\frac{d\Pi^{\mathrm{TR}}}{d\rho}$ equals that of the expression in the curly bracket, denoted by $D(\gamma)$. The definition of γ_2 implies that $2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)(3+\rho)$ is negative at $\gamma = \gamma_{\mathrm{A}}$ and zero at $\gamma = \gamma_2$. Additionally, $\hat{\theta}$ is zero at $\gamma = \gamma_{\mathrm{A}}$ and positive at $\gamma = \gamma_2$. Thus, $\frac{d\Pi^{\mathrm{TR}}}{d\rho} < 0$ at $\gamma = \gamma_{\mathrm{A}}$ and $\frac{d\Pi^{\mathrm{TR}}}{d\rho} > 0$ at $\gamma = \gamma_2$. Hence, to prove that there exists $\gamma_1 \in (\gamma_{\mathrm{A}}, \gamma_2)$ such that $\frac{d\Pi^{\mathrm{TR}}}{d\rho} < 0$ if and only if $\gamma < \gamma_1$, it suffices to show that $D(\gamma)$ is increasing in γ over $(\gamma_{\mathrm{A}}, \gamma_2)$. Indeed, we have

$$\begin{aligned}D'(\gamma) &= 2\mu + 2\lambda(1-\gamma)(1-\rho)(3+\rho) + \left(2\rho - 2\frac{\lambda}{\mu}(1-\gamma)(1-\rho)^2\right)\hat{\theta}\\ &\quad + \left(2\rho\gamma + \frac{\lambda}{\mu}(1-\gamma)^2(1-\rho)^2\right)\frac{d\hat{\theta}}{d\gamma}\\ &= 2\mu + 2\lambda(1-\gamma)(1-\rho)\left(3+\rho-(1-\rho)\frac{\hat{\theta}}{\mu}\right) + 2\rho\hat{\theta}\\ &\quad + \left(2\rho\gamma + \frac{\lambda}{\mu}(1-\gamma)^2(1-\rho)^2\right)\frac{d\hat{\theta}}{d\gamma},\end{aligned}$$

which is positive because $3 + \rho > \frac{(1-\rho)\hat{\theta}}{\mu}$ by $\hat{\theta} < \mu$ and $\frac{d\hat{\theta}}{d\gamma} > 0$ by Lemma 2.

We now prove $d\gamma_{\mathrm{P}}/d\rho < 0$. Differentiating (8) with respect to ρ yields

$$\frac{d\hat{\theta}}{d\rho} = \frac{2\lambda\mu(1-\gamma)^2[2\rho\gamma\mu + \lambda(1-\gamma)^2(1-\rho)^2]}{[2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)^2]^2} > 0.$$

Moreover, differentiating both sides of (A4) with respect to ρ yields

$$\frac{d\hat{\theta}}{d\rho} = 2\lambda(1-\rho)\Big(\frac{d}{d\gamma_{\mathrm{P}}}\frac{(1-\gamma_{\mathrm{P}})^2}{\gamma_{\mathrm{P}}}\Big)\frac{d\gamma_{\mathrm{P}}}{d\rho} - \frac{2\lambda(1-\gamma_{\mathrm{P}})^2}{\gamma_{\mathrm{P}}},$$

where $\frac{d}{d\gamma_{\mathrm{P}}}\frac{(1-\gamma_{\mathrm{P}})^2}{\gamma_{\mathrm{P}}} < 0$; hence, $\frac{d\gamma_{\mathrm{P}}}{d\rho}$ must be negative for $\frac{d\hat{\theta}}{d\rho}$ to be positive. Finally, it is immediate to see that $d\gamma_{\mathrm{A}}/d\rho < 0$. □

Proof of Proposition 4 To ease our notation, let us simply write e_1^{TR} instead of $e_1^{\mathrm{TR}}(N)$ and denote $\hat{\theta}(e_1^{\mathrm{TR}}(N))$ by $\hat{\theta}$ or by $\hat{\theta}(\gamma)$ when we want to be explicit about the value of γ. We have $\psi_{\mathrm{P}}(\gamma_{\mathrm{FB}}) = \gamma_{\mathrm{FB}}\hat{\theta}(\gamma_{\mathrm{FB}}) - 2\lambda(1-\gamma_{\mathrm{FB}})^2(1-\rho) = \gamma_{\mathrm{FB}}[\hat{\theta}(\gamma_{\mathrm{FB}}) - 2\lambda\mu]$, where the last equality holds by (14). Thus, the sign of $\psi_{\mathrm{P}}(\gamma_{\mathrm{FB}})$ is determined by $\hat{\theta}(\gamma_{\mathrm{FB}}) - 2\lambda\mu$. At $\lambda = 0$, $\hat{\theta}(\gamma_{\mathrm{FB}}) - 2\lambda\mu = \mu > 0$; while at $\lambda = 1/2$, $\hat{\theta}(\gamma_{\mathrm{FB}}) - 2\lambda\mu = \hat{\theta}(\gamma_{\mathrm{FB}}) - \mu < 0$ since $\hat{\theta} < \mu$ for $e_1 > 0$. Moreover, $\hat{\theta}(\gamma_{\mathrm{FB}}) - 2\lambda\mu$ is decreasing in λ because $\hat{\theta}$ is decreasing in λ. Therefore, there exists $\lambda^* \in (0, 1/2)$ such that $\lambda < \lambda^*$ if and only if $\psi_{\mathrm{P}}(\gamma_{\mathrm{FB}}) > 0 = \psi_{\mathrm{P}}(\gamma_{\mathrm{P}})$, which is equivalent to $\gamma_{\mathrm{FB}} > \gamma_{\mathrm{P}}$ because $\psi_{\mathrm{P}}(\gamma)$ is increasing in γ. □

Proof of Proposition 5 As in the proof of Proposition 4, let us simply write e_1^{TR} and $\hat{\theta}$ instead of $e_1^{\mathrm{TR}}(N)$ and $\hat{\theta}(e_1^{\mathrm{TR}}(N))$, respectively. Since $\Delta W = 0$ for $\gamma \leq \gamma_{\mathrm{A}}$, suppose $\gamma > \gamma_{\mathrm{A}}$. To derive ΔW, let us first characterize its components: By (A2),

$$c(e^{\mathrm{OP}}) - c(e_1^{\mathrm{TR}}) = \frac{1}{2}(e^{\mathrm{OP}} - e^{\mathrm{TR}})(e^{\mathrm{OP}} + e^{\mathrm{TR}}) = \frac{\lambda^2(1-\gamma)^2(1-\rho)}{4\mu}\hat{\theta}\Big(2 - \frac{1-\rho}{2\mu}\hat{\theta}\Big).$$

By using this and (A3), we can rewrite $\pi(N) - \pi(K)$ as follows:

$$\begin{aligned}\pi(N) - \pi(K) &= \frac{\lambda\hat{\theta}}{4\mu}\left[\gamma\hat{\theta} - 2\lambda(1-\gamma)^2(1-\rho) + \lambda(1-\gamma)^2(1-\rho)\Big(2 - \frac{1-\rho}{2\mu}\hat{\theta}\Big)\right] \\ &= \frac{\lambda\hat{\theta}^2}{8\mu^2}\left[2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)^2\right]. \qquad \text{(A6)}\end{aligned}$$

Plugging (A3) and (A6) into (3.4) yields

$$\Delta W = \frac{\hat{\theta}}{4\mu}\left[\hat{\theta}\left(\frac{\lambda[2\gamma\mu - \lambda(1-\gamma)^2(1-\rho)^2]}{2\mu} + (1-\lambda)\gamma\right) - 2\lambda(1-\lambda)(1-\gamma)^2(1-\rho)\right].$$

Let $\psi_{\mathrm{SB}}(\gamma)$ be the expressions in the square bracket in this equation. Since $\hat{\theta} > 0$, the sign of $\psi_{\mathrm{SB}}(\gamma)$ is the same as the sign of ΔW. First, since $\hat{\theta}$ is increasing in γ, so is $\psi_{\mathrm{SB}}(\gamma)$. Second, $\psi_{\mathrm{SB}}(\gamma_{\mathrm{A}}) = -2\lambda(1-\lambda)(1-\gamma)^2(1-\rho) < 0$ because $\hat{\theta} = 0$ at $\gamma = \gamma_{\mathrm{A}}$. Finally, $\psi_{\mathrm{SB}}(\gamma_{\mathrm{P}}) > 0$ because $\Delta W > 0$ ($\pi(N) - \pi(K) > 0$ and $\Delta\Pi = 0$) at $\gamma = \gamma_{\mathrm{P}}$. Therefore, $\gamma_{\mathrm{SB}} \in (\gamma_{\mathrm{A}}, \gamma_{\mathrm{P}})$ follows. □

References

Aghion, Philippe and Jean Tirole (1994) "The Management of Innovation," *Quarterly Journal of Economics*, Vol. 109, pp. 1185–1209, URL: `http://www.jstor.org/stable/2118360`.

Aguinis, Herman and Charles A Pierce (2008) "Enhancing the relevance of organizational behavior by embracing performance management research," *Journal of Organizational Behavior*, Vol. 29, No. 1, pp. 139–145.

Amabile, Teresa M., Teresa M. Amabile, Mary Ann Collins, Regina Conti, Elise Phillips, Martha Picariello, John Ruscio, and Dean Whitney eds. (1996) *Creativity in Context: Update to The Social Psychology of Creativity*, New York, NY: Routledge.

Anderson, Neil, Kristina Potočnik, and Jing Zhou (2014) "Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework," *Journal of Management*, Vol. 40, No. 5, pp. 1297–1333.

Azoulay, Pierre, Joshua S. Graff Zivin, and Gustavo Manso (2011) "Incentives and Creativity: Evidence from the Academic Life Sciences," *RAND Journal of Economics*, Vol. 42, No. 3, pp. 527–554, URL: http://dx.doi.org/10.1111/j.1756-2171.2011.00140.x.

Baer, Markus (2012) "Putting creativity to work: The implementation of creative ideas in organizations," *Academy of Management Journal*, Vol. 55, No. 5, pp. 1102–1119.

Bernstein, Ethan S (2012) "The transparency paradox: A role for privacy in organizational learning and operational control," *Administrative Science Quarterly*, Vol. 57, No. 2, pp. 181–216.

Brown, James R and Gustav Martinsson (2019) "Does Transparency Stifle or Facilitate Innovation?" *Management Science*, Vol. 65, No. 4, pp. 1600–1623.

Cato, Susumu and Akifumi Ishihara (2017) "Transparency and Performance Evaluation in Sequential Agency," *Journal of Law, Economics, and Organization*, Vol. 33, No. 3, pp. 475–506.

Crémer, Jacques (1995) "Arm's length relationships," *Quarterly Journal of Economics*, Vol. 110, No. 2, pp. 275–295.

Deci, Edward L. (1972) "The effects of contingent and noncontingent rewards and controls on intrinsic motivation," *Organizational Behavior and Human Performance*, Vol. 8, No. 2, pp. 217–229, URL: https://www.sciencedirect.com/science/article/pii/0030507372900475, DOI: http://dx.doi.org/https://doi.org/10.1016/0030-5073(72)90047-5https://doi.org/10.1016/0030-5073(72)90047-5.

Ederer, Florian and Gustavo Manso (2013) "Is Pay-for-Performance Detrimental to Innovation?" *Management Science*, Vol. 59, No. 7, pp. 1496–1513.

Gross, Daniel P. (2017) "Performance Feedback in Competitive Product Development," *RAND Journal of Economics*, Vol. 48, pp. 438–466.

Grossman, Sanford J. and Oliver D. Hart (1986) "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, Vol. 94, pp. 691–719.

Hastings, Reed and Erin Meyer (2020) *No rules rules: Netflix and the culture of reinvention*, New York, NY: Penguin Press.

Hellmann, Thomas and Veikko Thiele (2011) "Incentives and Innovation: A Multitasking Approach," *American Economic Journal: Microeconomics*, Vol. 3, pp. 78–128, URL: http://dx.doi.org/10.1257/mic.3.1.78, DOI: http://dx.doi.org/10.1257/mic.3.1.7810.1257/mic.3.1.78.

Holmstrom, Bengt (1989) "Agency Costs and Innovation," *Journal of Economic Behavior & Organization*, Vol. 12, No. 3, pp. 305–327.

Jehiel, Philippe (2015) "On Transparency in Organizations," *Review of Economic Studies*, Vol. 82, No. 2, pp. 736–761.

Levitt, Theodore (1963) "Creativity is not enough," *Harvard Business Review*, Vol. 41, No. 3, pp. 72–83.

Manso, Gustavo (2011) "Motivating Innovation," *Journal of Finance*, Vol. 66, pp. 1823–1860, URL: http://dx.doi.org/10.1111/j.1540-6261.2011.01688.x.

Murphy, Kevin R, Jeanette N Cleveland, and Madison E Hanscom (2018) *Performance appraisal and management*: SAGE Publications.

Onishi, Koichiro, Hideo Owan, and Sadao Nagaoka (2021) "How Do Inventors Respond to Financial Incentives? Evidence from Unanticipated Court Decisions on Employees' Inventions in Japan," *The Journal of Law and Economics*, Vol. 64, No. 2, pp. 301–339.

Prat, Andrea (2005) "The Wrong Kind of Transparency," *American Economic Review*, Vol. 95, No. 3, pp. 862–877.

Pulakos, Elaine D, Rose Mueller Hanson, Sharon Arad, and Neta Moye (2015) "Performance management can be fixed: An on-the-job experiential learning approach for complex behavior change," *Industrial and Organizational Psychology*, Vol. 8, No. 1, pp. 51–76.

Schnackenberg, Andrew K and Edward C Tomlinson (2016) "Organizational transparency: A new perspective on managing trust in organization-stakeholder relationships," *Journal of Management*, Vol. 42, No. 7, pp. 1784–1810.

Wirtz, Julia (2016) “Feedback and Learning in Tournaments,” November, mimeo.

Zhong, Rong Irene (2018) “Transparency and firm innovation,” *Journal of Accounting and Economics*, Vol. 66, No. 1, pp. 67–93.

Appendix B for Online Appendix

OA 0.1 The Uniform-Uniform Model

In this section, we provide proofs of Lemmas 3, 4, and 5, and Propositions 6, and 7 for the uniform-uniform model.

Proof of Lemma 3 The conditional distribution of s given θ is given by

$$F_{s|\theta}(s \mid \theta) = \Pr(\gamma\theta + (1-\gamma)e_1 + \varepsilon \le s \mid \theta) = F_\varepsilon(s - \gamma\theta - (1-\gamma)e_1)$$

and thus, its conditional density is written as $f_{s|\theta}(s \mid \theta) = f_\varepsilon(s-\gamma\theta-(1-\gamma)e_1)$. The joint density function of θ and s is given by $f_{\theta,s}(\theta, s) = f_\varepsilon(s-\gamma\theta-(1-\gamma)e_1)f_\theta(\theta)$. Moreover, the cumulative distribution function $F_s(s)$ of s is given by

$$F_s(s) = \Pr(\gamma\theta + (1-\gamma)e_1 + \varepsilon \le s) = \int_0^{2\mu} F_\varepsilon(s - \gamma\theta - (1-\gamma)e_1)dF_\theta(\theta)$$

and thus, its density is given by

$$f_s(s) = \int_0^{2\mu} f_\varepsilon(s-\gamma\theta-(1-\gamma)e_1)dF_\theta(\theta) = \frac{1}{4\mu\delta}\int_{\max\{0,\frac{s-(1-\gamma)e_1-\delta}{\gamma}\}}^{\min\{2\mu,\frac{s-(1-\gamma)e_1+\delta}{\gamma}\}} d\theta. \quad \text{(OA1)}$$

Since $s-\gamma\theta-(1-\gamma)e_1 \in (-\delta,\delta)$ holds if and only if $\theta \in \left(\frac{s-(1-\gamma)e_1-\delta}{\gamma}, \frac{s-(1-\gamma)e_1+\delta}{\gamma}\right)$, the conditional density function of θ given s is written as

$$f_{\theta|s}(\theta \mid s) = \frac{f_{\theta,s}(\theta,s)}{f_s(s)} = \frac{f_\varepsilon(s-\gamma\theta-(1-\gamma)e_1)f_\theta(\theta)}{\frac{1}{4\delta\mu}\int_{\max\{0,\frac{s-(1-\gamma)e_1-\delta}{\gamma}\}}^{\min\{2\mu,\frac{s-(1-\gamma)e_1+\delta}{\gamma}\}} d\theta} = \frac{1}{\int_{\max\{0,\frac{s-(1-\gamma)e_1-\delta}{\gamma}\}}^{\min\{2\mu,\frac{s-(1-\gamma)e_1+\delta}{\gamma}\}} d\theta} \quad \text{(OA2)}$$

if $\max\{0, \frac{s-(1-\gamma)e_1-\delta}{\gamma}\} \le \theta \le \min\{2\mu, \frac{s-(1-\gamma)e_1+\delta}{\gamma}\}$, otherwise, $f_{\theta|s}(\theta \mid s) = 0$.

Moreover, since $\underline{s}(e_1) + 2\delta < \overline{s}(e_1) - 2\delta$ holds under **A1**, we have (i) for $s > \overline{s}(e_1)-2\delta$, $\min\{2\mu, \frac{s-(1-\gamma)e_1+\delta}{\gamma}\} = 2\mu$ and $\max\{0, \frac{s-(1-\gamma)e_1-\delta}{\gamma}\} = \frac{s-(1-\gamma)e_1-\delta}{\gamma}$; (ii) for $s \in [\underline{s}(e_1)+2\delta, \overline{s}(e_1)-2\delta]$, $\min\{2\mu, \frac{s-(1-\gamma)e_1+\delta}{\gamma}\} = \frac{s-(1-\gamma)e_1+\delta}{\gamma}$ and $\max\{0, \frac{s-(1-\gamma)e_1-\delta}{\gamma}\} = \frac{s-(1-\gamma)e_1-\delta}{\gamma}$, and (iii) for $s < \underline{s}(e_1) + 2\delta$, $\min\{2\mu, \frac{s-(1-\gamma)e_1+\delta}{\gamma}\} = \frac{s-(1-\gamma)e_1+\delta}{\gamma}$ and

$\max\{0, \frac{s-(1-\gamma)e_1-\delta}{\gamma}\} = 0$. Thus, (OA1) and (OA2) can be rewritten as

$$f_s(s) = \begin{cases} \frac{\overline{s}(e_1)-s}{4\gamma\mu\delta} & \text{if } s \in [\overline{s}(e_1) - 2\delta, \overline{s}(e_1)] \\ \frac{1}{2\gamma\mu} & \text{if } s \in [\underline{s}(e_1) + 2\delta, \overline{s}(e_1) - 2\delta] \\ \frac{s-\underline{s}(e_1)}{4\gamma\mu\delta} & \text{if } s \in [\underline{s}(e_1), \underline{s}(e_1) + 2\delta] \end{cases} \tag{OA3}$$

and

$$f_{\theta|s}(\theta \mid s) = \begin{cases} \frac{\gamma}{\overline{s}(e_1)-s} & \text{if } s \in [\overline{s}(e_1) - 2\delta, \overline{s}(e_1)] \\ \frac{\gamma}{2\delta} & \text{if } s \in [\underline{s}(e_1) + 2\delta, \overline{s}(e_1) - 2\delta] \\ \frac{\gamma}{s-(1-\gamma)e_1+\delta} & \text{if } s \in [\underline{s}(e_1), \underline{s}(e_1) + 2\delta]. \end{cases}$$

Therefore, the conditional expectation of θ given s is given by

$$\begin{aligned} E[\theta|s] &= \int_{\max\left\{0, \frac{s-(1-\gamma)e_1-\delta}{\gamma}\right\}}^{\min\left\{2\mu, \frac{s-(1-\gamma)e_1+\delta}{\gamma}\right\}} \theta f_{\theta|s}(\theta \mid s) d\theta \\ &= \begin{cases} \frac{s-(1-\gamma)e_1+2\gamma\mu-\delta}{2\gamma} & \text{if } s \in [\overline{s}(e_1) - 2\delta, \overline{s}(e_1)] \\ \frac{s-(1-\gamma)e_1}{\gamma} & \text{if } s \in [\underline{s}(e_1) + 2\delta, \overline{s}(e_1) - 2\delta] \\ \frac{s-(1-\gamma)e_1+\delta}{2\gamma} & \text{if } s \in [\underline{s}(e_1), \underline{s}(e_1) + 2\delta]. \end{cases} \end{aligned}$$

It is easy to verify that $E[\theta|s]$ is continuous and increasing in s. □

Proof of Lemma 4 Suppose $i_1 = N$. The agent switches to $i_2 = K$ if and only if

$$\lambda\gamma E[\theta \mid s] + \lambda(1-\gamma)(e_1 + e^{\text{OP}}) - c(e^{\text{OP}}) \leq \lambda\gamma\mu + \lambda(1-\gamma)(\rho e_1 + e^{\text{OP}}) - c(e^{\text{OP}}),$$

which is reduced to $\gamma E[\theta \mid s] \leq \gamma\mu - (1-\gamma)(1-\rho)e_1$. Thus, when $(1-\gamma)(1-\rho)e_1 \geq \gamma\mu$ or $e_1 \geq \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$ is chosen, since $E[\theta \mid \underline{s}(e_1)] = 0$, even the agent who receives the worst possible news $\underline{s}(e_1)$ chooses $i_2 = N$.

Next, we consider the idea choice given $e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$. Let $\hat{s}(e_1)$ be s such that $E[\theta \mid s] = \mu - \frac{(1-\gamma)}{\gamma}(1-\rho)e_1 \in (0, \mu)$. Note that $\hat{s}(e_1)$ uniquely exists because $E[\theta \mid s]$ is continuous and strictly increasing in s, and its image is $[0, 2\mu]$.

We first show $\hat{s}(e_1) < \overline{s}(e_1) - 2\delta$. $E[\theta \mid s = \overline{s}(e_1) - 2\delta] = \frac{2\gamma\mu-\delta}{\gamma} > \mu$ holds by $\gamma\mu > \delta$. Moreover, $E[\theta \mid \hat{s}(e_1)] < \mu$ and $E[\theta \mid s]$ are increasing in s, implying that $\hat{s}(e_1) < \overline{s}(e_1) - 2\delta$.

At $s = \underline{s}(e_1) + 2\delta$, the conditional expectation is given by

$$E[\theta \mid s = \underline{s}(e_1) + 2\delta] = \frac{\delta}{\gamma},$$

which is equal to zero at $\delta = 0$, $\mu - \frac{(1-\gamma)}{\gamma}(1-\rho)e_1 = E[\theta \mid \hat{s}(e_1)]$ at $\delta = \gamma\mu - (1-\gamma)(1-\rho)e_1$, and μ at $\delta = \gamma\mu$; and it is strictly increasing in δ. This implies that $\hat{s}(e_1) \in (\underline{s}(e_1) + 2\delta, \overline{s}(e_1) - 2\delta)$ if and only if $\delta < \gamma\mu - (1-\gamma)(1-\rho)e_1$, and $\hat{s}(e_1) \leq \underline{s}(e_1) + 2\delta$ if and only if $\delta \geq \gamma\mu - (1-\gamma)(1-\rho)e_1$.

Thus, if $\delta < \gamma\mu - (1-\gamma)(1-\rho)e_1$ or $e_1 < \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}$, it follows from Lemma 3 that $E[\theta \mid \hat{s}(e_1)] = \frac{\hat{s}(e_1)-(1-\gamma)e_1}{\gamma} = \mu - \frac{(1-\gamma)}{\gamma}(1-\rho)e_1$, or equivalently $\hat{s}(e_1) = \gamma\mu + (1-\gamma)\rho e_1$. Similarly, if $\frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)} \leq e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, then $E[\theta \mid \hat{s}(e_1)] = \frac{\hat{s}(e_1)-(1-\gamma)e_1+\delta}{2\gamma} = \mu - \frac{(1-\gamma)}{\gamma}(1-\rho)e_1$ yields $\hat{s}(e_1) = 2\gamma\mu - (1-\gamma)(1-2\rho)e_1 - \delta$. □

Proof of Lemma 5 Suppose $i_1 = N$. Then, the agent switches to $i_2 = K$ if and only if $s \leq \hat{s}(e_1)$ and $e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$. Given any $\varepsilon > 0$ and $e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, let $\hat{\theta}(\varepsilon, e_1)$ be the ex-post cutoff quality of the new idea, satisfying

$$\hat{s}(e_1) = \gamma\hat{\theta}(\varepsilon, e_1) + (1-\gamma)e_1 + \varepsilon,$$

or

$$\hat{\theta}(\varepsilon, e_1) \equiv \frac{1}{\gamma}\{\hat{s}(e_1) - (1-\gamma)e_1 - \varepsilon\}. \tag{OA4}$$

By the definition, given the realized error ε and effort choice e_1, the agent receives the signal $s \leq \hat{s}(e_1)$ if and only if $\theta \leq \hat{\theta}(\varepsilon, e_1)$. By Lemma 4 we have

$$\hat{\theta}(\varepsilon, e_1) \equiv \begin{cases} \frac{1}{\gamma}[2\gamma\mu - 2(1-\gamma)(1-\rho)e_1 - \delta - \varepsilon] & \text{if } \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)} \leq e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)} \\ \frac{1}{\gamma}[\gamma\mu - (1-\gamma)(1-\rho)e_1 - \varepsilon] & \text{if } e_1 < \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}. \end{cases} \tag{OA5}$$

Let

$$\hat{\varepsilon}(e_1) \equiv \begin{cases} 2\gamma\mu - 2(1-\gamma)(1-\rho)e_1 - \delta & \text{if } \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)} \leq e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)} \\ \delta & \text{if } e_1 < \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}. \end{cases} \tag{OA6}$$

Then $\hat{\theta}(\varepsilon, e_1) \leq 0$ if only if $\varepsilon > \hat{\varepsilon}(e_1)$. To observe this, note that $\hat{\theta}(\varepsilon, e_1) = \frac{1}{\gamma}(\delta - \varepsilon) \geq 0$ for any ε at $e_1 = \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}$. Since $\hat{\theta}(\varepsilon, e_1)$ is decreasing in e_1, this implies that $\hat{\theta}(\varepsilon, e_1) > 0$ for any ε (thus, we let $\hat{\varepsilon}(e_1) = \delta$) if $e_1 < \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}$. However, with

$e_1 > \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}$, $\hat{\theta}(\varepsilon, e_1) < 0$ if and only if $\varepsilon > \hat{\varepsilon}(e_1) = 2\gamma\mu - 2(1-\gamma)(1-\rho)e_1 - \delta$. Hence, given $i_1 = N$ and $e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, the expected final output is given by

$$\begin{aligned} E[x_2 \mid i_1 = N] = & \int_{-\delta}^{\hat{\varepsilon}(e_1)} F_\theta(\hat{\theta}(\varepsilon, e_1))[\gamma\mu + (1-\gamma)(\rho e_1 + e^{\mathrm{OP}})]dF_\varepsilon \\ & + \int_{-\delta}^{\hat{\varepsilon}(e_1)} \int_{\hat{\theta}(\varepsilon,e_1)}^{2\mu} [\gamma\theta + (1-\gamma)(e_1 + e^{\mathrm{OP}})]dF_\theta dF_\varepsilon \\ & + \int_{\hat{\varepsilon}(e_1)}^{\delta} \int_{0}^{2\mu} [\gamma\theta + (1-\gamma)(e_1 + e^{\mathrm{OP}})]dF_\theta dF_\varepsilon \\ = & \gamma \Big\{ \int_{-\delta}^{\hat{\varepsilon}(e_1)} \Big(F_\theta(\hat{\theta}(\varepsilon, e_1))\mu + \int_{\hat{\theta}(\varepsilon,e_1)}^{2\mu} \theta dF_\theta \Big) dF_\varepsilon + (1 - F_\varepsilon(\hat{\varepsilon}(e_1))\mu \Big\} \\ & + (1-\gamma)\Big\{ e_1 \Big(1 - (1-\rho) \int_{-\delta}^{\hat{\varepsilon}(e_1)} F_\theta(\hat{\theta}(\varepsilon, e_1))dF_\varepsilon \Big) + e^{\mathrm{OP}} \Big\}. \end{aligned} \tag{OA7}$$

Upon choosing $i_1 = N$ and $e_1 \geq \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, since no switching occurs, the expected final output is given by $E[x_2 \mid i_1 = N] = \gamma\mu + (1-\gamma)(e_1 + e^{\mathrm{OP}})$. Since the agent's payoff given $i_1 = N$ is $\pi(N) = \lambda E[x_2 \mid i_1 = N] - \frac{1}{2}e_1^2$, differentiation yields

$$\frac{\partial \pi(N)}{\partial e_1} = \begin{cases} \lambda \frac{\partial E[x_2|i_1=N]}{\partial e_1} + \lambda \frac{\partial E[x_2|i_1=N]}{\partial \hat{\varepsilon}} \hat{\varepsilon}'(e_1) + \lambda \frac{\partial E[x_2|i_1=N]}{\partial \hat{\theta}(\varepsilon,e_1)} \frac{d\hat{\theta}(\varepsilon,e_1)}{de_1} - e_1 & \text{if } e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)} \\ \lambda(1-\gamma) - e_1 & \text{if } e_1 \geq \frac{\gamma\mu}{(1-\gamma)(1-\rho)}. \end{cases}$$

This condition can be simplified as follows:

Claim 1. *$\frac{\partial \pi(N)}{\partial e_1} = \lambda \frac{\partial E[x_2|i_1=N]}{\partial e_1} - e_1$ if $e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$.*

Proof. We first prove $\frac{\partial E[x_2|i_1=N]}{\partial \hat{\varepsilon}} \hat{\varepsilon}'(e_1) = 0$. When $e_1 < \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}$, $\hat{\varepsilon}'(e_1) = 0$ by $\hat{\varepsilon}(e_1) = \delta$. When $\frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)} \leq e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, it follows that

$$\begin{aligned} \frac{\partial E[x_2 \mid i_1 = N]}{\partial \hat{\varepsilon}} = & \gamma \Big\{ \Big(F_\theta(\hat{\theta}(\hat{\varepsilon}(e_1), e_1))\mu + \int_{\hat{\theta}(\hat{\varepsilon}(e_1),e_1)}^{2\mu} \theta dF_\theta \Big) f_\varepsilon(\hat{\varepsilon}(e_1)) - f_\varepsilon(\hat{\varepsilon}(e_1))\mu \Big\} \\ & - (1-\gamma)e_1(1-\rho)F_\theta(\hat{\theta}(\hat{\varepsilon}(e_1), e_1)) f_\varepsilon(\hat{\varepsilon}(e_1)), \end{aligned}$$

which is zero because $\hat{\theta}(\hat{\varepsilon}(e_1), e_1) = 0$.

It remains to show that $\frac{\partial E[x_2|i_1=N]}{\partial\hat{\theta}(\varepsilon,e_1)} = 0$. It follows from (OA7) that

$$\frac{\partial E[x_2 \mid i_1 = N]}{\partial\hat{\theta}(\varepsilon,e_1)} = \frac{\gamma}{2\mu}\int_{-\delta}^{\hat{\varepsilon}(e_1)}(\mu - \hat{\theta}(\varepsilon,e_1))dF_\varepsilon - \frac{(1-\gamma)(1-\rho)e_1}{2\mu}F_\varepsilon(\hat{\varepsilon}(e_1)).$$

When $e_1 < \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}$, by $\hat{\varepsilon}(e_1) = \delta$ and (OA5), the right-hand side is reduced to

$$\frac{\gamma}{2\mu}\int_{-\delta}^{\delta}\Big(\mu - \frac{1}{\gamma}[\gamma\mu - (1-\gamma)(1-\rho)e_1 - \varepsilon]\Big)dF_\varepsilon - \frac{(1-\gamma)(1-\rho)e_1}{2\mu} = 0.$$

When $\frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)} \leq e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, by (OA5) and (OA6), the right-hand side is reduced to

$$\begin{aligned}&\frac{1}{4\mu\delta}\Big\{\int_{-\delta}^{\hat{\varepsilon}(e_1)}[-\gamma\mu + 2(1-\gamma)(1-\rho)e_1 + \delta + \varepsilon]d\varepsilon - (1-\gamma)(1-\rho)e_1[\hat{\varepsilon}(e_1)+\delta]\Big\}\\ &\quad= \frac{1}{4\mu\delta}\{0\}[\hat{\varepsilon}(e_1)+\delta]\\ &\quad= 0.\end{aligned}$$

□

By Claim 1 and (OA7), we have

$$\frac{\partial\pi(N)}{\partial e_1} = \begin{cases}\lambda(1-\gamma)\Big(1-(1-\rho)\int_{-\delta}^{\hat{\varepsilon}(e_1)}F_\theta(\hat{\theta}(\varepsilon,e_1))dF_\varepsilon\Big) - e_1 & \text{if } e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)}\\ \lambda(1-\gamma) - e_1 & \text{if } e_1 \geq \frac{\gamma\mu}{(1-\gamma)(1-\rho)}.\end{cases} \tag{OA8}$$

For $e_1 < \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}$, plugging $\hat{\varepsilon}(e_1) = \delta$ into (OA8) yields

$$\begin{aligned}\frac{\partial\pi(N)}{\partial e_1} &= \lambda(1-\gamma)\Big(1-(1-\rho)\int_{-\delta}^{\delta}F_\theta(\hat{\theta}(\varepsilon,e_1))dF_\varepsilon\Big) - e_1\\ &= \lambda(1-\gamma) - \lambda(1-\gamma)(1-\rho)\frac{[\gamma\mu - (1-\gamma)(1-\rho)e_1]}{2\gamma\mu} - e_1 \qquad \text{(OA9)}\\ &= \frac{1}{2}\lambda(1-\gamma)(1+\rho) + \Big(\frac{\lambda}{2\gamma\mu}(1-\gamma)^2(1-\rho)^2 - 1\Big)e_1, \qquad \text{(OA10)}\end{aligned}$$

where the second equality holds by (OA5).

For $\frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)} \leq e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, plugging (OA5) into (OA8) yields

$$\begin{aligned}\frac{\partial\pi(N)}{\partial e_1} &= \lambda(1-\gamma) - \lambda(1-\gamma)(1-\rho)\frac{1}{4\gamma\mu\delta}\int_{-\delta}^{\hat{\varepsilon}(e_1)}[2\gamma\mu - 2(1-\gamma)(1-\rho)e_1 - \delta - \varepsilon]d\varepsilon - e_1 \\ &= \lambda(1-\gamma) - \lambda(1-\gamma)(1-\rho)\frac{\hat{\varepsilon}(e_1)+\delta}{4\gamma\mu\delta}\left\{2\gamma\mu - 2(1-\gamma)(1-\rho)e_1 - \frac{\hat{\varepsilon}(e_1)+\delta}{2}\right\} - e_1 \\ &= \lambda(1-\gamma) \\ &\quad - \lambda(1-\gamma)(1-\rho)\frac{[\gamma\mu-(1-\gamma)(1-\rho)e_1]}{2\gamma\mu}\frac{[\gamma\mu-(1-\gamma)(1-\rho)e_1]}{\delta} - e_1,\end{aligned} \tag{OA11}$$

where the last equality holds by (OA6).

Subtracting (OA9) from (OA11) yields

$$-\lambda(1-\gamma)(1-\rho)\frac{[\gamma\mu-(1-\gamma)(1-\rho)e_1]}{2\gamma\mu}\left\{\frac{[\gamma\mu-(1-\gamma)(1-\rho)e_1]}{\delta} - 1\right\},$$

which is positive if $\frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)} \leq e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, where $\delta \geq \gamma\mu - (1-\gamma)(1-\rho)e_1$ holds. Thus, the following inequalities hold:

$$\begin{cases} \frac{\partial\pi(N)}{\partial e_1} \geq \lambda(1-\gamma) - \lambda(1-\gamma)(1-\rho)\frac{[\gamma\mu-(1-\gamma)(1-\rho)e_1]}{2\gamma\mu} - e_1 & \text{if } e_1 \leq \frac{\gamma\mu}{(1-\gamma)(1-\rho)} \\ \frac{\partial\pi(N)}{\partial e_1} = \lambda(1-\gamma) - e_1 & \text{if } e_1 \geq \frac{\gamma\mu}{(1-\gamma)(1-\rho)}. \end{cases} \tag{OA12}$$

There are three cases to consider: Case (i) when $\frac{\partial\pi(N)}{\partial e_1} \geq 0$ at $e_1 = \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, Case (ii) when $\frac{\partial\pi(N)}{\partial e_1} < 0$ at $e_1 = \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$ and $\frac{\partial\pi(N)}{\partial e_1} \geq 0$ at $e_1 = \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}$, and Case (iii) when $\frac{\partial\pi(N)}{\partial e_1} < 0$ both at $e_1 = \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$ and $e_1 = \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}$.

Case (i) Suppose $\left.\frac{\partial\pi(N)}{\partial e_1}\right|_{e_1=\frac{\gamma\mu}{(1-\gamma)(1-\rho)}} = \lambda(1-\gamma) - e_1 = \lambda(1-\gamma) - \frac{\gamma\mu}{(1-\gamma)(1-\rho)} \geq 0$, or $\gamma \leq \gamma_A$. Then, since the right-hand side of the first line in (OA12) is linear in e_1 and positive both at $e_1 = 0$ and $e_1 = \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, $\frac{\partial\pi(N)}{\partial e_1} > 0$ holds for all $e_1 \leq \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$. Since $\frac{\partial\pi(N)}{\partial e_1} = \lambda(1-\gamma) - e_1$ is decreasing in e_1 and $\left.\frac{\partial\pi(N)}{\partial e_1}\right|_{e_1=e^{\mathrm{OP}}} = 0$, $e_1^{\mathrm{TR}}(N) = e^{\mathrm{OP}}$.

Cases (ii) and (iii) Suppose $\left.\frac{\partial \pi(N)}{\partial e_1}\right|_{e_1=\frac{\gamma\mu}{(1-\gamma)(1-\rho)}} < 0$, or $\gamma > \gamma_{\mathrm{A}}$. We first show that $\frac{\partial \pi(N)}{\partial e_1}$ is decreasing in e_1. For $e_1 < \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}$, it follows from (OA10) that $\frac{\partial \pi(N)}{\partial e_1}$ is decreasing because $\frac{\lambda}{2\gamma\mu}(1-\gamma)^2(1-\rho)^2 < \frac{\lambda}{\gamma\mu}(1-\gamma)^2(1-\rho) < 1$, where the last inequality holds by $\gamma > \gamma_{\mathrm{A}}$. For $\frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)} \leq e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, by (OA11) we have

$$\begin{aligned}
\frac{\partial^2 \pi(N)}{\partial e_1^2} &= \frac{\lambda}{\gamma\mu\delta}(1-\gamma)^2(1-\rho)^2[\gamma\mu-(1-\gamma)(1-\rho)e_1]-1 \\
&= \frac{\lambda}{\delta}(1-\gamma)^2(1-\rho)^2 - 1 - \frac{\lambda}{\gamma\mu\delta}(1-\gamma)^3(1-\rho)^3 e_1 \\
&\leq \frac{\lambda}{\delta}(1-\gamma)^2(1-\rho)^2 - 1 - \frac{\lambda}{\gamma\mu\delta}(1-\gamma)^3(1-\rho)^3 \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)} \\
&= -1 + \frac{\lambda}{\gamma\mu}(1-\gamma)^2(1-\rho)^2 \\
&< 0, \qquad \text{(OA13)}
\end{aligned}$$

where the first inequality follows by $\frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)} \leq e_1$ and the last inequality follows by $\gamma > \gamma_{\mathrm{A}}$. For $e_1 \geq \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, $\frac{\partial \pi(N)}{\partial e_1} = \lambda(1-\gamma) - e_1$, which is decreasing in e_1.

Since $\frac{\partial \pi(N)}{\partial e_1}$ is continuous and decreasing in e_1, the first-order conditions uniquely characterizes $e_1^{\mathrm{TR}}(N)$. It follows that

$$\begin{aligned}
\left.\frac{\partial \pi(N)}{\partial e_1}\right|_{e_1=\frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}} &= \frac{\lambda}{2}(1-\gamma)(1+\rho) + \left(\lambda\frac{1}{2\gamma\mu}(1-\gamma)^2(1-\rho)^2 - 1\right)\frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)} \\
&= \frac{\gamma\mu-\delta}{\gamma\mu(1-\gamma)(1-\rho)}\left\{\frac{\lambda\gamma\mu(1-\gamma)^2(1+\rho)(1-\rho)}{2(\gamma\mu-\delta)} + \frac{\lambda(1-\gamma)^2(1-\rho)^2}{2} - 1\right\},
\end{aligned}$$

which is positive if and only if

$$\begin{aligned}
\gamma\mu &\leq \frac{\lambda\gamma\mu}{2(\gamma\mu-\delta)}(1-\gamma)^2(1+\rho)(1-\rho) + \lambda\frac{1}{2}(1-\gamma)^2(1-\rho)^2 \\
&= \lambda(1-\gamma)^2(1-\rho)\left\{\frac{\gamma\mu}{2(\gamma\mu-\delta)}(1+\rho) + \frac{1}{2}(1-\rho)\right\} \\
&= \lambda(1-\gamma)^2(1-\rho)\left\{\frac{2\gamma\mu-(1-\rho)\delta}{2(\gamma\mu-\delta)}\right\}. \qquad \text{(OA14)}
\end{aligned}$$

Thus, when (OA14) is satisfied, which corresponds to Case (ii), e_1 satisfying $\frac{\partial \pi(N)}{\partial e_1} = 0$ must lie in $\frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)} \leq e_1 < \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$, implying that equating (OA11) to zero uniquely characterizes $e_1^{\mathrm{TR}}(N)$, which satisfies

$$e_1^{\mathrm{TR}}(N) = \lambda(1-\gamma) - \lambda(1-\gamma)(1-\rho)\frac{[\gamma\mu - (1-\gamma)(1-\rho)e_1^{\mathrm{TR}}(N)]^2}{2\gamma\mu\delta}. \quad \text{(OA15)}$$

When (OA14) is violated (i.e., Case (iii)), e_1 satisfying $\frac{\partial \pi(N)}{\partial e_1} = 0$ must lie in $e_1 < \frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)}$, implying that equating (OA9) to zero uniquely characterizes $e_1^{\mathrm{TR}}(N)$.

Finally, since $\gamma\mu - \left\{\frac{2\gamma\mu-(1-\rho)\delta}{2\gamma\mu-2\delta}\right\}\lambda(1-\gamma)^2(1-\rho)$ is increasing in γ, inequality (OA14) is equivalent to $\gamma \leq \gamma^*$. Moreover, for $\delta > 0$, by $\frac{2\gamma^*\mu-(1-\rho)\delta}{2\gamma^*\mu-2\delta} > 1$ we have

$$0 = \gamma^*\mu - \left\{\frac{2\gamma^*\mu-(1-\rho)\delta}{2\gamma^*\mu-2\delta}\right\}\lambda(1-\gamma^*)^2(1-\rho) < \gamma^*\mu - \lambda(1-\gamma^*)^2(1-\rho).$$

Since $\gamma\mu - \lambda(1-\gamma)^2(1-\rho)$ is increasing in γ and $\gamma_{\mathrm{A}}\mu - \lambda(1-\gamma_{\mathrm{A}})^2(1-\rho) = 0$ by the definition of γ_{A}, $\gamma^* > \gamma_{\mathrm{A}}$ follows. For $\delta = 0$, clearly $\gamma_{\mathrm{A}} = \gamma^*$. To show that γ^* is increasing in δ, differentiating $\gamma\mu - \left\{\frac{2\gamma\mu-(1-\rho)\delta}{2\gamma\mu-2\delta}\right\}\lambda(1-\gamma)^2(1-\rho)$ with respect to δ yields $-\frac{1}{2}\lambda(1-\gamma)^2(1-\rho)\frac{(1+\rho)(\gamma\mu)}{(\gamma\mu-\delta)^2} < 0$. Since $\gamma\mu - \left\{\frac{2\gamma\mu-(1-\rho)\delta}{2\gamma\mu-2\delta}\right\}\lambda(1-\gamma)^2(1-\rho)$ is increasing in γ, γ^* must increase if δ increases. □

Proof of Proposition 6 In the proof of Propositions 6 and 7, we simply write e_1^{TR}, $\hat{s}$, $\underline{s}$, and $\overline{s}$ instead of $e_1^{\mathrm{TR}}(N)$, $\hat{s}(e_1^{\mathrm{TR}}(N))$, $\underline{s}(e_1^{\mathrm{TR}}(N))$, and $\overline{s}(e_1^{\mathrm{TR}}(N))$, respectively. By Lemmas 4 and 5, $\frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)} < e_1^{\mathrm{TR}} < \frac{\gamma\mu}{(1-\gamma)(1-\rho)}$ and $\hat{s} \in [\underline{s}, \underline{s} + 2\delta]$. Thus, by Lemma 4, $\hat{s} = 2\gamma\mu - (1-\gamma)(1-2\rho)e_1^{\mathrm{TR}} - \delta$; and by (OA3), $F_s(s) = \int_{\underline{s}}^{s} \frac{t-(1-\gamma)e_1+\delta}{4\gamma\mu\delta}dt = \frac{(s-(1-\gamma)e_1+\delta)^2}{8\gamma\mu\delta}$, yielding

$$F_s(\hat{s}) = \frac{(\hat{s} - \underline{s})^2}{8\gamma\mu\delta}. \quad \text{(OA16)}$$

At $e_1 = e_1^{\mathrm{TR}}$, applying the implicit function theorem to (OA15) yields

$$\begin{aligned}\frac{de_1^{\mathrm{TR}}}{d\rho} &= \frac{\partial^2\pi(N)/\partial e_1\partial\rho}{-\partial^2\pi(N)/\partial e_1^2}\bigg|_{e_1=e_1^{\mathrm{TR}}} \\ &= \frac{\frac{\lambda(1-\gamma)}{2\gamma\mu\delta}A^2 - \frac{\lambda(1-\gamma)^2(1-\rho)}{\gamma\mu\delta}Ae_1^{\mathrm{TR}}}{1-\frac{\lambda(1-\gamma)^2(1-\rho)^2}{\gamma\mu\delta}A} \\ &= \frac{\lambda(1-\gamma)A^2 - 2\lambda(1-\gamma)^2(1-\rho)Ae_1^{\mathrm{TR}}}{2\gamma\mu\delta - 2\lambda(1-\gamma)^2(1-\rho)^2A} \\ &= \frac{\lambda(1-\gamma)A[\gamma\mu - 3(1-\gamma)(1-\rho)e_1^{\mathrm{TR}}]}{2\gamma\mu\delta - 2\lambda(1-\gamma)^2(1-\rho)^2A}, \end{aligned} \tag{OA17}$$

where $A \equiv \gamma\mu - (1-\gamma)(1-\rho)e_1^{\mathrm{TR}} > 0$ and $-\frac{\partial^2\pi(N)}{\partial e_1^2}\Big|_{e_1=e_1^{\mathrm{TR}}} = 1 - \frac{\lambda(1-\gamma)^2(1-\rho)^2}{\gamma\mu\delta}A \geq 0$ by the SOC (OA13). Then, applying the envelop theorem yields

$$\begin{aligned}\frac{d\Pi^{\mathrm{TR}}}{d\rho} &= \frac{\partial\Pi^{\mathrm{TR}}}{\partial\rho} + \frac{\partial\Pi^{\mathrm{TR}}}{\partial e_1}\frac{de_1^{\mathrm{TR}}}{d\rho} \\ &= (1-\gamma)F_s(\hat{s})e_1^{\mathrm{TR}} + \frac{e_1^{\mathrm{TR}}}{\lambda}\frac{de_1^{\mathrm{TR}}}{d\rho} \\ &= \frac{1-\gamma}{2\gamma\mu\delta}A^2e_1^{\mathrm{TR}} + e_1^{\mathrm{TR}}\frac{(1-\gamma)A[\gamma\mu - 3(1-\gamma)(1-\rho)e_1^{\mathrm{TR}}]}{2\gamma\mu\delta - 2\lambda(1-\gamma)^2(1-\rho)^2A}, \end{aligned} \tag{OA18}$$

where the second equality holds by $\frac{\partial\pi(N)}{\partial e_1} = \lambda\frac{\partial\Pi^{\mathrm{TR}}}{\partial e_1} - c'(e_1) = 0$ at $e_1 = e_1^{\mathrm{TR}}$ and the third equality holds by (OA16), (OA17), and $\frac{\hat{s}-\underline{s}}{2} = \frac{2\gamma\mu-(1-\gamma)(1-2\rho)e_1^{\mathrm{TR}}-\delta-((1-\gamma)e_1^{\mathrm{TR}}-\delta)}{2} = A$. It follows that (OA18) is positive if and only if

$$A + \frac{[\gamma\mu - 3(1-\gamma)(1-\rho)e_1^{\mathrm{TR}}]}{1-\frac{\lambda(1-\gamma)^2(1-\rho)^2}{\gamma\mu\delta}A} > 0.$$

Multiplying this by the denominator of the second term, which is positive, yields

$$
\begin{aligned}
&A - \frac{1}{\gamma\mu\delta}\lambda(1-\gamma)^2(1-\rho)^2 A^2 + \gamma\mu - 3(1-\gamma)(1-\rho)e_1^{\mathrm{TR}} \\
&= 2\gamma\mu - 4(1-\gamma)(1-\rho)e_1^{\mathrm{TR}} - \frac{1}{\gamma\mu\delta}\lambda(1-\gamma)^2(1-\rho)^2 A^2 \\
&= 2\gamma\mu - 4(1-\gamma)(1-\rho)e_1^{\mathrm{TR}} + 2(1-\gamma)(1-\rho)e_1^{\mathrm{TR}} - 2\lambda(1-\gamma)^2(1-\rho) \\
&= 2\underbrace{\left[\gamma\mu - \lambda(1-\gamma)^2(1-\rho) - (1-\gamma)(1-\rho)\, e_1^{\mathrm{TR}}\right]}_{\equiv \phi_\rho(\gamma)}.
\end{aligned}
$$

where the second equality holds by (OA15).

Since the sign of $d\Pi^{\mathrm{TR}}/d\rho$ is the same as that of $\phi_\rho(\gamma)$, we shall investigate the properties of $\phi_\rho(\gamma)$ below. We first show that $\frac{de_1^{\mathrm{TR}}}{d\gamma} \leq 0$, which implies that $\phi_\rho(\gamma)$ is increasing over $(\gamma_{\mathrm{A}}, \gamma^*)$. Since $\frac{de_1^{\mathrm{TR}}}{d\gamma} = \frac{\partial^2\pi(N)/\partial e_1\partial\gamma}{-\partial^2\pi(N)/\partial e_1^2}\Big|_{e_1=e_1^{\mathrm{TR}}}$ follows by applying the implicit function theorem to (OA15) and $-\frac{\partial^2\pi(N)}{\partial e_1^2}\Big|_{e_1=e_1^{\mathrm{TR}}} \geq 0$, it suffices to show that $\frac{\partial^2\pi(N)}{\partial e_1\partial\gamma}\Big|_{e_1=e_1^{\mathrm{TR}}}$ is negative. By (OA11) and (OA15), we have

$$
\begin{aligned}
\frac{\partial^2\pi(N)}{\partial e_1\partial\gamma}\bigg|_{e_1=e_1^{\mathrm{TR}}} &= -\lambda + \lambda(1-\rho)\frac{A^2}{2\gamma^2\mu\delta} - \lambda(1-\gamma)(1-\rho)\frac{A[\mu + (1-\rho)e_1^{\mathrm{TR}}]}{\gamma\mu\delta} \\
&= -\frac{\lambda}{2\gamma^2\mu\delta}\left\{2\gamma^2\mu\delta - (1-\rho)A^2 + 2(1-\gamma)(1-\rho)A[\gamma\mu + \gamma(1-\rho)e_1^{\mathrm{TR}}]\right\} \\
&< -\frac{\lambda}{2\gamma^2\mu\delta}\left\{2\gamma^2\mu A - (1-\rho)A^2 + 2(1-\gamma)(1-\rho)A[\gamma\mu + \gamma(1-\rho)e_1^{\mathrm{TR}}]\right\} \\
&= -\frac{\lambda A}{2\gamma^2\mu\delta}\left\{2\gamma^2\mu - (1-\rho)A + 2(1-\gamma)(1-\rho)[\gamma\mu + \gamma(1-\rho)e_1^{\mathrm{TR}}]\right\} \\
&= -\frac{\lambda A}{2\gamma^2\mu\delta}\left\{[2\gamma\rho + (1-\rho)]\gamma\mu + (2\gamma+1)(1-\gamma)(1-\rho)^2 e_1^{\mathrm{TR}}\right\} \\
&< 0,
\end{aligned}
$$

where the first inequality holds by $A < \delta$ because $\frac{\gamma\mu-\delta}{(1-\gamma)(1-\rho)} < e_1^{\mathrm{TR}}$. Thus, $\phi_\rho(\gamma)$ is increasing over $(\gamma_{\mathrm{A}}, \gamma^*)$.

Finally, $\phi_\rho(\gamma_A) = -(1-\gamma_A)(1-\rho)\,e_1^{TR} < 0$. Moreover,

$$\phi_\rho(\gamma^*) = \gamma^*\mu - \lambda(1-\gamma^*)^2(1-\rho) - (1-\gamma^*)(1-\rho)\,e_1^{TR} = \delta - \lambda(1-\gamma^*)^2(1-\rho),$$

where the last equality holds by $e_1^{TR} = \frac{\gamma^*\mu-\delta}{(1-\gamma^*)(1-\rho)}$ at $\gamma = \gamma^*$. Let $\delta_\rho \equiv \lambda(1-\gamma^*)^2(1-\rho)$. Then for $\delta \le \delta_\rho$, $\phi_\rho(\gamma) < 0$ over (γ_A, γ^*). For $\delta > \delta_\rho$, there exists $\gamma_\rho \in (\gamma_A, \gamma^*)$ such that $\phi_\rho(\gamma) < 0$ for $\gamma \in (\gamma_A, \gamma_\rho)$ and $\phi_\rho(\gamma) \ge 0$ for $\gamma \in [\gamma_\rho, \gamma^*)$. □

Proof of Proposition 7 Note that e_1^{TR} and $\hat{s}$ remains the same, as in Proposition 6. With the notation $A \equiv \gamma\mu - (1-\gamma)(1-\rho)e_1^{TR}$, (15) can be rewritten as follows:

$$\Pi^{TR} = \gamma\Big(\mu - \int_{\underline{s}}^{\hat{s}} E\left[\theta \mid s\right] dF_s\Big) + AF_s(\hat{s}) + (1-\gamma)\,(e_1^{TR} + e^{OP}).$$

In what follows, we compute each term of (16).

$$\begin{aligned}
\frac{\partial \Pi^{TR}}{\partial \delta} &= -\gamma \frac{\partial}{\partial \delta}\int_{\underline{s}}^{\hat{s}} E\left[\theta \mid s\right] f_s(s) ds + A\frac{\partial}{\partial \delta} F_s(\hat{s}) \\
&= \gamma \frac{\partial \underline{s}}{\partial \delta} E\left[\theta \mid \underline{s}\right] \underbrace{f_s(\underline{s})}_{=0} - \gamma \int_{\underline{s}}^{\hat{s}} \frac{\partial}{\partial \delta} E\left[\theta \mid s\right] f_s(s) ds + A \frac{\partial}{\partial \delta}\frac{(\hat{s}-\underline{s})^2}{8\gamma\mu\delta} \\
&= -\gamma \int_{\underline{s}}^{\hat{s}} \frac{\partial}{\partial \delta}\frac{(s-\underline{s})^2}{8\gamma^2\mu\delta} ds + A\frac{\partial}{\partial \delta}\frac{(\hat{s}-\underline{s})^2}{8\gamma\mu\delta} \\
&= -\int_{\underline{s}}^{\hat{s}} \frac{2(s-\underline{s})\delta - (s-\underline{s})^2}{8\gamma\mu\delta^2} ds + \underbrace{A}_{=\frac{\hat{s}-\underline{s}}{2}} \frac{2(\hat{s}-\underline{s})\delta - (\hat{s}-\underline{s})^2}{8\gamma\mu\delta^2} \\
&= -\int_0^{\hat{s}-\underline{s}} \frac{2x\delta - x^2}{8\gamma\mu\delta^2} dx + \frac{2(\hat{s}-\underline{s})^2\delta - (\hat{s}-\underline{s})^3}{16\gamma\mu\delta^2} \\
&= -\frac{(\hat{s}-\underline{s})^2\delta}{8\gamma\mu\delta^2} + \frac{(\hat{s}-\underline{s})^3}{24\gamma\mu\delta^2} + \frac{2(\hat{s}-\underline{s})^2\delta - (\hat{s}-\underline{s})^3}{16\gamma\mu\delta^2} \\
&= -\frac{(\hat{s}-\underline{s})^3}{48\gamma\mu\delta^2},
\end{aligned}$$

where the second equality holds by (OA16), the third equality holds by (OA3) and Lemma 3, and the fifth equality holds by integration by substitution. Moreover, we can derive $\frac{de_1^{\mathrm{TR}}}{d\delta} = \frac{\partial^2\pi(N)/\partial e_1\partial\delta}{-\partial^2\pi(N)/\partial e_1^2}\Big|_{e_1=e_1^{\mathrm{TR}}}$ and $\frac{\partial\Pi^{\mathrm{TR}}}{\partial e_1} = \frac{e_1^{\mathrm{TR}}}{\lambda}$ analogously to the derivation for (OA17), yielding

$$\frac{\partial\Pi^{\mathrm{TR}}}{\partial e_1}\frac{de_1^{\mathrm{TR}}}{d\delta} = \frac{e_1^{\mathrm{TR}}}{\lambda}\frac{\frac{\lambda(1-\gamma)(1-\rho)A^2}{2\gamma\mu\delta^2}}{1-\frac{\lambda(1-\gamma)^2(1-\rho)^2}{\gamma\mu\delta}A}.$$

Combining these yields

$$\begin{aligned}
\frac{d\Pi^{\mathrm{TR}}}{d\delta} &= -\frac{(\hat{s}-\underline{s})^3}{48\gamma\mu\delta^2} + \frac{e_1^{\mathrm{TR}}}{\lambda}\frac{\frac{\lambda(1-\gamma)(1-\rho)A^2}{2\gamma\mu\delta^2}}{1-\frac{\lambda(1-\gamma)^2(1-\rho)^2}{\gamma\mu\delta}A} \\
&= -\frac{A^3}{6\gamma\mu\delta^2} + e_1^{\mathrm{TR}}\frac{\frac{(1-\gamma)(1-\rho)A^2}{2\gamma\mu\delta^2}}{1-\frac{\lambda(1-\gamma)^2(1-\rho)^2}{\gamma\mu\delta}A} \\
&= \frac{A^2}{6\gamma\mu\delta^2}\left(-A + \frac{3(1-\gamma)(1-\rho)e_1^{\mathrm{TR}}}{1-\frac{\lambda(1-\gamma)^2(1-\rho)^2}{\gamma\mu\delta}A}\right) \\
&= \frac{A^2}{6\gamma\mu\delta^2}\frac{-\gamma\mu + \frac{\lambda(1-\gamma)^2(1-\rho)^2}{\gamma\mu\delta}A^2 + 4(1-\gamma)(1-\rho)e_1^{\mathrm{TR}}}{1-\frac{\lambda(1-\gamma)^2(1-\rho)^2}{\gamma\mu\delta}A} \\
&= \frac{A^2}{6\gamma\mu\delta^2}\frac{-\gamma\mu + 2\lambda(1-\gamma)^2(1-\rho) + 2(1-\gamma)(1-\rho)e_1^{\mathrm{TR}}}{1-\frac{\lambda(1-\gamma)^2(1-\rho)^2}{\gamma\mu\delta}A},
\end{aligned}$$

where the fifth equality holds by (OA15) and $1-\frac{\lambda(1-\gamma)^2(1-\rho)^2}{\gamma\mu d}A \geq 0$ by the SOC (OA13). Thus, the sign of $\frac{d\Pi^{\mathrm{TR}}}{d\delta}$ is determined by

$$\begin{aligned}
\phi_\delta(\gamma) &\equiv -\gamma\mu + 2\lambda(1-\gamma)^2(1-\rho) + 2(1-\gamma)(1-\rho)e_1^{\mathrm{TR}} \\
&= -\gamma\mu + 2(1-\gamma)(1-\rho)[e_1^{\mathrm{TR}} + \lambda(1-\gamma)] \\
&= -\gamma\mu + 2(1-\gamma)(1-\rho)[e_1^{\mathrm{TR}} + e^{\mathrm{OP}}].
\end{aligned}$$

Since the sign of $d\Pi^{\mathrm{TR}}/d\delta$ is the same as that of $\phi_\delta(\gamma)$, we shall investigate the properties of $\phi_\delta(\gamma)$ below. First, $\phi_\delta(\gamma)$ is decreasing over $(\gamma_{\mathrm{A}},\gamma^*)$: $\phi'_\delta(\gamma) = -\mu - 2(1-\rho)[e_1^{\mathrm{TR}} + e^{\mathrm{OP}}] + 2(1-\gamma)(1-\rho)\left(\frac{de_1^{\mathrm{TR}}}{d\gamma} + \frac{de^{\mathrm{OP}}}{d\gamma}\right) < 0$ because $\frac{de_1^{\mathrm{TR}}}{d\gamma} < 0$

and $\frac{de^{\text{OP}}}{d\gamma} < 0$. Second,

$$\phi_\delta(\gamma_{\text{A}}) = -\lambda(1-\gamma_{\text{A}})^2(1-\rho)+2(1-\gamma_{\text{A}})(1-\rho)[2\lambda(1-\gamma_{\text{A}})] = 3\lambda(1-\gamma_{\text{A}})^2(1-\rho) > 0,$$

where the first equality holds because $\gamma_{\text{A}}\mu - \lambda(1-\gamma_{\text{A}})^2(1-\rho) = 0$ and $e_1^{\text{TR}} = e^{\text{OP}} = \lambda(1-\gamma_{\text{A}})$ at $\gamma = \gamma_{\text{A}}$. Finally, it follows from $e_1^{\text{TR}} = \frac{\gamma^*\mu-\delta}{(1-\gamma^*)(1-\rho)}$ at $\gamma = \gamma^*$ and $\gamma^*\mu = \frac{2\gamma^*\mu-(1-\rho)\delta}{2\gamma^*\mu-2\delta}\lambda(1-\gamma^*)^2(1-\rho)$ that

$$\begin{aligned}\phi_\delta(\gamma^*) &= -\tfrac{2\gamma^*\mu-(1-\rho)\delta}{2\gamma^*\mu-2\delta}\lambda(1-\gamma^*)^2(1-\rho) + 2(1-\gamma^*)(1-\rho)\left\{\tfrac{\gamma^*\mu-\delta}{(1-\gamma^*)(1-\rho)} + \lambda(1-\gamma^*)\right\} \\ &= \tfrac{2\gamma^*\mu-(3+\rho)\delta}{2\gamma^*\mu-2\delta}\lambda(1-\gamma^*)^2(1-\rho) + 2(\gamma^*\mu-\delta),\end{aligned}$$

which becomes positive at $\delta \to 0$ and negative as $\delta \to \gamma^*\mu$ while $\frac{d\phi_\delta(\gamma^*)}{d\delta} = \frac{\partial\phi_\delta(\gamma^*)}{\partial\delta} + \frac{\partial\phi_\delta}{\partial\gamma}\frac{\partial\gamma^*}{\partial\delta} < 0$ by $\frac{\partial\phi_\delta(\gamma^*)}{\partial\delta} < 0$, $\frac{\partial\phi_\delta}{\partial\gamma} < 0$, and $\frac{\partial\gamma^*}{\partial\delta} > 0$. Thus, the proposition follows by letting δ_δ be δ such that $\phi_\delta(\gamma^*) = 0$. □

OA 0.2 The Normal-Normal Model

In this section we provide a sufficient condition for the second-order condition regarding e_1 and a proof of Lemma 8 for the normal-normal model.

Sufficient condition for the second-order condition Differentiating the right-hand side of (20) yields

$$\begin{aligned}\frac{\partial^2\pi(N)}{\partial e_1^2} &= -\lambda(1-\gamma)(1-\rho)f_\omega(\hat{\omega}(e_1^{\text{TR}}(N)))\hat{\omega}'(e_1^{\text{TR}}(N)) - 1 \\ &= \lambda(1-\gamma)^2(1-\rho)^2\frac{\gamma^2\sigma_\theta^2+\sigma_\varepsilon^2}{\gamma^2\sigma_\theta^2}f_\omega(\hat{\omega}(e_1^{\text{TR}}(N))) - 1 \\ &\leq \lambda(1-\gamma)^2(1-\rho)^2\frac{\sqrt{\gamma^2\sigma_\theta^2+\sigma_\varepsilon^2}}{\gamma^2\sigma_\theta^2\sqrt{2\pi}} - 1,\end{aligned}$$

where the the last inequality holds because $f_\omega(\cdot)$ is maximized at its mean $\gamma\mu$ and $f_\omega(\gamma\mu) = 1/\sqrt{2\pi(\gamma^2\sigma_\theta^2+\sigma_\varepsilon^2)}$. Thus, the following is a sufficient condition for the second-order condition in order for $e_1^{\mathrm{TR}}(N)$ satisfying (20) to be optimal:

$$1 \geq \lambda(1-\gamma)^2(1-\rho)^2 \frac{\sqrt{\gamma^2\sigma_\theta^2+\sigma_\varepsilon^2}}{\gamma^2\sigma_\theta^2\sqrt{2\pi}}. \tag{OA19}$$

Proof of Lemma 8 As in the proof of Proposition 1, let $\pi(N; e_1, \hat{\omega}(e_1))$ be the agent's expected payoff from choosing e_1 and following the switching strategy $\hat{\omega}(e_1)$ while adopting $i_1 = N$. We have

$$\begin{aligned}\pi(N) &= \pi(N; e_1^{\mathrm{TR}}(N), \hat{\omega}(e_1^{\mathrm{TR}}(N))) \\ &> \pi(N; e_1^{\mathrm{OP}}, \hat{\omega}(e_1^{\mathrm{OP}})) \\ &\geq \lim_{\hat{\omega}\to-\infty} \pi(N; e_1^{\mathrm{OP}}, \hat{\omega}) = \pi(K),\end{aligned}$$

where the first inequality holds because $e_1^{\mathrm{TR}}(N)$ is optimal, the second inequality holds because $\hat{\omega}(e_1)$ is optimal, and the last inequality holds because of the agent's payoff from adopting $i_1 = N$ and never switching ($\hat{\omega} \to -\infty$) is equal to the agent's equilibrium payoff from adopting $i_1 = K$. □