DISCUSSION PAPER SERIES

DP16768

Organizing Data Analytics

Ricardo Alonso and Odilon Câmara

ORGANIZATIONAL ECONOMICS

CEPR

ISSN 0265-8003

Organizing Data Analytics

Ricardo Alonso and Odilon Câmara

Discussion Paper DP16768
Published 02 December 2021
Submitted 25 November 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Organizing Data Analytics

Abstract

We develop a theory of credible skepticism in organizations to explain the main trade-offs in organizing data generation, analysis and reporting. In our designer-agent-principal game, the designer selects the information privately observed by the agent who can misreport it at a cost, while the principal can audit the report. We study three organizational levers: tampering prevention, tampering detection and the allocation of the experimental-design task. We show that motivating informative experimentation while discouraging misreporting are often conflicting organizational goals. To incentivize experimentation, the principal foregoes a flawless tampering detection/prevention system and separates the tasks of experimental design and implementation.

JEL Classification: D8, D83, M10

Keywords: strategic experimentation, Bayesian persuasion, tampering, Organizational design, Information Technology, Audit

Ricardo Alonso - r.alonso@lse.ac.uk
London School of Economics and CEPR

Odilon Câmara - ocamara@marshall.usc.edu
University of Southern California

Acknowledgements
For their suggestions, we thank Raquel Campos, Heski Bar-Isaac, Rahul Deb, Andrea Galeotti, Chad Kendall, Jin Li, Tony Marino, Doehong Min, Anh Nguyen, Luca Picariello, Marco Pagano, Eduardo Perez-Richet, João Ramos, Lucas Siga, Vasiliki Skreta, Wing Suen, Roland Strausz, Denis Shishkin, Teck Yong Tan and Giorgio Zanarone. We thank attendees to several seminars and participants of the 2017 LA Theory Conference, 2018 Informs, 2020 Kent-Bristol-City Workshop, 2020 Seminar in Communication and Persuasion and 2021 SAET.

Organizing Data Analytics*

RICARDO ALONSO†

London School of Economics and CEPR

ODILON CÂMARA‡

University of Southern California

November 24, 2021

Abstract

We develop a theory of credible skepticism in organizations to explain the main trade-offs in organizing data generation, analysis and reporting. In our designer-agent-principal game, the designer selects the information privately observed by the agent who can misreport it at a cost, while the principal can audit the report. We study three organizational levers: tampering prevention, tampering detection and the allocation of the experimental-design task. We show that motivating informative experimentation while discouraging misreporting are often conflicting organizational goals. To incentivize experimentation, the principal foregoes a flawless tampering detection/prevention system and separates the tasks of experimental design and implementation.

JEL classification: D8, D83, M10.

Keywords: Strategic experimentation, Bayesian persuasion, tampering, organizational design, information technology, audit.

*This manuscript incorporates some results presented by the authors under the title "Tampering with Information." For their suggestions, we thank Raquel Campos, Heski Bar-Isaac, Rahul Deb, Andrea Galeotti, Chad Kendall, Jin Li, Tony Marino, Doehong Min, Anh Nguyen, Luca Picariello, Marco Pagano, Eduardo Perez-Richet, João Ramos, Lucas Siga, Vasiliki Skreta, Wing Suen, Roland Strausz, Denis Shishkin, Teck Yong Tan and Giorgio Zanarone. We thank attendees to several seminars and participants of the 2017 LA Theory Conference, 2018 Informs, 2020 Kent-Bristol-City Workshop, 2020 Seminar in Communication and Persuasion and 2021 SAET.

†LSE, Houghton Street, London WC2A 2AE, United Kingdom. R.Alonso@lse.ac.uk

‡USC FBE Dept, 701 Exposition Blvd, Ste. 231, HOH-231, MC-1422, Los Angeles, CA 90089-1422. ocamara@marshall.usc.edu.

1 Introduction

Employees need to recognize that not all numbers are created equal–some are more reliable than others.

Shah, Horne and Capellá, "Good Data Won't Guarantee Good Decisions", HBR (April 2012)

The Digital and ICT revolution has made organizations awash with data by drastically reducing the costs of data gathering, storage, access, and analysis. It has also changed how managers make decisions, relying less on opinions and intuition and more on insights derived from this data.[1] In spite of these improvements, the information that reaches decision makers is still hampered by incentive conflicts: conflicts of interest over decisions result in disagreement over which data to collect and how to analyze it, and creates frictions when communicating its findings to decision makers. In this paper, we develop a theory of the organization of data analytics in the face of these frictions.

To highlight the role of data, consider a CEO (principal) who decides on the scale of a project and relies on the opinion of a manager (agent) who is inclined to overstate the quality of his own projects. At first glance, the recent advancements in data analytics and easier access to data could be a solution to this problem–KMPG (KPMG, 2016) and McKinsey (McKinsey and Co, 2017) highlight the importance of using data to counterbalance biases, and advocate the adoption of a data-driven, test-and-learn culture.[2] However, evidence shows that companies are struggling to capture value from analytics (McKinsey and Co, 2018a). One of the problems is that organizations still need to rely on people to design the experiments, manage the data and report the results, and their preferences might not be

[1]Brynjolfsson, Hitt, and Kim (2011) and Brynjolfsson and McElheran (2016) report rapid and widespread adoption of Data-Driven Decision Making (DDM) practices in organizations, where the rate of adoption is heavily influenced by a series of complementary organizational practices. See also Goldfarb and Tucker (2019) for a discussion on the type of cost reductions brought about by digital economic activity.

[2]Most data analytics is used for process or product improvement or is related to other types of innovation—see Wu, Hitt, and Lou (2020). Earlier work by Pfeffer and Sutton (Pfeffer and Sutton, 2006) promotes what they call evidence-based management: organizations should encourage trial programs, pilot studies, and experimentation.

aligned with the goals of the organization. These frictions in data management manifest themselves, for example, in decision-makers' mistrust of data–a recent survey by KPMG reports that only one third of respondents trust the insights generated from their business operations.[3]

We see data analytics as comprising two tasks: (i) *experimental design* that specifies which data to collect and how to process it, and (ii) *analysis* whereby the actual experiment is implemented according to the specified design. If the principal were to directly specify which experiment to run–or could delegate this role to a trusted data scientist–while the analysis is still performed by a biased manager, then fostering trust in analytics would call for policies that ensure data accuracy and integrity—e.g., through regular examination of data, access management, and audit trails—or that prevent data tampering (or minimize its effect).[4]

The key conflict in our model is that the principal must delegate the experimental design to a biased manager, who can potentially limit how much the organization learns from the data. This scenario is consistent with many relevant applications. For example, if the principal is a high-level executive of a large organization, she may lack the time or the expertise to personally design all aspects of all the experiments.[5] Even if she could hire unbiased data scientists, empirical evidence shows that firms cannot simply delegate experimentation to them—lack of visibility and knowledge of the business unit forces firms to still rely on the participation of business managers when designing experiments.[6] To circumvent this problem, analytics centers typically employ "translators"–employees sourced from the business units with business knowledge–to work with the data scientists (McKinsey

[3]See KPMG (2016, 2018). There are multiple factors behind this mistrust of data: data breaches and inaccuracies–questioning the integrity of data–but also the lack of experience with certain advanced analytics that lead managers to regard them as a "black box" and doubt the value of their results (KPMG, 2018).

[4]The adoption of new technologies such as blockchain can eliminate data tampering within organizations, thus giving decision makers access to information that is known to be correct (Tapscott and Tapscott, 2017).

[5]For large companies implementing a new analytics program, McKinsey (McKinsey and Co, 2017) suggests coming up with as many as 100 possible use cases.

[6] McKinsey (McKinsey and Co, 2017, 2018a) finds that if a firm's analytics team works on an island, isolated from business, then its impact might be very limited. Pilots carried out in small labs with limited connection to the business typically fail to provide the needed answers. Data scientists might lack a deeper understanding of the business. Consequently, McKinsey recommends that the design of analytics solutions needs to have business participation from the start.

and Co., 2018b). Since these translators have a longstanding relationship with business units and their careers depend on the latter's success, we consider them to have the same preferences as the biased manager.

We are interested in understanding optimal organizational practices under delegated experimentation. We first consider a principal that "*organizes to innovate*:" she needs to test a new project to decide whether to scale-up (decision d_H) or toss out the project and retain the status-quo (decision d_S). The agent in our model can be interpreted as a business unit manager who will run the test and report its results. The agent always prefers to scale-up,[7] while the principal prefers to scale-up if the state is "high" and keep the status-quo when the state is "low." We model data analytics as a *designer-agent-principal* game:[8] (i) to persuade the principal, an uninformed designer strategically designs an experiment that reveals information about the payoff-relevant state—as in Kamenica and Gentzkow (2011), KG henceforth. We consider both the case in which the agent designs the experiment (integration) or the design task is allocated to an analyst outside the business unit but with the same preference (separation);[9] (ii) the agent runs the experiment and privately observes its outcome; before submitting his report, he finds how costly it is to misrepresent (tamper) this outcome. This simple setup captures the main experimentation-tampering trade-off: the principal needs to strike a balance between inducing the designer to select a more informative experiment while at the same time restraining the agent from tampering.

To manage this trade-off, we explore several organizational levers: (i) *task allocation*—namely, whether to integrate or separate the role of the designer–and (ii) *data governance*, comprising two types of policies: *tampering detection* policies—how much to audit the agent's report where auditing intensity is given by the probability $\lambda \in [0, 1]$ that the principal also observes the actual experimental outcome, thus both detecting tampering and muting its effect on decision making–and *tampering prevention* policies that raise the costs of tampering.[10]

[7]Business managers may favor scaling-up because of a preference for empire building, or because their human capital is tied to this decision, or as a result of improved outside opportunities.

[8]Alternatively, borrowing from the strategic communication literature, this would be a *designer-sender-receiver* game.

[9]As argued above, effective delegation of the design of experiments requires the involvement of employees with knowledge of the business unit (McKinsey and Co., 2018b).

[10]For instance, through a by-law that defines the punishment for tampering, or through data security

Our first result highlights the role of tampering detection: for a fixed and imperfect prevention policy, increasing auditing intensity reduces the informativeness of the designer's experiment. To wit, managers may gather "just enough" evidence if decision makers find them more reliable. In fact, a perfect audit $\lambda = 1$ deters tampering altogether but results in an experiment from which the principal derives no surplus–as she observes the true experimental outcome, when she decides to scale-up she is indifferent between the two decisions; see KG. With an imperfect audit, the agent tampers more but the designer must select a more informative experiment to compensate for the tampering. Thus, and consistent with our theme of "credible skepticism," the principal commits to an imperfect audit ($\lambda^* < 1$) that optimally trades-off the increased tampering with the increased informativeness.

Our second result shows that she prefers to separate the design task from the analysis task as integrating tasks leads the designer to also incur the tampering costs, thus forcing him to economize on them by shifting to less informative experiments. Both insights resonate with organizations that centralize design in a Center of Excellence (CoE) while coming short of implementing water-tight auditing measures.[11]

Our third result considers the optimal tampering prevention policy given an imperfect audit. We show that the firm makes low tampering costs sufficiently likely as this incentivizes tampering and provides decision makers with commitment power to reject self-serving recommendations.

We then turn to the optimal data governance (i.e., the joint design of tampering prevention and detection policies). We argue that promoting a moderate sense of mistrust can create a culture of "healthy" skepticism in the organization: the principal can refrain from adopting agents' self-serving recommendations issued with weak supporting evidence, forcing the latter to provide stronger evidence backing them. However, to credibly do so, the principal *both* makes low tampering costs sufficiently likely and limits her auditing intensity–tampering prevention and detection act as complements. Under this optimal scheme, *the designer always selects a fully informative experiment.* That is, organizations in our model

measures that make tampering more or less costly and more or less easy to detect. All in all, these policies shape the distribution of the tampering costs faced by the agent.

[11]For instance, McKinsey and Co. reports on several firms centralizing data analytics around a CoE tasked with homogenizing data analytics and supporting the different business units. See McKinsey and Co. (2018b).

would take actions to maximize experimentation while being subject to moderate levels of data misrepresentation. We show that this optimal organization can be implemented through a decoupled internal-external audit system.

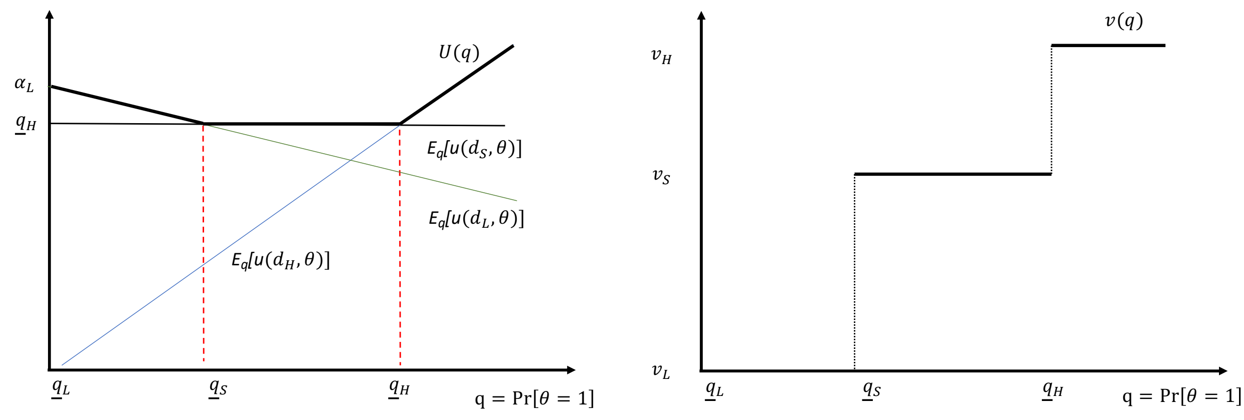
Finally, we look at a principal that "*organizes for scale*"–i.e., she entertains both scaling-up (d_H) and down (d_L) as viable alternatives to the status-quo (d_S). It is still true that separating tasks and committing to an imperfect audit leads to stronger evidence backing a d_H recommendation, but the principal must guard against "adverse switches"–i.e., a switch from an experiment recommending d_H or d_L to a less informative "status-quo" experiment recommending d_H or d_S– and it may now be optimal for her to integrate tasks and to perfectly audit the agent's report (set $\lambda^* = 1$). Adverse switches occur because agents promoting d_H are more willing to compromise on the status quo than the principal. She can avoid such compromise by ruling out the status-quo as an option; in fact, if the principal can commit to ruling-out decisions (i.e., to reduce her discretion), then she prefers an imperfect audit ($\lambda^* < 1$). Thus, in this context, discretion and auditing intensity act as complements.

We present the model in Section 2. Section 3 characterizes the equilibrium in the communication subgame for a fixed organizational structure. Sections 4 and 5 cover our main insights on the optimal organization of data analytics for the case in which the principal organizes to innovate. Section 6 studies the case in which the principal organizes for scale. We conclude with a discussion of the related literature in Section 7. All proofs are in the Appendices.

2 Model

To model the different tasks of data analytics, we introduce a "designer-agent-principal" game in which the data designer (he) specifies what information the agent (he) will privately observe and report to the principal (she) prior to making the decision.

Preferences and Prior Beliefs: Players are expected utility maximizers. The state space is binary, with typical realization $\theta \in \Theta = \{0, 1\}$, and players hold a common prior $\mu = \Pr[\theta =$



(a) Principal's indirect utility

(b) Agents' indirect utility

Figure 1: Principal and Agents' payoffs

1]. The principal selects d from $\{d_L, d_S, d_H\}$, and has preferences characterized by $u(d,\theta)$, [12]

$$u(d,\theta) = \begin{cases} \underline{q}_H & \text{for } d = d_S, \\ \theta & \text{for } d = d_H, \\ \alpha_L & \text{for } d = d_L \text{ and } \theta = 0, \\ \alpha_L - \mathbb{I}_{\left\{\alpha_L \geq \underline{q}_H\right\}} \frac{\alpha_L - \underline{q}_H}{\underline{q}_S} & \text{for } d = d_L \text{ and } \theta = 1, \end{cases}$$

with $0 = \underline{q}_L \leq \underline{q}_S < \underline{q}_H < 1$. In words, the principal decides either to keep the status quo d_S, to scale-up operations by choosing d_H, or to scale-down by choosing d_L–Figure 1 represents the principal's payoff as a function of her posterior q. Our main focus will be on the case that $\alpha_L < \underline{q}_H$ so that the principal "*organizes to innovate*" as she effectively chooses between d_S or d_H (i.e., whether to approve the "innovation" d_H), selecting d_H only if $q \in \left[\underline{q}_H, 1\right]$. Second, if $\alpha_L \geq \underline{q}_H$, then the principal "*organizes for scale*:" if she deviates from d_S, then she could either scale-up (d_H), or scale-down (d_L), selecting d_L only if $q \in \left[\underline{q}_L, \underline{q}_S\right]$. In either case, $\underline{q}_i$ represents the minimum posterior belief for which the principal still selects d_i.

We capture the conflict of interest between the agents and the principal by positing that the designer and the agent receive a state-independent payoff $v(d_i, \theta) = v_i$ with $0 = v_L <$

[12] $\mathbb{I}_A$ represents the indicator function of the set A; i.e., $\mathbb{I}_A(x) = 1$ if $x \in A$ and $\mathbb{I}_A(x) = 0$ if $x \notin A$.

$v_S < v_H$, so that they benefit from persuading the principal to choose "higher" decisions. To focus on the more interesting case, we assume that $\mu \in (\underline{q}_S, \underline{q}_H)$ so that the principal retains the status quo in the absence of any report, and we let $\Delta_i \equiv v_H - v_i$, $i \in \{L, S\}$ be the designer/agent's gain from inducing his preferred decision d_H when the alternative is d_i.

Strategic Experimentation, Communication and Tampering: All players process information according to Bayes' rule. We consider three stages to this model of data analytics.

First, in the experimental-design stage, the designer specifies which data to gather and how it will be processed: he selects an experiment π, consisting of a finite outcome space $S(\pi)$ and a family of likelihood functions over $S(\pi)$, $\{\pi(\cdot|\theta)\}_{\theta \in \Theta}$, with $\pi(\cdot|\theta) \in \Delta(S(\pi))$. Given the common prior, we can without loss set $S(\pi) \subset \Delta(\Theta)$, so that $\pi = \{q, \Pr[q]\}_{q \in S(\pi)}$ is expressed as a distribution over posterior beliefs q induced by observing the experimental outcome, with $S(\pi)$ indexing these outcomes. We say that the designer "experiments more" when he selects a Blackwell-more informative experiment. We make two important assumptions regarding experimental design. First, as in KG, the designer can choose *any experiment* that is correlated with the state. Second, experiments are costless to the designer. This can be the case, for instance, if a fully informative experiment is originally available to the designer and he can garble its outcome at no cost.

Second, the design stage is followed by an analysis/communication stage. The agent privately observes the outcome $s \in S(\pi)$–we refer to s as the agent's "type"– and sends a message $m \in S(\pi)$ to the principal, which is potentially subject to misrepresentation: the agent can tamper with the true outcome s by reporting instead $s' \in S(\pi), s' \neq s$. We will work with a reduced-form model of tampering: the agent incurs a cost c if he tampers, with c unknown at the design stage and distributed according to $F(c)$, and independent of the experiment π. These costs are shaped by the principal's tampering prevention policies and can be physical costs–e.g., effort in "doctoring the books" or "creating a credible alternative story"–or represent punishments if caught misrepresenting–with the severity of the punishment varying with the tampering method–or even psychic costs of misrepresentation.[13] We let $\bar{F}(c) \equiv 1 - F(c)$ and $f \equiv dF/dc$ be its density, whenever it exists.

We make two assumptions regarding these tampering costs. First, they are always borne

[13] Gneezy (2005) and Abeler, Nosenzo and Raymond (2020) show experimentally that individuals have some innate preference for honesty.

by the agent when he tampers. In Section 5.5, we show that the main insights of our analysis hold if the agent incurs the cost c only if tampering is uncovered through auditing. Second, the agent bears the same cost independently of the actual message sent. In other words, the decision of how to misrepresent the state only depends on the equilibrium inference of the principal, rather than the costs/punishments specifically associated to different messages.

In the third stage, the decision making stage, the principal observes both the designer's experiment and the agent's message. Key in our model is the principal's ability to evaluate the truthfulness of this message, and undo the effect of any misrepresentation, by auditing the experiment. We assume that with probability λ the audit is conclusive and the principal learns the actual experimental outcome, while with probability $1 - \lambda$ the audit is inconclusive and she gains no new information. Importantly, what can be learned from an audit is constrained by the informativeness of π. Thus, auditing differs from seeking a "second opinion" in which the principal may have access to a separate information source.[14] If the audit is conclusive, the principal is informed (of s) and selects (a possibly mixed) $d_I(m, s)$ which depends on the message m and the outcome s. If the audit is inconclusive, she is uninformed and selects $d_U(m)$. To lighten the exposition, we say that "the message/recommendation is (un)audited" when the audit is (in)conclusive.

Organizational Design: Organizational members perform two tasks—experimental design and analysis—and the principal has several organizational levers to incentivize them. First, she defines the firm's data governance comprising the tampering prevention and detection policies. In terms of tampering detection, she sets the auditing intensity $\lambda \in [0, 1]$. For instance, she can assign resources at the outset that are used later to audit the agent's report, thus, dictating the likelihood of a conclusive audit. An important assumption is that the principal can commit to an imperfect audit, i.e., to $\lambda < 1$. Otherwise, once the designer selects an experiment, she can completely eliminate the effect of tampering by perfectly auditing the agent's message. In anticipation of a perfect audit, however, the designer would select the perfect commitment experiment as in KG. To see this, and for future reference, consider experiments $\{\underline{q}_L, \underline{q}_H\}$ and $\{\underline{q}_S, \underline{q}_H\}$ with

[14]See, for instance, Kolotilin (2018), Kolotilin, Mylovanov, Zapechelnyuk, and Li (2017), and Guo and Shmaya (2019) for information-design models where the receiver is privately informed.

$$p_i^C \equiv \Pr[s = \underline{q}_i] = \frac{q_H - \mu}{\underline{q}_H - \underline{q}_i}, \tag{1}$$

the probability of outcome $s = \underline{q}_i$, $i = L, S$. If $\lambda = 1$, then the designer selects the experiment $\{\underline{q}_i, \underline{q}_H\}$ that minimizes $p_i^C \Delta_i$. If the organization hopes to induce more experimentation, it must be able to guarantee that the success rate of an audit is limited to λ. Our interpretation is that further resources cannot be deployed once the auditing intensity is announced, so that λ cannot be increased neither in reaction to the chosen experiment, nor to the reported outcome. In terms of tampering prevention, she can enact different data encryption and authentication systems to preserve data integrity or security measures to make access to data storage systems costly. All these policies effectively shape the distribution of tampering costs F.

Second, the principal can choose to either integrate design and analysis, by letting the same agent perform both tasks, or to separate them, by allocating each to a different agent.[15] Let k denote the principal's task allocation, with $k \in (\mathcal{I}, \mathcal{S})$. Instead of changing the number of agents for each task allocation, we keep our designer-agent-principal game throughout all task allocations and assume that the designer also bears the tampering costs incurred by the agent under integration ($k = \mathcal{I}$), while he does not bear them under separation ($k = \mathcal{S}$). In terms of organizational structure, task separation would correspond to a firm in which experimental design is centralized in a corporate headquarters and the designer mandates each operating unit which analysis to perform, while the actual data collection and reporting is decentralized to those units.[16]

Timing: The principal selects whether to separate or integrate tasks and data governance

[15] A maintained assumption of our analysis is that task allocation does not affect the agents preferences over decisions. That is, task allocation cannot be used to reduce the conflict of interest between principal and agents.

[16] For instance, the design of customer surveys or the specification of which data to be collected by Enterprise Resource Planning (ERP) systems could be performed by an enterprise-wide data architect, while the analysis of the results is performed at the divisional level. Integration would have both tasks been decentralized to lower level units, so that local agents have discretion in deciding which data to collect and which analysis to perform. As an example in the public sector, David Cameron created the Behavioral Insights Team (BIT) under the supervision of the Cabinet Office (see Alonso and Câmara, 2016 for details). In an example of task-integration, the BIT would both design and conduct small-scale experiments for the UK Government.

(λ, F). Then, the designer publicly selects $\pi = \{q, \Pr[q]\}_{q \in S(\pi)}$—this is the design subgame. The communication subgame follows: Nature draws θ and the agent privately observes outcome $s \in S(\pi)$, generated according to π, and the cost c, and selects a message $m \in S(\pi)$. The principal observes the actual outcome s with probability λ and, given the outcome of the audit and the agent's message, she updates her beliefs according to Bayes' rule, selects a decision, payoffs are realized and the game ends. We look for Perfect Bayesian Equilibria that constitute a Perfect Bayesian Equilibrium in every subgame.

3 Communication Equilibria

We start the organizational-design analysis by studying how agents respond to a given organizational structure. That is, we study the equilibria in the *designer-agent-principal* game corresponding to a fixed task allocation, cost distribution and auditing intensity. We work backwards by first characterizing the equilibria in the communication subgame for any π, which will determine both expected tampering and the distribution over the principal's decisions.

3.1 Equilibrium Tampering

The agent decides whether to tamper by comparing the gain from misrepresenting his type to the realized tampering cost. Let $\bar{\Delta}$ be the maximum gain from tampering in any equilibrium (both on- and off- the equilibrium path).[17] Throughout the paper, we consider pairs (λ, F) that guarantee the existence of tampering costs that make tampering unprofitable for every

[17] If the audit is inconclusive, the largest gain from tampering comes from inducing the principal to select d_H when truthful communication would have led to his least preferred decision. Tampering incentives are also shaped by the principal's decision after a conclusive audit: if the agent's type is a "threshold" type $\underline{q}_i$, $i = \{S, H\}$, then the principal may punish/reward him after a conclusive audit by randomizing differently between decisions as a function of his report. Then, the maximum gain from tampering is $\bar{\Delta} = (1-\lambda)\Delta_S$ if the principal organizes to innovate, but $\bar{\Delta} = \Delta_L - \lambda\Delta_S$ if she organizes for scale. To see this last case, consider the experiment $\left\{\underline{q}_S, q\right\}$ with $q > \underline{q}_H$, and the following sequentially-rational decision making: after a conclusive audit, the principal selects d_H if $s = q$ but if $s = \underline{q}_S$ she selects d_L if the agent truthfully reported $m = \underline{q}_S$ and d_S if he tampered $m = q$. If the audit is inconclusive, and the only tampered outcome is q, the principal's consistent belief after an unaudited $m = \underline{q}_S$ must be precisely $\underline{q}_S$, in which case she selects d_L. Then, the gain from tampering after $s = \underline{q}$ is $\lambda v_S + (1-\lambda)v_H - v_L = \Delta_L - \lambda\Delta_S$.

agent's type, thus ensuring a positive probability of truthful reporting for every experiment and experimental outcome. This limits the scope of the principal to discipline the agent's tampering by holding "optimistic or pessimistic" beliefs after an off-the-equilibrium-path message.

Assumption 1 (All messages on-path) *The tampering cost distribution satisfies*

$$\bar{F}(\bar{\Delta}) > 0. \tag{2}$$

In the absence of threshold types–i.e., whenever $\underline{q}_i \notin S(\pi), i = \{S, H\}$–the agent that tampers sends a message that induces the "highest" decision after an inconclusive audit. Thus, if different types choose to tamper by sending different messages, it must be that they all induce the same unaudited decision (or mixtures over decisions). With this observation, we now characterize tampering behavior in the communication subgame.

Proposition 1. *In any equilibrium of the communication subgame following the choice* $\pi = \{q, \Pr[q]\}_{q \in S(\pi)}$ *we have:*

(i) For each $q \in S(\pi)$*, there exists* $\bar{c}(q)$*, with* $\bar{F}(\bar{c}(q)) > 0$*, such that* $m^*(q, c) = q$ *if* $c > \bar{c}(q)$ *and* $m^*(q, c) \neq q$ *if* $c < \bar{c}(q)$*;*

(ii) Let $M_T(\pi) \subset S(\pi)$ *be the set of "tampered outcomes:"*

$$M_T(\pi) = \{q \in S(\pi) : \exists (q_z, c),\ m^*(q_z, c) = q, q_z \neq q\}.$$

If $d_I(m, \underline{q}_i)$ *is independent of* m *whenever* $\underline{q}_i \in S(\pi), i = \{S, H\}$*, then for* $q, q' \in M_T(\pi)$ *(a)* $d_U(q) = d_U(q')$*, and (b)* $\bar{c}(q) = 0$*.*

Proposition 1-i shows that the agent's tampering behavior is monotonic: he reports truthfully if the realized cost exceeds an outcome-dependent threshold, $\bar{c}(q)$, and will surely tamper if the cost falls below this threshold. Proposition 1-ii(a) makes formal the above-mentioned property that "tampered outcomes"–messages that are transmitted by some other type with positive probability—may induce different posterior beliefs but must all lead to the same unaudited mixture over decisions; this is true as long as the principal does not condition his audited decision on the agent's message. Additionally, there shouldn't be any gain from tampering for a "tampered-outcome" type; that is, types that others would like to mimic always report truthfully. This is Proposition 1-ii(b).

4 Organizing to Innovate: Equilibrium Experimentation

Suppose that the firm "organizes to innovate"–i.e., the principal decides whether to "approve" d_H or retain d_S. To characterize the designer's equilibrium choice, we introduce the set of "status-quo" experiments Π_S: binary experiments of the form $\{0, q\}$ such that, in equilibrium, an inconclusive audit leads the principal to posterior $\underline{q}_H$ following $m = q$.

Definition (Status-quo Experiments) *Define the class of status-quo experiments* Π_S *indexed by* $\tau \in [0,1]$, *as*

$$\Pi_S = \left\{ \pi = \{0, q\} : q \geq \underline{q}_H \text{ and } \bar{F}\left(\bar{c}_S(\tau)\right) = p_S^C \left(\frac{q}{q-\mu} \right), \tau \in [0,1] \right\}, \tag{3}$$

with

$$\bar{c}_S(\tau) = \tau \left(1 - \lambda\right) \Delta_S, \tag{4}$$

alongside the equilibrium behavior

$$d_U\left(0\right) = d_I\left(m, 0\right) = d_S; d_I\left(m, q\right) = d_H; d_U\left(q\right) = \tau d_H + (1-\tau) d_S.$$

In equilibrium, the principal selects d_H with probability τ and d_S with probability $1 - \tau$ following $m = q$ and an inconclusive audit. This also determines the incentives to tamper after an unfavorable outcome—which given scale-up probability τ translate to threshold (4)—and the probability $\bar{F}\left(\bar{c}_S(\tau)\right)$ that following $s = 0$ the agent reports truthfully−see (3). This also implies that the set Π_S depends on the auditing intensity λ; for instance, if $\lambda = 1$ then the commitment experiment is the only status-quo experiment. Moreover, experiments in Π_S can be equivalently indexed by either (i) the posterior q, (ii) the probability of outcome $s = 0$, with $\Pr\left[s = 0\right] = \frac{q-\mu}{q}$, (iii) the scale-up probability τ, or (iv) the induced tampering threshold $\bar{c}_S$. Finally, all experiments in Π_S are ordered according to their informativeness: trivially, an experiment with a higher q (equivalently higher $\bar{c}_S$, higher τ, or higher $\Pr[s = 0]$) corresponds to a Blackwell-more informative experiment.

We now present our main equilibrium characterization. Given F, let $v_S(\tau, \mu; \lambda, k)$ be the designer's equilibrium payoff in a communication subgame after he selects $\pi_S\left(\tau\right) \in \Pi_S$, with $k \in \{\mathcal{S}, \mathcal{I}\}$ the principal's task allocation and λ her auditing intensity, and let

$$V_S\left(\mu; \lambda, k\right) \equiv \max_{\tau \in [0,1]} v_S(\tau, \mu; \lambda, k), \tag{5}$$

be his maximum expected utility from a status-quo experiment.

Proposition 2. *Let* $\lambda > 0$ *and* $\mu \in (0, \underline{q}_H)$*. Then,*
(i) there is always an equilibrium of the design subgame in which the designer selects a status-quo experiment,
(ii) if the designer obtains payoff V^* *in some equilibrium of the design subgame, then* $V^* = V_S(\mu; \lambda, k)$*.*

Proposition 2 justifies our restriction to status-quo experiments when analyzing the principal's organizational-design problem. This is based on two observations. First, there is always an equilibrium in which the designer responds by selecting an experiment in Π_S. We prove this claim in the appendix by constructing from an arbitrary $\pi' = \{q, \Pr[q]\}_{q \in S(\pi')}$ a status-quo experiment that gives the designer a (weakly) higher payoff. Second, all equilibria of the design subgame give the same expected payoff; thus, to find the designer's optimal payoff we can restrict attention to status-quo experiments.

For the remainder, let $\{0, q_S^*(\lambda, k)\} \in \Pi_S$ be the designer's optimal experiment, with $\bar{c}_S^*(\lambda, k)$ the induced tampering threshold.

4.1 Designer's Equilibrium Payoffs

To solve for the designer's optimal experiment using Proposition 2, we now characterize $v_S(\tau(\bar{c}), \mu; \lambda, k)$–the designer's payoff as a function of the induced tampering threshold. To this end, define $\eta(\bar{c})$ as the product of the expected tampering cost conditional on tampering times the odds of tampering,

$$\eta(\bar{c}) \equiv E[c | c \leq \bar{c}] \frac{F(\bar{c})}{\overline{F}(\bar{c})} = \frac{\int_0^{\bar{c}} c dF(c)}{\overline{F}(\bar{c})}. \tag{6}$$

Lemma 1. *For* $\mu \in \left[0, \underline{q}_H\right]$*, consider* $\{0, q\} \in \Pi_S$ *that induces threshold* $\bar{c}$*. Then,*

$$v_S(\tau(\bar{c}), \mu; \lambda, k) = v_H - (1-\lambda)\Delta_S - m_S(\bar{c}; \lambda, k)\left(\underline{q}_H - \mu\right), \tag{7}$$

with m_S *the slope of the designer's payoff with respect to the prior* μ*:*

$$m_S(\bar{c}; \lambda, k) \equiv \frac{1}{\underline{q}_H}\left(\frac{\lambda \Delta_S}{\overline{F}(\bar{c})} + \bar{c} + \mathbb{I}_{\{k=\mathcal{I}\}}\eta(\bar{c})\right). \tag{8}$$

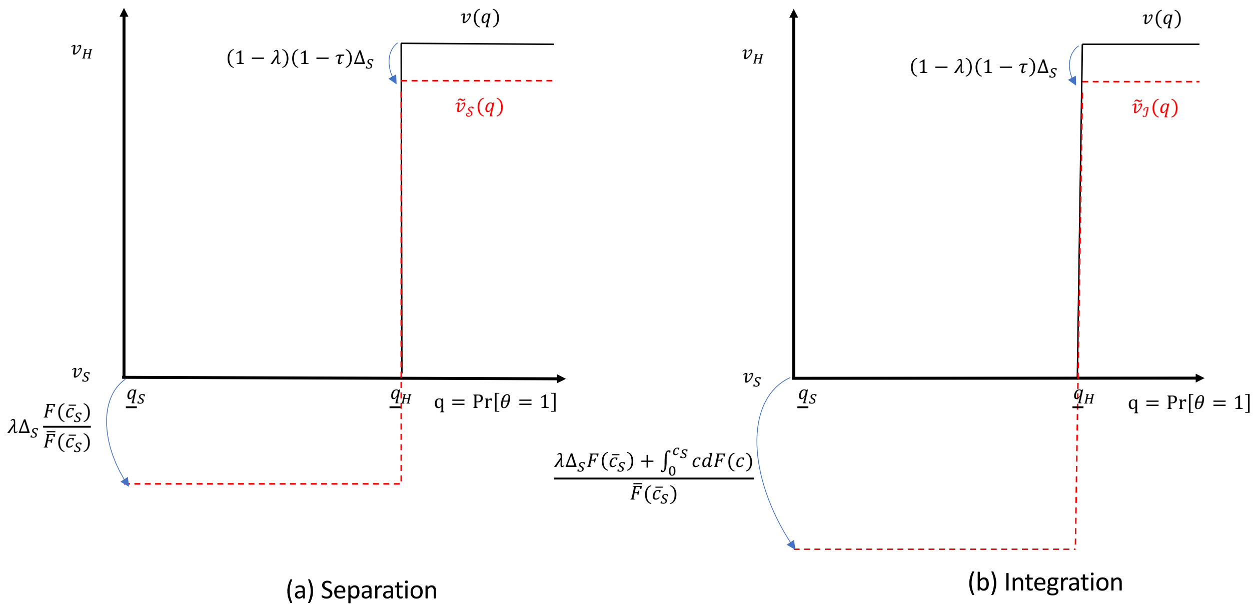


Figure 2: Payoffs for the equivalent "full commitment" game.

To understand (7), for each $\{0, q\} \in \Pi_S$ we introduce an equivalent "full commitment" game with a "modified" designer's utility so that the designer's payoff from $\{0, \underline{q}_H\}$ gives him the same expected payoff–see Figure 2.[18] This allows us to appeal to the geometric intuition of the full commitment case when studying the designer's preferences in the imperfect commitment case. Fix $\pi(\bar{c}) \in \Pi_S$ and define the task allocation-dependent utility $\tilde{v}_{i,k}$ for d_i,

$$\tilde{v}_{H,k} \equiv v_H - (1-\lambda)(1-\tau(\bar{c}))\Delta_S, \tag{9}$$

$$\tilde{v}_{S,k} \equiv v_S - \lambda\Delta_S \frac{F(\bar{c})}{\bar{F}(\bar{c})} - \mathbb{I}_{\{k=\mathcal{I}\}}\eta(\bar{c}). \tag{10}$$

Figure 2 describes the relation between $v(q)$ in the original game and the indirect utility $\tilde{v}_{i,k}(q)$ in the equivalent "full commitment" game.

To show payoff-equivalence, start with the designer's payoff under full commitment from $\{0, \underline{q}_H\}$, $\left(1 - p_S^C\right) v_H + p_S^C v_S$, with p_S^C given by (1). Equilibrium tampering alters this payoff in two ways. First, upon observing $\underline{q}_H$, the principal now keeps the status quo with proba-

[18] We cannot apply the concavification argument when $\lambda < 1$ because the probability of a message corresponding to a "tampered outcome," as well as the principal's posterior belief when unaudited, are determined in equilibrium by the agent's global tampering behavior, which in turn depends on the entire distribution of experimental outcomes.

bility $(1-\lambda)(1-\tau(\bar{c}))$ −this explains (9). Second, tampering and auditing change the distribution of outcomes so that the probability of observing 0 increases from p_S^C to $p_S^C/\bar{F}(\bar{c})$—see (3). This higher probability of a low outcome reduces the designer's payoff by

$$\lambda\left[\left(1-p_S^C\right)v_H+p_S^C v_S\right]-\lambda\left[\left(1-\frac{p_S^C}{\bar{F}(\bar{c})}\right)v_H+\frac{p_S^C}{\bar{F}(\bar{c})}v_S\right]=\lambda\Delta_S\frac{F(\bar{c})}{\bar{F}(\bar{c})},$$

which explains (10) for $k=\mathcal{S}$. Finally, tampering costs are incurred if $s=0$, which occurs with probability $p_S^C/\bar{F}(\bar{c})$, and explains (10) for $k=\mathcal{I}$.

Expressions (9-10) capture the trade-off that the designer faces: a higher scale-up probability τ increases the payoff after an unaudited message q (outcome $\underline{q}_H$ in the equivalent "full commitment" game), but a higher scale-up probability can only result from a higher tampering threshold. This forces the designer to offer a more informative experiment to sustain the higher τ, thus increasing the likelihood of observing an unfavorable outcome.

5 Organizing to Innovate: Organizational Design

We now turn to the issue of organizational design. To understand the trade-offs that the principal faces, consider her expected utility from experiment $\pi=\{0,q\}\in\Pi_S$:

$$U(\lambda,k;\pi) \quad = \quad \underline{q}_H+\Pr[s=q]\,\lambda(q-\underline{q}_H). \tag{11}$$

The principal benefits from π only after auditing a scale-up recommendation so that the more convincing the evidence in favor of d_H (the larger $q-\underline{q}_H$ is) the greater her gain. Given equilibrium behavior, her expected utility increases with both the auditing intensity and with the odds of tampering,

$$U(\lambda,k)=\underline{q}_H+\left(\underline{q}_H-\mu\right)\lambda\frac{F(\bar{c}^*)}{\bar{F}(\bar{c}^*)}. \tag{12}$$

This expression showcases our main insight: fostering experimentation while discouraging tampering are conflicting goal. The principal can always eliminate frictions in communication by perfectly auditing the experiment or making tampering sufficiently costly—for a fixed experiment and costless auditing, she will certainly do so—but this will reduce the information she receives regarding decision d_H. In fact, a fallible data governance allows her to credibly withhold scaling-up if the evidence in favor of d_H are not convincing, forcing the designer to provide more compelling evidence. Thus, she would like to incentivize tampering by having a "shadow of a doubt" on the claims of the agent, but such skepticism can only be credible if $\lambda < 1$.

5.1 Optimal Task Allocation

Expression (12) clarifies that, given equilibrium behavior, the principal is always willing to trade-off more distortions in communication for more informative experimentation. How do the different organizational levers help her motivate experimentation? We first show that, fixing a data governance policy, the principal weakly prefers to separate tasks as this leads the designer to choosing a more informative experiment.

Proposition 3. *We have* $q_S^*(\lambda, \mathcal{S}) \geq q_S^*(\lambda, \mathcal{I})$.

The intuition is straightforward: assigning the tasks of experimental design and analysis to the same agent forces him to economize on tampering costs when choosing an experiment; when organizing to innovate, reducing tampering costs can only be achieved through a reduction in scale-up probability, leading to a less informative experiment.[19]

We now consider the optimal choice of data governance. We discuss the optimal choice of tampering detection and prevention separately before considering the optimal joint design.

5.2 Optimal Tampering Detection

Consider a fix tampering prevention policy that leads to a cost distribution F. The standard rationale for auditing data analytics is both to ensure data integrity and to dissuade tampering. This remains true in our model and implies that, for a *fixed* experiment, increasing λ can only increase the information that reaches the principal.

However, once experimental design is delegated, varying λ also changes the designer's incentives to experiment. Indeed, reducing λ both: (i) changes the set of status-quo experiments, allowing for more informative experiments; and (ii) (weakly) reduces the principal's scale-up probability for each experiment. In fact, moving away from a perfect audit always leads the designer to select a Blackwell-more informative experiment.

[19] Decreasing scale-up probability lowers both the equilibrium tampering threshold and the likelihood that a tampering outcome occurs, leading to lower expected tampering costs.

Proposition 4. *Fix $k \in \{\mathcal{S}, \mathcal{I}\}$. Then,*

(i) $v_S(\tau(\bar{c}), \mu; \lambda, k)$ is supermodular in $(\bar{c}, -\lambda)$.

(ii) Outcome $q_S^*(\lambda, k)$ is non-increasing in λ.

In words, increasing λ : (a) decreases the designer's incremental payoff from an experiment with a higher tampering threshold; and (b) decreases the informativeness of its experiment. To see this, consider the marginal change in the designer's payoff (7) when, as a result of intensifying auditing, scale-up probability increases to preserve the same tampering threshold $\bar{c}$. The designer's marginal expected payoff conditional on an inconclusive audit decreases by v_S, but increases after a conclusive audit by $\Delta_S \Pr[s = q] + v_S$. Overall, the designer's marginal payoff is proportional to the probability that π generates a favorable outcome, which decreases with $\bar{c}$. Therefore, increasing λ reduces the designer's marginal payoff from more informative experiments. Moreover, a more intense auditing also reduces the set of status-quo experiments. Proposition 4-ii then shows that the combined effect of a higher λ unambiguously discourages experimentation.

5.2.1 The Optimality of Lax Auditing

To derive properties of the optimal audit, we introduce the notion of designer's responsiveness to auditing.

Definition (Responsiveness) *The designer under a $k-$allocation is responsive to $\lambda-$auditing if he strictly prefers experiment $\{0, q\}$ to $\{0, \underline{q}_H\}$ for some $q > \underline{q}_H$.*

From (12), the principal gains from experimentation only if lax auditing compels the designer to select an experiment that will be tampered with positive probability. We thus reach one of our main results: if auditing is costless and the designer is responsive to auditing, then in every equilibrium the principal commits to an imperfect audit, i.e. $\lambda^* < 1$.

Proposition 5. *Suppose that the principal can select λ at no cost, and let λ^* denote her equilibrium choice. Then, $\lambda^* < 1$ in every equilibrium if and only if the designer is responsive to auditing. In particular, if she separates tasks and $f(0) > 0$, then $\lambda^* < 1$.*

Note that the conditions for $\lambda^* < 1$ are not too stringent: as long as there is a positive probability that tampering is costless, the optimal audit should be imperfect.

5.2.2 Optimal Auditing

How much should the principal audit the agent's report given that she separates tasks? To derive her optimal audit λ^*, we first characterize the designer's equilibrium experiment for any $\lambda \in (0,1)$. To this end, define $L(c) \equiv f(c)/\left(\bar{F}(c)\right)^2$ and $\bar{c}_{FI}$ implicitly by $\bar{F}\left(\bar{c}_{FI}\right) = p_S^C/\left(1-\mu\right)$, so that $\bar{c}_{FI}$ is the threshold induced by a fully informative experiment.[20]

Lemma 2. *Fix $\lambda \in (0,1)$ and suppose that $L(c) - (\phi_S/\lambda)$ is single-crossing in $[0, \Delta_S]$, from negative to positive, with $\phi_S \equiv \left(1-p_S^C\right)/\left(p_S^C \Delta_S\right)$. Then, the designer under separation selects the commitment experiment, corresponding to $\bar{c}^* = 0$, if $L(0) \geq \phi_S/\lambda$. Otherwise, he selects an experiment that induces tampering threshold*

$$\bar{c}^* = \min\left[L^{-1}(\phi_S/\lambda), (1-\lambda)\,\Delta_S, \bar{c}_{FI}\right]. \tag{13}$$

Consistent with Proposition 4-ii, the designer's optimal experiment induces less tampering, but is less informative, as auditing intensifies–the equilibrium tampering threshold (13) decreases with λ. The single-crossing condition on $L(c)$ guarantees that the designer's expected utility is quasiconcave in the tampering threshold and is always satisfied, for instance, if the hazard rate $f(c)/\bar{F}(c)$ is increasing. The equilibrium threshold $\bar{c}^*$ is the minimum of three possible choices. The term $\bar{c}_{FI} = \bar{F}^{-1}\left(p_S^C/\left(1-\mu\right)\right)$ corresponds to a fully informative experiment, while $(1-\lambda)\,\Delta_S$ corresponds to the case that the principal rubberstamps the agent's recommendation. The first term in (13) reflects the designer's choice when it leads to a lower approval probability. In fact, if $L(c)$ is large—in particular, $L(0) \geq \phi_S/\lambda$—then the principal only approves when she audits and the designer's experiment induces $\bar{c}^* = 0$. Therefore, imperfect, albeit intense, auditing—specifically, when $\lambda \geq \phi_S/f(0)$—can still completely crowd-out valuable experimentation. This imposes an upper bound on the range of auditing intensities that the principal might entertain.

From (12), the principal's problem for a fixed prevention policy inducing F is

$$\lambda^* \in \arg\max_{\lambda\in[0,1]} \lambda \frac{F\left(\bar{c}^*\right)}{\bar{F}\left(\bar{c}^*\right)}, \text{ with } \bar{c}^* \text{ given by (13).} \tag{14}$$

The optimal auditing will, in general, be sensitive to F and the preferences of agents. To illustrate (14), we study a case where F leads to tampering costs that are uniformly distributed.

[20] Recall that $\underline{q}_S = 0$ when organizing to innovate.

Example: Uniform Distribution. Let $\Delta_S = 1$ with $v_S = 0$, and suppose that c is uniformly distributed in $[0, 1]$, so that $\bar{F}[(1-\lambda)\Delta_S] = \lambda$. From (7), the designer's utility under separation when the experiment induces $\bar{c} \in [0, 1-\lambda]$ is

$$v_S(\tau(\bar{c}), \mu; \lambda, S) = \lambda + \bar{c} - m_S(\bar{c}; \lambda, S)\left(\underline{q}_H - \mu\right) = \lambda + \left(1 - p_S^C\right)\bar{c} - p_S^C \frac{\lambda}{1-\bar{c}},$$

which is concave in $\bar{c}$. Denote by $\bar{c}_{crit} \equiv 1 - \sqrt{\lambda \frac{p_S^C}{1-p_S^C}}$ its unconstrained maximum. Then, mirroring (13), the designer's optimal experiment leads to a tampering threshold

$$\bar{c}^* = \min\left\{\max\{0, \bar{c}_{crit}\}, 1-\lambda, \frac{\mu}{1-\mu}\frac{1-\underline{q}_H}{\underline{q}_H}\right\}.$$

We can now fully characterize the equilibrium experiment as a function of λ. Recall that $\frac{1-p_S^C}{p_S^C}$ are the approval odds of the innovation if the principal were to set $\lambda = 1$. If $\lambda \geq \frac{1-p_S^C}{p_S^C}$, then $\bar{c}_{crit} \leq 0$ and the designer selects $\left\{0, \underline{q}_H\right\}$, i.e., selects the commitment experiment. If $\lambda \leq \frac{p_S^C}{1-p_S^C}$, then $\bar{c}_{crit} \geq 1-\lambda$ and the designer selects the most informative status-quo experiment. This would lead to a fully informative experiment or to an experiment for which the principal's rubberstamps an unaudited scale-up recommendation. Finally, if $\frac{p_S^C}{1-p_S^C} \leq \lambda \leq \frac{1-p_S^C}{p_S^C}$, then $\bar{c}^* = \min\left[\bar{c}_{crit}, \frac{\mu}{1-\mu}\frac{1-\underline{q}_H}{\underline{q}_H}\right]$ and the designer limits the informativeness of the experiment, leading to intermediate approval probabilities after an inconclusive audit.

Figure 3 describes two cases, with $\frac{1-p_S^C}{p_S^C}$ taking values 2 and 1/2.[21] If $\frac{1-p_S^C}{p_S^C} = 2$, then the innovation idea is a *good prospect*: it is likely to be perceived after experimentation as a profitable alternative to the current status quo. Then, the designer reacts to more intense auditing by switching to experiments that are less informative (consistent with Proposition 4) but that lead to a higher probability of approval. Figure 3-a shows the principal's utility, which is maximized for $\lambda = 0.57$. So, for good-prospect ideas, the principal engages in somewhat intense auditing and the designer restricts experimentation as, for such intense auditing, the principal is willing to rubberstamp the agent's recommendations.

If $\frac{1-p_S^C}{p_S^C} = 1/2$, then the innovation idea is a *poor prospect*: it is unlikely that experimentation will uncover evidence showing it to be more profitable than the status quo. Again, the designer reacts to more intense auditing by experimenting less but approval probability is now non-monotonic: it increases for low values of λ—as the designer always selects a

[21] In both cases, we take $\mu = 1/4$. We have $\underline{q}_H = 3/8$ if $\frac{1-p_S^C}{p_S^C} = 2$, while $\underline{q}_H = 3/4$ if $\frac{1-p_S^C}{p_S^C} = 1/2$,

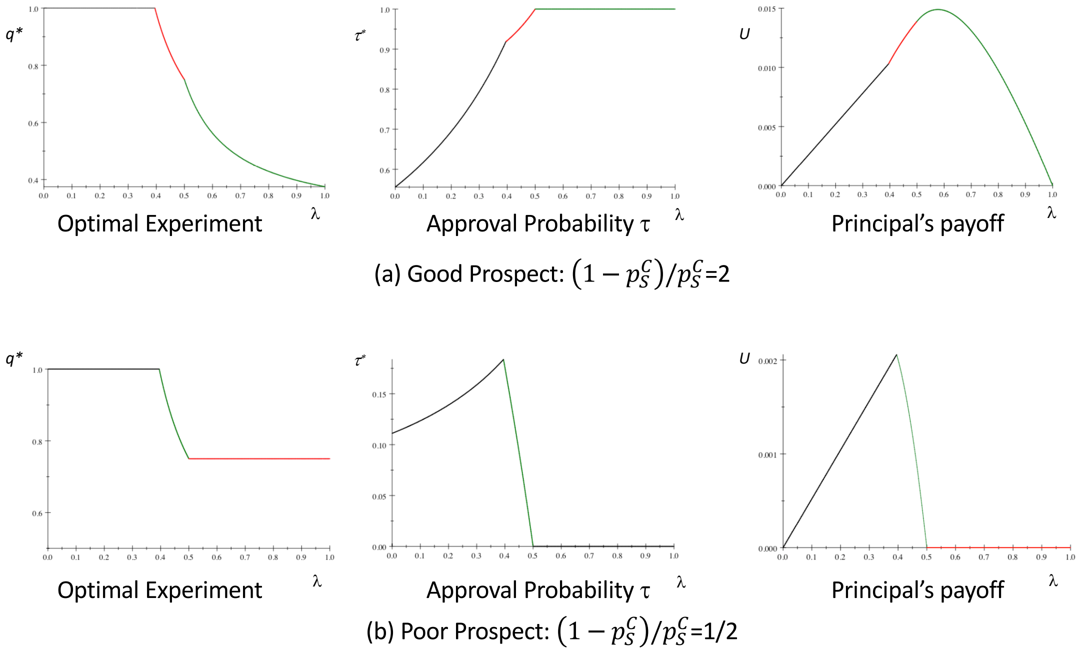
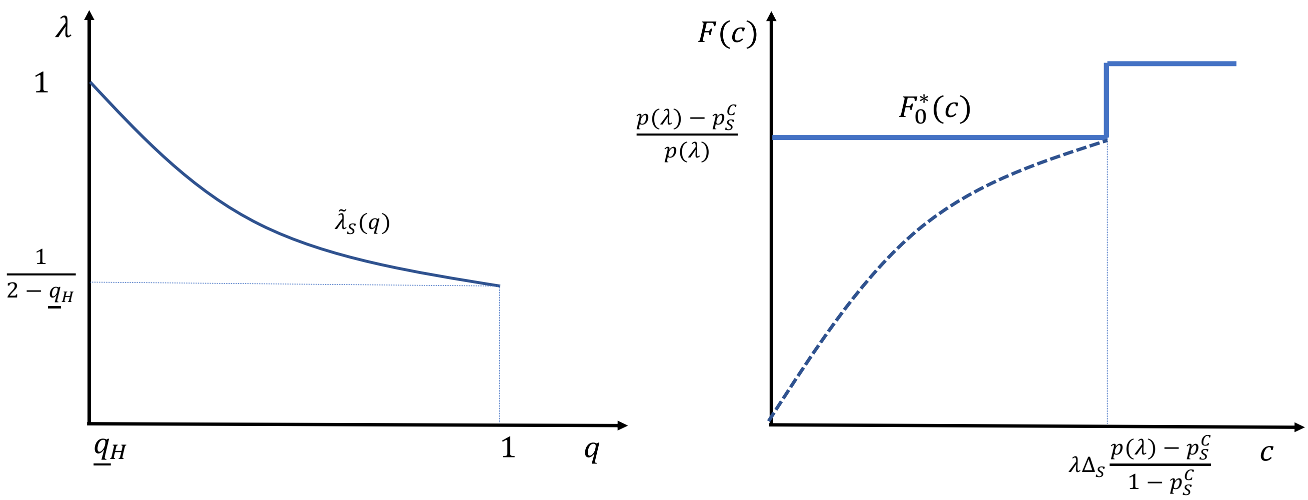


Figure 3: Equilibrium experimentation as a function of auditing for (a) a good prospect, and (b) a poor prospect.

fully informative experiment and increased auditing simply raises approval probability—but it monotonically decreases when the designer actually switches to a less informative experiment. In fact, for $\lambda > 1/2$, the designer selects the commitment experiment so that experimentation creates no value for the principal. Figure 3-b describes the principal's utility which is maximized for $\lambda = 0.39$. So, for poor-prospect ideas, the principal seldom audits the experiment and the designer in response does not reduce experimentation—i.e., the designer's experiment fully reveals the state. Nevertheless, such lax auditing implies that approval largely relies on the principal vetting the agent's recommendation.

5.3 Optimal Tampering Prevention

Consider now a fixed auditing intensity λ and suppose that the principal is unconstrained in her choice of F. Consistent with our theme of "credible skepticism" to motivate experimentation, she will incentivize some tampering in equilibrium by selecting a prevention policy that actually makes low tampering costs sufficiently likely.



(a) Maximum λ that induces {0,q}.

(b) Minimum cost distribution to induce $\{0, q_S^*(\lambda, \mathcal{S})\}$

Figure 4: Auditing and optimal tampering cost distribution.

Proposition 6. *Fix* $1/(2-\underline{q}_H) < \lambda < 1$ *and suppose that tasks are separated. Then,*
(i) For any prevention policy, we have

$$q_S^*(\lambda, \mathcal{S}) \leq \frac{\lambda}{2\lambda - 1}\underline{q}_H. \tag{15}$$

(ii) There is a multiplicity of prevention policies that achieve the bound in (15), all of them satisfying the following inequality

$$F(c) \geq \frac{c}{c + \lambda\frac{p_S^C \Delta_S}{1-p_S^C}} \text{ for } c \leq \lambda\Delta_S\frac{p(\lambda)-p_S^C}{1-p_S^C} \text{ with } p(\lambda) = 1 - \frac{(2\lambda-1)}{\lambda}p_S^C. \tag{16}$$

We prove this proposition by solving an auxiliary problem: to find the maximum auditing intensity that induces the selection of experiment $\pi = \{0, q\}$ for some cost distribution. The solution is $\widetilde{\lambda}_S(q) = q/(2q - \underline{q}_H)$–depicted in Figure 4(a)–which is obtained by ensuring that switching to the commitment experiment $\{0, \underline{q}_H\}$ is never profitable for the designer. Inverting this relation, we then obtain the most informative experiment consistent with auditing intensity λ.

There are many different distributions that would lead the designer to select $\{0, \frac{\lambda}{2\lambda-1}\underline{q}_H\}$– in all of them the distribution $F(c)$ exceeds some lower bound–see dotted line in Figure 4-b. That is, tampering for low realizations must be sufficiently likely so that the principal can commit to high approval rates only if experiments are sufficiently informative. Our argument didn't require the distribution to be smooth or to have a density. One distribution that satisfies (16) is supported only on two cost realizations, 0 and $\lambda\Delta_S\frac{p(\lambda)-p_S^C}{1-p_S^C}$–see the solid-line F_0^* in Figure 4-b–with $\Pr[c=0]=\left(p(\lambda)-p_S^C\right)/p(\lambda)$, and the agent only tampers if $c=0$; thus, expected tampering costs are zero.

5.4 Data Governance

Tampering prevention and detection are perfect substitutes when it comes to dissuading tampering. Can an organization improve its performance by simultaneously controlling both? One of our main insights is that optimal data governance calls for *both* a lax auditing intensity and a fallible tampering prevention system.

Proposition 7. *Consider a principal that specifies both the task allocation and data governance, then:*
(i) She sets a prior-independent auditing intensity

$$\lambda_{opt}^*=\frac{1}{2-\underline{q}_H}. \tag{17}$$

(ii) The designer selects a fully informative experiment $\pi=\{0,1\}$.
(iii) There is a multiplicity of optimal cost distributions but, among them, the following minimizes expected tampering costs,

$$F_{opt}^*(c)=\begin{cases}\frac{\mu\left(1-\underline{q}_H\right)}{\underline{q}_H(1-\mu)} \text{ for } c\in[0,\frac{1-\underline{q}_H}{2-\underline{q}_H}),\\ 1 \text{ for } c\geq\frac{1-\underline{q}_H}{2-\underline{q}_H}.\end{cases} \tag{18}$$

For this cost distribution, the principal is indifferent between separating and integrating tasks.
(iv) The principal can implement (18) through a dual internal-external audit: Tampering is always costless, but an internal audit privately verifies the agent's report with probability $\frac{\underline{q}_H-\mu}{\underline{q}_H(1-\mu)}$ *and rectifies a tampered report.*

An important principle in organizing data analytics is that, under delegated experimentation, the organization must also allow, to some extent, tampering by agents. To do so optimally, the organization both raises the likelihood of low tampering costs and engages in lax auditing–the optimal auditing intensity (17) is always lower than 1, but higher than 1/2, and increases with the principal's approval threshold.

We prove this proposition by appealing to Proposition 6 and optimizing over λ. We also obtain that $q_S^*(\lambda_{opt}^*, \mathcal{S}) = 1$–this is Proposition 7-ii. To wit, under an optimal organization, the designer has no incentive to garble an experiment that reveals the underlying state and the principal rubberstamps any scale-up recommendation after an inconclusive audit. If the organization wants to minimize the costs imposed upon agents—say because of concerns with increasing hiring costs—then the optimal distribution makes tampering either costless or completely deters tampering (see Figure 4). This distribution also makes task allocation irrelevant, as expected tampering costs are always zero.

Our model also relates to the literature that studies optimal government regulation of markets with externalities. This literature has pointed out that sometimes it is optimal for the government to simultaneously impose ex ante policies (e.g., safety standards), which constrain what can be done before the externality is generated, and ex post policies (e.g., exposure to tort liability), which defines what may happen after the externality is generated --- e.g., Kolstad, Ulen, and Johnson, 1990; Marino, 1988; Shavell, 1984a, 1984b. In our setup, the principal also finds it optimal to use a combination of an ex ante policy (tampering prevention) and an ex post policy (tampering detection).

The optimal organization that satisfies (18) can be afforded an intuitive implementation: the agent faces no cost of tampering but his report is subjected to a *decoupled internal-external audit*. First, the report is internally audited, albeit the probability of elucidating the true outcome is restricted to $\frac{\underline{q}_H - \mu}{\underline{q}_H(1-\mu)}$. If the internal audit is conclusive, however, it ensures that the report is consistent with the experimental outcome. Second, this report is subjected to an imperfect external audit, which is conclusive with probability $1/(2-\underline{q}_H)$. In summary, the organization conducts more intense internal audits for poor prospects—i.e., when scaling-up is less likely under a perfect audit—but commits to a prior-independent external audit.

Importantly, the outcome of the internal audit must be unknown to the principal. This decoupling of audits is essential to incentivize experimentation: if the outcome of the internal audit were known to the principal, the designer in anticipation would then select the commitment experiment $\left\{0, \underline{q}_H\right\}$. The accounting literature is also concerned with the possible effects of internal control audits and, in particular, whether the public disclosure of internal control audits should be mandatory. For example, Lennox and Wu (forthcoming) study the effects of regulation mandating the disclosure of internal control audits in China.

They present evidence that mandatory disclosure of internal control audits can significantly reduce the quality of information.

5.5 Tampering costs incurred only if audit is conclusive.

A conclusive audit reveals both the true experimental outcome and whether tampering took place. Suppose then that tampering costs are only borne if the agent is found to have tampered. We can adjust our analysis by setting the expected tampering cost to λc when the cost realization is c. A full analysis of this case can be found in the online Appendix B.

For a fixed $\lambda < 1$, Proposition 1 and 2 still apply, albeit with different threshold values $\bar{c}(q)$. To see how this affects the designer's payoff, consider $\{0, q\} \in \Pi_S$ that induces tampering threshold $\bar{c}$. Then, expressions (3-4) translate to

$$\bar{F}\left(\bar{c}\right) p = p_S^C \text{ with } p = \frac{q-\mu}{q},$$
$$\bar{c} = \frac{1-\lambda}{\lambda}\tau\Delta_S.$$

The first condition is identical to (4): the tampering threshold determines the probability of truthful reporting given an unfavorable outcome–i.e., $\bar{F}\left(\bar{c}\right)$– so that the probability of observing message $m = 0$ must be equal to p_S^C, irrespective of whether expected tampering costs are c or λc. However, the second condition shows that the corresponding scale-up probability must be lower to account for the lower tampering cost.

Nevertheless, the comparative statics of experimentation with respect to the two organizational levers remain the same: separating tasks or decreasing auditing intensity always increases experimentation.[22] What are the organizational implications if the agent's tampering cost reduces to λc? It is still true that the principal prefers to separate tasks and to commit to an imperfect audit whenever the designer is responsive to auditing–however, the conditions for designer responsiveness are now more stringent. Moreover, if the principal can freely shape the distribution of tampering costs through tampering prevention policies and can commit to ruling out decisions, then the same organizational design as in Proposition 7 remains optimal-see online Appendix B.

[22]The only notable difference with respect to Proposition 4 is that the designer's payoff is no longer supermodular in (minus) the tampering threshold and the auditing intensity.

6 Organizing for Scale

Suppose now that the firm "organizes for scale:" the principal's alternative to scaling-up is either to retain the status-quo or to scale-down. We show in Section 6.1 that the designer may now select an experiment recommending d_H or d_L instead of one recommending d_H or d_S, and in Section 6.2 we discuss how this extra choice affects data governance.

6.1 Equilibrium Experimentation and Designer's Payoffs

Looking at the communication subgame, the fact that all messages are on-path–see Assumption 1– does not rule out multiple equilibria as the principal may randomize after a conclusive or inconclusive audit. For instance, if the designer selects $\{\underline{q}_S, q\}$ then the principal could choose different (mixed) decisions as a function of the agent's report after a conclusive audit yields $s = \underline{q}_S$, or as a function of the audit after the agent reports $m = \underline{q}_S$. Thus, the designer's payoff from experiment $\{\underline{q}_S, q\}$ varies with the equilibrium in the communication subgame, which affects the designer's preferences over experiments.

To characterize the designer's equilibrium choice, we introduce the set of robust experiments Π_i^R, $i \in \{L, S\}$. These are binary experiments of the form $\{\underline{q}_i, q\}$ such that the principal when indifferent–e.g., after message $m = \underline{q}_i$ or after a conclusive audit determines $s = \underline{q}_i$– selects the most favorable action to the agent.

Definition (Robust Experiments) *Define the class $i \in \{L, S\}$ of robust experiments Π_i^R indexed by $\tau \in [0, 1]$, as*

$$\Pi_i^R = \left\{ \pi = \{\underline{q}_i, q\} : q \geq \underline{q}_H \text{ and } \bar{F}\left(\bar{c}_i(\tau)\right) = p_i^C \left(\frac{q - \underline{q}_i}{q - \mu} \right), \tau \in [0, 1] \right\} \tag{19}$$

with $\bar{c}_i(\tau)$ given by

$$\bar{c}_S(\tau) = \tau \left(1 - \lambda\right) \Delta_S \text{ and } \bar{c}_L(\tau) = \bar{c}_S(\tau) + \left(1 - \lambda\right) \left(\Delta_L - \Delta_S\right), \ \tau > 0, \tag{20}$$

and $\bar{c}_i(\tau) = 0$ if $\tau = 0$, alongside the equilibrium behavior

$$d_U\left(\underline{q}_i\right) = d_I\left(m, \underline{q}_i\right) = d_i; d_I\left(m, q\right) = d_H; d_U\left(q\right) = \tau d_H + (1 - \tau) d_S.$$

Finally, let $\Pi^R \equiv \Pi_L^R \cup \Pi_S^R$ denote the set of robust experiments.

We will refer to Π_S^R as "status-quo" (robust) experiments and to Π_L^R as "up-or-down" (robust) experiments[23] and describe "up-or-down" experiments Π_L^R as the "more informative" class.[24] To present our main equilibrium characterization, let $v_i(\tau, \mu; \lambda, k)$ $i \in \{L, S\}$, be the designer's equilibrium payoff after he selects $\pi_i(\tau) \in \Pi_i^R$, and let

$$\begin{aligned} V_i(\mu; \lambda, k) &\equiv \max_{\tau \in [0,1]} v_i(\tau, \mu; \lambda, k), \\ \bar{V}(\mu; \lambda, k) &\equiv \max\{V_S(\mu; \lambda, k), V_L(\mu; \lambda, k)\}. \end{aligned}$$

Proposition 8. *Let $\lambda > 0$ and $\mu \in (\underline{q}_S, \underline{q}_H)$. Then,*
(i) there is always an equilibrium of the design subgame in which the designer selects a robust experiment,
(ii) if the designer obtains payoff V^* in some equilibrium of the design subgame, then $V^* \geq \bar{V}(\mu; \lambda, k)$. If $V^* > \bar{V}(\mu; \lambda, k)$ and π^* is an equilibrium experiment, then $\underline{q}_S \in S(\pi^*)$.

Proposition 8 justifies our restriction to robust equilibria when the principal organizes for scale. First, there is always an equilibrium in which the designer responds by selecting an experiment in Π^R. Second, the designer can always guarantee himself $\bar{V}(\mu; \lambda, k)$ in any equilibrium. Indeed, experiment $\{\underline{q}_S + \varepsilon, q\}$, with ε small, results in a unique designer's payoff as the principal never randomizes after a conclusive audit. This uniqueness of equilibrium payoff guarantees that $V^* \geq \bar{V}(\mu; \lambda, k)$. However, if $\underline{q}_S \in S(\pi)$, the principal could use her indifference after a conclusive audit shows $s = \underline{q}_S$ to minimize tampering by announcing that she would treat favorably truth-telling and unfavorably tampering. In fact, whenever $\lambda \geq \Delta_S/\Delta_L$ the designer can obtain the commitment payoff from experiment $\{\underline{q}_S, \underline{q}_H\}$–this is the case if the principal threatens to scale-down if the agent is caught tampering. This proves that $V^* > \bar{V}(\mu; \lambda, k)$ in this case. We defer to Section 6.3 a discussion of organizational design with non-robust experiments.

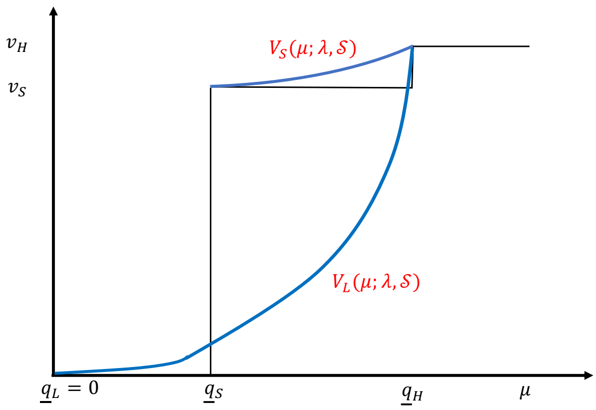
Lemma 3 extends the payoff characterization in Lemma 1 to this case.

Lemma 3. *For $\mu \in \left[\underline{q}_i, \underline{q}_H\right]$, consider $\{\underline{q}_i, q\} \in \Pi_i^R$ that induces threshold $\bar{c}$. Then,*

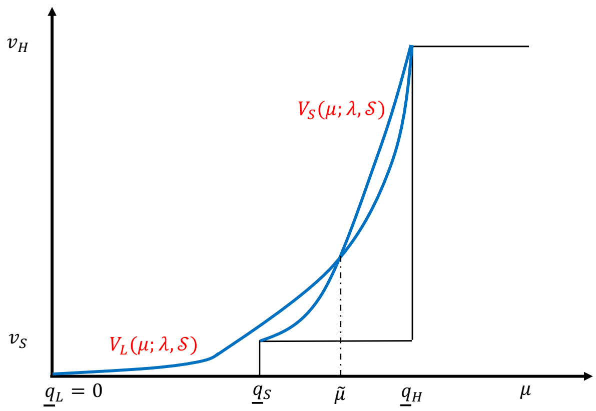
$$v_i(\tau(\bar{c}), \mu; \lambda, k) = v_H - (1-\lambda)\Delta_i + \bar{c} - m_i(\bar{c}; \lambda, k)(\underline{q}_H - \mu), \tag{21}$$

[23]To streamline the exposition, we drop the qualifier "robust" in this section when describing experiments in Π^R.

[24]This terminology captures the fact that for each experiment in Π_S^R there is always an experiment in Π_L^R that is Blackwell-more informative.



(a) Status-quo experiment optimal if $\lambda = 1$.



(b) Up-or-down experiment optimal if $\lambda = 1$.

Figure 5: Designer's equilibrium payoff for different experiment classes.

with m_i the slope of the designer's payoff with respect to the prior μ:

$$m_i(\bar{c};\lambda,k) \equiv \frac{\Delta_i}{\underline{q}_H - \underline{q}_i}\left(\frac{\lambda}{\overline{F}(\bar{c})} + \frac{\bar{c}}{\Delta_i} + \mathbb{I}_{\{k=\mathcal{I}\}}\frac{\eta(\bar{c})}{\Delta_i}\right). \tag{22}$$

Figure 5 depicts the designer's payoff under separation if he selects a status-quo experiment, $V_S(\mu;\lambda,\mathcal{S})$, and an up-or-down experiment $V_L(\mu;\lambda,\mathcal{S})$, for some $\lambda < 1$. $\bar{V}(\mu;\lambda,\mathcal{S})$ is then computed as the upper envelope of these payoffs. These graphs highlight a feature that distinguishes our model with tampering from a model with commitment: while in the commitment case this payoff is always concave, here the designer's equilibrium payoff is (locally) strictly convex in the prior for $\mu \in \left[\underline{q}_S, \underline{q}_H\right]$. This is a reflection that he actually changes the experiment (switching to one with a higher τ and hence higher $\bar{c}$) as μ increases. In fact, as shown in Figure 5-b, a higher prior can lead to switches in the optimal class of experiment. Indeed, in spite of an "up-or-down" experiment being optimal in Figure 5-b when $\lambda = 1$, if $\lambda < 1$ then the designer may actually switch to a less informative "status-quo" experiment.

For the remainder, we let $\bar{c}_i^*(\lambda,k)$ be the agent's optimal tampering threshold when restricted to an experiment in Π_i^R and $i^*(\lambda,k)$ be the class of the designer's optimal experiment.

6.2 Organizing for Scale: Organizational Design

To understand the trade-offs in organizational design, consider the principal's expected utility from experiment $\pi = \{\underline{q}_i, q\} \in \Pi_i^R$:

$$U(\lambda, k; \pi) = \underline{q}_H + \Pr[s = q]\lambda(q - \underline{q}_H) + \\ + \Pr[s = \underline{q}_i][1 - (1-\lambda)\Pr[m = q | s = \underline{q}_i]](E_{\underline{q}_i}[u(d_i, \theta)] - \underline{q}_H). \quad (23)$$

The principal now benefits from π in two ways. First, she selects d_H after auditing a scale-up recommendation. Second, when $\pi \in \Pi_L^R$, she selects d_L after $s = \underline{q}_L$ except when the agent tampers and her audit is inconclusive. Therefore, the gain from scaling-down can be increased by discouraging tampering. Thus, she needs to strike a balance between inducing the designer to experiment more while restraining at the same time the agent from tampering.

Using (23), we can write the principal's equilibrium expected utility before the design subgame:[25]

$$U(\lambda, k) = \underline{q}_H + \left(\underline{q}_H - \mu\right)\left(\lambda \frac{F(\bar{c}^*)}{\bar{F}(\bar{c}^*)} + \mathbb{I}_{\{i^* = L\}}\left(1 + \lambda \frac{F(\bar{c}^*)}{\bar{F}(\bar{c}^*)}\right)\left(\frac{\alpha_L - \underline{q}_H}{\underline{q}_H - \underline{q}_L}\right)\right). \quad (24)$$

Expression (24) clarifies that, given equilibrium behavior, the principal is again always willing to trade-off more distortions in communication for more informative experimentation. The reason is twofold. First, she strictly benefits from selecting d_H only if the audit is conclusive, thus making any distortion in communication irrelevant. Second, an experiment with a higher tampering threshold $\bar{c}^*$ lowers her expected utility conditional on $s = \underline{q}_L$, but also makes this outcome more likely; the combined effect on her expected utility is proportional to $1 + \lambda F(\bar{c}^*)/\bar{F}(\bar{c}^*)$ which increases in $\bar{c}^*$.

6.2.1 Optimal Task Allocation and Tampering Detection

How should the principal allocate tasks and design data governance? From (24), it remains true that she would separate tasks and commit to an imperfect audit in an effort to incentivize experimentation. When organizing for scale, however, she must guard against *adverse*

[25]Recall that for a $k-$allocation and $\lambda-$auditing, $i^* = i^*(\lambda, k)$ is the class of the designer's optimal experiment, and $\bar{c}^* = \bar{c}^*_{i^*}(\lambda, k)$ the agent's tampering threshold.

class switches—namely, a designer's switch from an "up-or-down" experiment to a "status-quo" experiment—in response to an imperfect audit. This terminology is motivated by the following observation: under task separation, a switch to a "status-quo" experiment makes the principal (weakly) worse off.

Lemma 4. *Let $\phi_i \equiv \left(1 - p_i^C\right) / \left(p_i^C \Delta_i\right)$, $i = L, S$, and suppose that $L(c) - (\phi_i/\lambda)$ is single-crossing in $[0, \Delta_i]$, and a perfect audit leads the designer to select an "up-or-down" experiment. Then, if tasks are separated, for any $\lambda < 1$ the principal is weakly better off if the designer is constrained to "up-or-down" experiments.*

The concern with adverse class switches underlies the main differences between organizing to innovate and organizing for scale. First, to ensure that the designer selects an "up-or-down" experiment, she may now prefer to integrate tasks. Second, she may still favor an imperfect audit as long as it does not trigger an adverse class switch.

Proposition 9. *(i) Fix λ and F and, letting*

$$W\left(\Delta, \bar{c}\right) \equiv \frac{\int_0^{\bar{c}} c dF(c)}{\lambda\Delta + \bar{c}\overline{F}\left[\bar{c}\right]}, \tag{25}$$

suppose that $W\left(\Delta_S, (1-\lambda)\Delta_S\right) > W\left(\Delta_L, (1-\lambda)\Delta_L\right)$. Then, there exist $\underline{q}_S$ and μ with $\underline{q}_S < \mu < \underline{q}_H$ such that the principal integrates tasks.
(ii) Suppose that either (ii-a) $p_L^C > (\Delta_S/\Delta_L)\, p_S^C$ and $f(0) > 0$, or (ii-b) $p_L^C < (\Delta_S/\Delta_L)\, p_S^C$, the designer is responsive to auditing and $i^*\left(k, \lambda\right) = L$. Then $\lambda^* < 1$ in every equilibrium of the organizational design game.

Integration can be optimal if the principal's auditing leads the designer under separation to select a status-quo experiment while he would instead select an "up-or-down" experiment under integration. Proposition 9-i provides a sufficient condition for such case. Proposition 9-ii provides sufficient conditions for optimal auditing to be imperfect. First, if $p_L^C > (\Delta_S/\Delta_L)\, p_L^C$ then the designer already selects an status-quo experiment under a perfect audit, eliminating any concern that an imperfect audit might trigger a switch; then, a similar argument as in Proposition 5 guarantees that $\lambda^* < 1$. If $p_L^C < (\Delta_S/\Delta_L)\, p_L^C$, however, the designer selects experiment $\{\underline{q}_L, \underline{q}_H\}$ when the audit is perfect and lowering auditing intensity might induce an adverse switch. In this case, $\lambda^* < 1$ obtains as long as there is an auditing intensity that avoids a switch to a status-quo experiment.

6.2.2 Discretion, Auditing, and Experimentation.

The only reason for setting $\lambda^* = 1$ is that any imperfect audit that motivates valuable experimentation also leads to an adverse class switch. Nevertheless, the principal can always avoid adverse switches if she can commit ex-ante to ruling out the status-quo, thus committing to selecting from extreme options.

Corollary 1. *Suppose that $f(0) > 0$. If the principal can ex-ante commit to ruling out decisions, then in every equilibrium we have $\lambda^* < 1$ and the principal prefers to separate tasks.*

This result resonates with insights from the delegation literature in which the principal rules out intermediate decisions to improve the informational content of delegated decision making (see, e.g., Szalay (2005) and Alonso and Matouschek (2008)). For example, Szalay (2005) shows that, to boost incentives to acquire information, a principal may rule out the (agent's) optimal decision if uninformed. In our case, agents interested in promoting a specific option (i.e., scaling up) are more willing to compromise on the status quo than the principal. Then, limiting discretion boosts experimentation by eliminating such compromise. Note also that restricting the principal's choice is accompanied by less intense monitoring: i.e., she sets $\lambda^* < 1$ if she is able to rule-out decisions. Thus, in this context, discretion and monitoring act as complements.

6.2.3 Data Governance

If $p_L^C < (\Delta_S/\Delta_L)\, p_S^C$, so that the designer selects an up-or-down experiment when $\lambda = 1$, then we show below that the optimal data governance in Proposition 7 remains optimal. If $p_L^C \geq (\Delta_S/\Delta_L)\, p_S^C$, however, then to optimally induce an "up-or-down" experiment the principal integrates tasks and the optimal cost distribution must make tampering strictly costly. The reason is that, if tasks are separated, the designer always selects a status-quo experiment if he prefers the status-quo experiment when $\lambda = 1$ (see discussion in Online Appendix B). The principal must then incentivize a class switch by integrating tasks and equilibrium tampering must be strictly costly. Nevertheless, and consistent with Corollary 1, if she can commit to ruling out decisions–so that she can induce an "up-or-down experiments" by ruling out decision d_S– then the optimal organization would always lead to full experimentation and rely on an internal-external dual audit.

Proposition 10. *Suppose that either (i)* $p_L^C < (\Delta_S/\Delta_L)\, p_S^C$*, so that the designer selects an up-or-down experiment when data governance is perfect, or, (ii)* $p_L^C \geq (\Delta_S/\Delta_L)\, p_S^C$ *and she can ex-ante commit to ruling out decisions. Then, the optimal data governance satisfies (17) and (18) in Proposition 7 and induces full experimentation by the designer.*

6.3 Organizing for scale with non-robust experimentation.

Our analysis has focused on equilibria in which the principal resolves any indifference after a conclusive audit by selecting the agents' preferred decision. While this is without loss when the principal organizes to innovate, these equilibria provide only a lower bound on the designer's equilibrium payoff when organizing for scale–see Proposition 8. Thus, one wonders if our insights may change if the principal instead conditions her audited decision on whether the agent tampered. We show that, while optimal auditing intensity may vary, our main qualitative results hold when we allow for "non-robust" equilibria.

Proposition 8-ii shows that if the designer's equilibrium payoff exceeds that of a robust experiment, then $\underline{q}_S \in S(\pi^*)$, i.e., the designer's experiment is (possibly a mixture involving) an experiment $\{\underline{q}_S, q\}$. The following lemma shows that the designer's maximum payoff from an experiment $\{\underline{q}_S, q\}$ comes when the principal punishes, to some extent, tampering.

Lemma 5. *Suppose that the designer selects* $\{\underline{q}_S, q\}$ *and let* $\bar{c}_S$ *be the tampering threshold associated with the robust experiment* $\{\underline{q}_S, q\} \in \Pi^R$*; i.e.,* $\bar{c}_S$ *satisfies* $\bar{F}\left(\bar{c}_S\right) = p_S^C \left(\frac{q-\underline{q}_S}{q-\mu}\right)$*–see (19). If* $\tau_I(q)$ *denotes the probability that the principal selects* d_S *after the agent is found tampering, then at the equilibrium of* $\{\underline{q}_S, q\}$ *which maximizes the designer's payoff we have* $\tau_I(q) = \tau_I^*(q)$ *with*

$$\tau_I^*(q) \equiv \min\left\{\tau_I(q) : \bar{c}_S = (1-\lambda)\tau_U(q)\Delta_S - \lambda(1-\tau_I(q))(\Delta_L - \Delta_S), \tau_U(q) \in [0,1]\right\}.$$

In particular, if the principal rubberstamps a scale-up recommendation for $\{\underline{q}_S, q\} \in \Pi_S$*, then* $\tau_I^*(q) = 1$*.*

If $\tau_I^*(q) < 1$ then the principal punishes tampering by lowering the probability of selecting d_S if the agent tampered. The designer then obtains a higher payoff from $\{\underline{q}_S, q\}$ not because it reduces the tampering threshold–indeed, the tampering threshold $\bar{c}_S$ is the same as that of the robust status-quo experiment–but rather because it allows the principal to scale-up more often after an unaudited scale-up recommendation. In turn, the principal could use the designer's preference for equilibria with $\tau_I(q) = \tau_I^*(q)$ as follows: she punishes tampering (i.e. sets $\tau_I(q) = \tau_I^*(q)$) if the designer's experiment is sufficiently informative–i.e., if the designer selects $\{\underline{q}_S, q\}$–while she is lenient if the designer selects a less informative status-quo experiment (i.e., $\tau_I(q') = 1$, if $q' < q$). While moving to non-robust experiments may lead the principal to choose a different auditing intensity, our main qualitative results of Section 5 nevertheless hold for non-robust experimentation.

Proposition 11. *Suppose that we restrict attention to PBE of the design subgame that maximize the principal's expected payoff. Then: (i) relative to Proposition 9, the principal is now more likely to set $\lambda^* < 1$ and to separate tasks, and (ii) the auditing intensity and cost distribution in Proposition 7 remain optimal.*

The only reason for the principal to set $\lambda^* = 1$ is to avoid an adverse class switch. However, the principal can induce a (weakly) more informative non-robust status-quo experiments for the same λ, so that the previous class switch may now be profitable for the principal. This explains Proposition 11-i. Moreover, the principal can increase the designer's payoff from experiment $\{\underline{q}_S, q\}$ only if she does not rubberstamp a scale-up recommendation for a robust experiment–i.e., only if $\tau_U(q) < 1$ for $\tau_I(q) = 1$. Nevertheless, the optimal organization in Proposition 7 is based on finding cost distributions that implement $\{\underline{q}_S, q\}$ with $\tau_U(q) = 1$, thus the principal cannot increase the designer's payoff for such experiments by punishing tampering. That is, if the principal can shape the distribution of tampering costs through tampering prevention policies and select her preferred PBE for each subgame, then the auditing intensity (17) and cost distribution (18) remain optimal.

7 Discussion and Concluding Remarks

In this paper, we develop a model of data analytics and argue that organizations that delegate experimentation to their agents must also create a culture of "credible skepticism" by limiting decision-makers' ability to assess the truthfulness of the information they receive. We now discuss these findings in the context of several strands of the literature, after which we conclude.

7.1 Related Literature

Literature on decision-making processes in organizations:

Our analysis contributes to the study of decision making processes in organizations and, in particular, to how organizations optimally react to the incentive conflicts that members face (see Gibbons, Matouschek, and Roberts (2013) and Bolton and Dewatripont (2013) for excellent surveys of this literature). For instance, in models of strategic delegation, the organization would like to assign authority to a party whose preferences may differ from those of the organization as these affects the production and communication of information (for instance, in Dessein, 2002, delegation to a biased intermediary can improve cheap-talk communication with experts).[26] One recent example is Nayeem (2017), who quantifies the value of appointing a decision maker that is harder to convince to approve a project–e.g., as his preference for a "good project" are weaker than those of the organization. That is, there is value in appointing a "skeptic" for project approval. In our model, however, the principal cannot credibly delegate the decision to someone else nor commit to biasing decisions in favor of agents. Skepticism arises not because of differing preferences, but as an attitude to (rationally) doubt the claims made by others.

Our paper also contributes to the literature that studies how "light monitoring" of agents' recommendations may avoid crowding-out their efforts to experiment (see, e.g., Aghion and Tirole (1997)). In our case, imperfect auditing allows the principal to refrain from adopting the agent's self-serving recommendation, thus, spurring experimentation.

The literature on task allocation has emphasized that task separation can allow for the provision of higher power incentives in each task (Holmstrom and Milgrom, 1991, Dewatripont, Jewitt, and Tirole, 2000) or improve information acquisition (Dewatripont and Tirole, 1999). Moreover, in sequential tasks, task separation may increase the information generated in the first task to incentivize the second (Lewis and Sappington, 1997, Landier, Sraer, and Thesmar (2009)), or can be optimal under effort externalities between tasks (Schmitz, 2013). We also find that task separation allows for stronger incentives to experiment, even though we do not allow for explicit incentives, as separation provides a "coarse" instrument

[26] More generally, decision makers may be able to commit to ex-post biasing decisions in favor of experts, e.g., in a relational setting as in Alonso and Matouschek, 2007.

to lower the costs of experimentation.

Literature on Information acquisition and Communication

We contribute to the literature that studies models of delegated expertise (Demski and Sappington (1987))–in particular, models in which a decision maker relies on the information actively gathered and communicated by experts. For instance, Pei (2015), Argenziano, Severinov, and Squintani (2016), and Deimen and Szalay (2019) consider models where an agent decides what information to gather if communication with the principal takes the form of cheap talk, while Che and Kartik (2009) considers certifiable disclosure.[27] Argenziano et al. (2016) and Deimen and Szalay (2019) use the threat of off-path "bad" communication (e.g., a reversion to a "babbling" equilibrium) if the expert acquires less information to motivate information acquisition. In Pei (2015), communication is "frictionless:" the agent reveals all the information gathered if acquiring a less informative signal is always feasible (and less costly) (see also Gentzkow and Kamenica (2016)). In Che and Kartik (2009), incentives to acquire information come from players having different priors: an expert has a stronger incentive to be informed relative to the common prior case as he expects that better information will lead the principal to, on average, embrace his point of view.[28]

A main insight in these papers is that frictions in communication can be used to discipline agents if they underinvest in information acquisition.[29] While this insight resonates with our main finding, our mechanism is markedly different. In contrast to Pei (2015), Argenziano et al. (2016), and Che and Kartik (2009), the agent faces no explicit cost in acquiring more information in our model–this matches our main application where data becomes available to the organization automatically through its normal operation. In contrast to Deimen and Szalay (2019), we consider an explicit cost of misrepresentation when the agent communicates the results, as well as the principal's ability to audit the agent's message and to allocate tasks to different agents.

[27]Our communication stage is also related to models of communication with lying costs–e.g., Kartik, Ottaviani, and Squintani (2007) and Kartik (2009). Relative to these models, our communication model is simpler, as we consider a message independent tampering cost, but we incorporate an information acquisition stage prior to communication.

[28]Alonso and Câmara, 2016 also show that differences of opinion generically give rise to incentives to persuade a principal.

[29]Frictions in communication can also improve information transmission, see e.g., Blume et al (2007)

Literature on relaxing the commitment assumption in models of Bayesian persuasion.

Our paper contributes to the literature that relaxes the sender-commitment assumption in Kamenica and Gentzkow's model of Bayesian Persuasion.[30] Papers in this recent literature differ on the modeling of imperfect commitment. For instance, Guo and Shmaya (Forthcoming) consider a model of costly miscalibration: the sender decides the statistical properties of an experiment and can deviate from the "asserted" meaning for each outcome at a cost related to the difference between the asserted meaning and its true meaning. That is, they allow for a sender's private experimental design rather than our public experimental design subject to private output-tampering. Min (2020) considers the output-tampering case but tampering only occurs with some exogenous probability and explores the effect of changes in this probability in Crawford and Sobel (1982) uniform-quadratic case. In these papers, there is no tampering or misrepresentation in equilibrium.[31] Instead, in our paper tampering is a generic equilibrium phenomenon resulting from the principal's choice of auditing intensity. Perez-Richet and Skreta (2021) study test design under costly state falsification: a designer selects a test and an agent can change its input at a cost. That is, in contrast to our setup with output-tampering, the agent engages in input-tampering. Fréchette, Lizzeri, and Perego (2019) analyze experiments in which the level of commitment can vary across treatments, albeit the ability to tamper is exogenously given, while it is an equilibrium outcome in our paper.

Closest to our modeling of limited commitment are Lipnowski, Ravid, and Shishkin (2018) and Nguyen and Tan (2018). Lipnowski et al. (2018) consider an information design setup with output-tampering where the tampering probability depends on the actual message/state and provide an elegant geometric characterization of the sender's value of persuasion. While in our setup the tampering probability is also message/state dependent, it arises endogenously from the agent's equilibrium incentives to tamper. Furthermore, while they characterize the optimal level of credibility from the sender's perspective, our focus lies on the receiver/principal's perspective. Nguyen and Tan (2018) also study public experimen-

[30]See also the literature on strategic sample selection, e.g., Tillio, Ottaviani, and Sørensen (2017), Tillio, Ottaviani, and Sørensen (2021), Adda, Decker, and Ottaviani (2020), Felgenhauer and Loerke (2017) and Libgober (Forthcoming).

[31]Tampering-proof equilibria are the focus of Min (2020), while Guo and Shmaya (Forthcoming) show that there is always a Sender-optimal equilibrium with a calibrated strategy–i.e., such that receiver correctly anticipates its meaning. See also Sobel (2020) for an analysis that distinguishes between "lying" and "deception".

tation subject to private output-tampering. They consider a setup with a fixed experimental outcome space and message space, and a communication technology where each message carries a cost that depends both on the message and the experimental outcome. They focus on conditions on this technology for the Sender's preferred equilibrium to be supported without tampering (Condition 1 in Nguyen and Tan (2018)). Out setup does not satisfy Condition 1 (as the tampering cost is the same regardless of the message sent) and, thus, we cannot apply their results.

One overarching theoretical difference with this literature is that we endogenize the sender's commitment power by allowing the receiver to select among different organizational practices; for instance, how much to audit the sender's message. Thus, while the literature shows that exogenously relaxing the sender's commitment can be beneficial for the receiver, we show the extent to which imperfect commitment is an equilibrium outcome of the receiver's organizational practices.

7.2 Concluding Remarks

The ICT revolution–by lowering the costs of data acquisition, storage and processing–has made managers more reliant on the insights derived from analyzing these data rather than the intuitions and opinions of other members of the organization. It would then seem that many of the trade-offs that drive the optimal organization to process information are no longer relevant. We argue that unresolved conflict still makes organizational structure meaningful as members handling data still decide which data to use and how to analyze it. We show that this poses a fundamental trade-off: dissuading misrepresentation also reduces data utilization, limiting the insights that agents derive from the data. Optimal organizations are then based on a culture of "credible skepticism:" decision makers have limited ability to audit the data and analytics behind the recommendations issued by agents, which invites tampering and misrepresentation in equilibrium.

The adoption of new technologies such as blockchain can eliminate tampering by effectively imposing an infinitely high tampering cost (Tapscott and Tapscott, 2017). Nevertheless, under delegated experimentation, this is never optimal for the firm as the optimal distribution of tampering costs must lead to some tampering in equilibrium. We showed that this optimal organization can be implemented through a decoupled internal-external audit: tampering is costless, but an internal (imperfect) audit can limit its effect by rectifying the tampered outcome with the true outcome. Then, an external audit is triggered with some probability without knowing whether the internal audit rectified the report. This system of consecutive audits strikes a perfect balance between experimentation and tampering and minimizes the tampering costs of agents. Importantly, under an optimal internal-external audit, the designer engages in full experimentation.

To focus on the trade-off between experimentation and misrepresentation, we offer a streamlined model. In particular, decision makers do not have access to alternative sources of information (i.e., they do not "seek a second opinion") nor do they induce competition between agents to persuade them. We see these extensions as promising and leave them for future work.

References

Adda, J., Decker, C., & Ottaviani, M. (2020). P-hacking in clinical trials and how incentives shape the distribution of results across phases. *Proceedings of the National Academy of Sciences*, *117*(24), 13386–13392.

Aghion, P., & Tirole, J. (1997). Formal and real authority in organizations. *Journal of Political Economy*, *105*(1), 1–29.

Alonso, R., & Câmara, O. (2016). Bayesian Persuasion with Heterogeneous Priors. *Journal of Economic Theory*, *165*, 672–706.

Alonso, R., & Matouschek, N. (2007). Relational delegation. *The RAND Journal of Economics*, *38*(4), 1070–1089.

Alonso, R., & Matouschek, N. (2008). Optimal delegation. *The Review of Economic Studies*, *75*(1), 259–293.

Argenziano, R., Severinov, S., & Squintani, F. (2016). Strategic information acquisition and transmission. *American Economic Journal: Microeconomics*, *8*(3), 119–155.

Bolton, P., & Dewatripont, M. (2013). Authority in organizations a survey. In R. Gibbons & J. Roberts (Eds.), *The handbook of organizational economics* (pp. 342–372).

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance? *Available at SSRN https://ssrn.com/abstract=18194*

Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, *106*(5), 133–39.

Che, Y.-K., & Kartik, N. (2009). Opinions as incentives. *Journal of Political Economy*, *117*(5), 815–860.

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, *50*(6), 1431.

Deimen, I., & Szalay, D. (2019). Delegated expertise, authority, and communication. *American Economic Review*, *109*(4), 1349–1374.

Demski, J. S., & Sappington, D. E. M. (1987). Delegated expertise. *Journal of Accounting Research*, *25*(1), 68.

Dessein, W. (2002). Authority and communication in organizations. *Review of Economic Studies*, *69*(4), 811–838.

Dewatripont, M., Jewitt, I., & Tirole, J. (2000). Multitask agency problems: Focus and task clustering. *European Economic Review*, *44*(4-6), 869–877.

Dewatripont, M., & Tirole, J. (1999). Advocates. *Journal of Political Economy*, *107*(1), 1–39.

Felgenhauer, M., & Loerke, P. (2017). Bayesian persuasion with private experimentation. *International Economic Review*, *58*(3), 829–856.

Fréchette, G., Lizzeri, A., & Perego, J. (2019). *Rules and commitment in communication: An experimental analysis*. Working Paper.

Gentzkow, M., & Kamenica, E. (2016). Disclosure of endogenous information. *Economic Theory Bulletin*, *5*(1), 47–56.

Gibbons, R., Matouschek, N., & Roberts, J. (2013). Decisions in organizations. In R. Gibbons & J. Roberts (Eds.), *The handbook of organizational economics* (pp. 373–431).

Goldfarb, A., & Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, *57*(1), 3–43.

Guo, Y., & Shmaya, E. (Forthcoming). Costly miscalibration. *Theoretical Economics*.

Guo, Y., & Shmaya, E. (2019). The Interval Structure of Optimal Disclosure. *Econometrica*, *87*(2), 653–675.

Holmstrom, B., & Milgrom, P. (1991). Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design. *The Journal of Law, Economics, and Organization*, *7*(special_issue), 24–52.

Kamenica, E., & Gentzkow, M. (2011). Bayesian Persuasion. *The American Economic Review*, *101*(6), 2590–2615.

Kartik, N. (2009). Strategic communication with lying costs. *Review of Economic Studies*, *76*(4), 1359–1395.

Kartik, N., Ottaviani, M., & Squintani, F. (2007). Credulity, lies, and costly talk. *Journal of Economic Theory*, *134*(1), 93–116.

Kolotilin, A. (2018). Optimal Information Disclosure: A Linear Programming Approach. *Theoretical Economics*, *13*(2), 607–635.

Kolotilin, A., Mylovanov, T., Zapechelnyuk, A., & Li, M. (2017). Persuasion of a Privately Informed Receiver. *Econometrica*, *85*(6), 1949–1964.

Kolstad, C. D., Ulen, T. S., & Johnson, G. V. (1990). Ex post liability for harm vs. ex ante safety regulation: Substitutes or complements? *The American Economic Review*, *80*(4), 888–901.

KPMG. (2016). *Building trust in analytics*. KPMG International Data Analytics.

KPMG. (2018). *Guardians of trust*. KPMG International Data Analytics.

Landier, A., Sraer, D., & Thesmar, D. (2009). Optimal dissent in organizations. *Review of Economic Studies*, *76*(2), 761–794.

Lewis, T., & Sappington, D. M. (1997). Information management in incentive problems. *Journal of Political Economy*, *105*(4), 796–821.

Libgober, J. (Forthcoming). False positives and transparency. *American Economic Journal: Microeconomics*.

Lipnowski, E., Ravid, D., & Shishkin, D. (2018). Persuasion via weak institutions. *SSRN Electronic Journal 3168103*.

Marino, A. M. (1988). Monopoly, liability and regulation. *Southern Economic Journal*, *54*(4), 913–927.

McKinsey, & Co. (2017). *Advanced analytics: Nine insights from the c-suite*. McKinsey Analytics.

McKinsey, & Co. (2018a). *Breaking away: The secrets to scaling analytics*. McKinsey Analytics.

McKinsey, & Co. (2018b). *Building an effective analytics organization*. McKinsey on Payments.

Min, D. (2020). Bayesian persuasion under partial commitment. *Unpublished*.

Nayeem, O. A. (2017). Bend them but don't break them: Passionate workers, skeptical managers, and decision making in organizations. *American Economic Journal: Microeconomics*, *9*(3), 100–125.

Nguyen, A., & Tan, T. Y. (2018). Bayesian persuasion with costly messages. *SSRN Electronic Journal 3298275*.

Pei, H. D. (2015). Communication with endogenous information acquisition. *Journal of Economic Theory*, *160*, 132–149.

Perez-Richet, E., & Skreta, V. (2021). Test design under falsification. *CEPR Discussion Papers 15627*.

Pfeffer, J., & Sutton, R. I. (2006). Evidence-based management. *Harvard business review*, *84*(1), 62.

Schmitz, P. W. (2013). Job design with conflicting tasks reconsidered. *European Economic Review*, *57*, 108–117.

Shavell, S. (1984a). A model of the optimal use of liability and safety regulation. *The RAND Journal of Economics*, *15*(2), 271–280.

Shavell, S. (1984b). Liability for harm versus regulation of safety. *The Journal of Legal Studies*, *13*(2), 357–374.

Sobel, J. (2020). Lying and deception in games. *Journal of Political Economy*, *128*(3), 907–947.

Szalay, D. (2005). The economics of clear advice and extreme options. *The Review of Economic Studies*, *72*(4), 1173–1198.

Tapscott, D., & Tapscott, A. (2017). How blockchain will change organizations. *MIT Sloan Management Review*, *58*(2), 10.

Tillio, A. D., Ottaviani, M., & Sørensen, P. N. (2017). Persuasion bias in science: Can economics help? *The Economic Journal*, *127*(605), F266–F304.

Tillio, A. D., Ottaviani, M., & Sørensen, P. N. (2021). Strategic sample selection. *Econometrica*, *89*(2), 911–953.

Wu, L., Hitt, L., & Lou, B. (2020). Data analytics, innovation, and firm productivity. *Management Science*, *66*(5), 2017–2039.

A Appendix

Proof of Proposition 1: Given $\pi = \{q, \Pr[q]\}_{i \in S(\pi)}$, with type space $S(\pi)$, consider a PBE of the communication subgame where the agent's reporting strategy is $m^*(q, c)$, which leads to decisions $d_I^*(m, s)$ and $d_U^*(m)$. Proposition 1-i follows immediately as the gain from tampering is the same for all agents that observe the same experimental outcome: if type (q, c) finds it profitable to send $q_z \neq q$ instead of q, all types with $c' < c$ will strictly prefer to tamper.

For part (ii), consider the set of tampered outcomes M_T defined in the proposition. Suppose that $q, q' \in M_T$ but the distributions $d_U^*(q)$ and $d_U^*(q')$ lead to different expected payoffs for the agent.[32] If $\underline{q}_i \notin S(\pi)$, $i \in \{S, H\}$, then the principal never mixes after a conclusive audit and the agent's payoff in this event is independent of the message sent. This is also the case if the audited decision $d_I^*(m, \underline{q}_i)$ is independent of m whenever $\underline{q}_i \in S(\pi), i \in \{S, H\}$. This implies that the agent only benefits from tampering in the event that the audit is inconclusive, but if $d_U^*(q)$ and $d_U^*(q')$ yield a different payoff, then $m^*(q, c)$ cannot be part of an equilibrium. Therefore, we must have that $d_U^*(q) = d_U^*(q')$ for $q, q' \in M_T$. Finally, suppose that $q \in M_T$, $q \neq \underline{q}_i, i \in \{S, H\}$. Then, the sender never gains from tampering, as the audited decision is independent of m and the unaudited decision would be the same if he had instead truthfully reported his type. ■

Proof of Proposition 2: (i) Consider an arbitrary finite experiment $\tilde{\pi} = \{q, \Pr[q]\}_{q \in S(\tilde{\pi})}$. Then, we show that there exists $\widetilde{\pi}_S \in \Pi_S$ that (weakly) improves the designer's payoff relative to $\tilde{\pi}$. Therefore, if π^* maximizes the designer's payoff when restricted to Π_S, then selecting π^* is a PBE of the design subgame, as the designer's expected utility cannot be improved by any alternative $\tilde{\pi}$.

Define $S_T(\tilde{\pi})$ as the set of tampering types:

$$S_T(\tilde{\pi}) = \{q \in S(\tilde{\pi}) : \Pr[m^*(q, c) \neq q] > 0\}, \tag{A.1}$$

and recall that, from Proposition 1, $M_T(\tilde{\pi})$ is the set of tampered outcomes. Thus, type $q \in S_T(\tilde{\pi})$ will tamper with positive probability while some tampering type will report $q' \in M_T(\tilde{\pi})$ with positive probability. Since $d_I^*(m, 0) = d_S$ whenever $0 \in S(\pi)$, Proposition 1 shows that $S_T(\tilde{\pi}) \cap M_T(\tilde{\pi}) = \emptyset$. We first show that tampering types correspond to low realizations of the experiment while tampered outcomes are associated with high realizations, i.e.,

$$q_{S_T} \equiv \max\{q : q \in S_T(\tilde{\pi})\} < \min\{q : q \in M_T(\tilde{\pi})\} \equiv q_{M_T}. \tag{A.2}$$

To see this, let d_U' be the decision following an unaudited tampered outcome—see Proposition 1-ii(a)—and suppose, by contradiction, that there are $q' < q''$ with $q' \in M_T(\tilde{\pi})$ and $q'' \in S_T(\tilde{\pi})$. Assumption 1 implies that message $m = q''$ is sent with positive probability and, as $q'' \notin M_T(\tilde{\pi})$, we must have that the posterior belief of the principal if the audit

[32] As the principal only mixes after an inconclusive audit when her posterior is either $\underline{q}_S$ (thus, mixing between d_L and d_S) or $\underline{q}_H$ (thus, mixing between d_S and d_H), the agent must obtain a different expected payoff after an inconclusive audit when reporting q and q' if these distributions are different.

is inconclusive must be q''. Since $q' < q'' \leq q_{S_T}$, and Proposition 1-ii(b) shows that type $s = q' \in M_T(\tilde{\pi})$ sends $m = q'$, the principal's posterior belief after an unaudited $m = q'$ must be strictly lower than q_{S_T}. But then, we must have $q_{S_T} \notin S_T(\tilde{\pi})$ as type q_{S_T} prefers to induce decision $d_U(q_{S_T})$ rather than tamper to induce d'_U, thus reaching a contradiction.

Next, partition $S(\tilde{\pi})$ by defining $X_S(\tilde{\pi}) = S(\tilde{\pi}) \cap \left(\underline{q}_S, \underline{q}_H\right)$ and $X_H(\tilde{\pi}) = S(\tilde{\pi}) \cap [\underline{q}_H, 1]$. We now show that (A.2) implies that all messages in $X_i(\tilde{\pi})$ lead to the same unaudited (mixture over) decision(s)–which means that all types in $X_i(\tilde{\pi})$ face the same gain from tampering and must therefore have the same tampering threshold. Proposition 1-ii(a) implies that this is true if $X_i(\tilde{\pi}) \subset M_T(\tilde{\pi})$. We will show by contradiction that there cannot be tampered outcomes as well as non-tampered outcomes in $X_i(\tilde{\pi})$. To see this, suppose that q_{M_T} defined in (A.2) satisfies $q_{M_T} \in X_i(\tilde{\pi})$ and there is some $q' \in X_i(\tilde{\pi})$ but $q' \notin M_T(\tilde{\pi})$. Then we must have $q' < q_{M_T}$, but $d^*_U(q') = d_i$ as the posterior after an unaudited message q' is precisely q'. However, the posterior after unaudited $q_{M_T} \in M_T(\tilde{\pi})$ must be strictly lower than q_{M_T}. But then we must have that $d^*_U(q_{M_T}) = d_i$, otherwise tampering types would send message q' instead of q_{M_T}. Thus, for all $q, q' \in X_i(\tilde{\pi})$, $d^*_U(q) = d^*_U(q')$.

We now construct the binary experiment $\tilde{\pi}_c = \{\tilde{q}^{X_S}, \tilde{q}^{X_H}\}$ that in equilibrium gives the designer the same expected utility as the equilibrium of $\tilde{\pi}$. We do so by replacing all realizations in $X_i(\tilde{\pi})$, $i = S, H$, with a realization $s = \tilde{q}^{X_i}$ that is its conditional expectation, i.e.,

$$\tilde{q}^{X_i} = \frac{\sum_{q \in X_i(\tilde{\pi})} \Pr[q]\, q}{\sum_{q \in X_i(\tilde{\pi})} \Pr[q]}, \quad \Pr\left[\tilde{q}^{X_i}\right] = \sum_{q \in X_i(\tilde{\pi})} \Pr[q],$$

and adjusting the equilibrium (mixture over) messages to

$$m_c(\tilde{q}^{X_i}, c) = \frac{\sum_{q \in X_i(\tilde{\pi})} \Pr[q] \left(\sum_{j=\{S,H\}} \sum_{q' \in X_j(\tilde{\pi})} \Pr[m(q,c) = q']\, \tilde{q}^{X_j}\right)}{\sum_{q \in X_i(\tilde{\pi})} \Pr[q]}.$$

That is, the probability that type $s = \tilde{q}^{X_i}$ sends message $m = \tilde{q}^{X_j}$ when the cost realization is c, is the conditional probability that a type in $X_i(\tilde{\pi})$ would send a message corresponding to a type in $X_j(\tilde{\pi})$. We complement the definition by having threshold type $\underline{q}_H$ send message $m = \tilde{q}^{X_S}$ whenever they were sending a message $m \in X_S(\tilde{\pi})$. As all messages in $X_i(\tilde{\pi})$ led to the same unaudited decision, the same decision must now be optimal for the principal with experiment $\tilde{\pi}_c$, as the tampering threshold corresponding to $\tilde{q}^{X_i}$ is the same as the threshold for $q \in X_i$. Thus, the designer's expected payoff from $\tilde{\pi}$ and $\tilde{\pi}_c$ coincide.

Finally, we construct an experiment $\tilde{\pi}_S \in \Pi_S$ that weakly improves upon $\tilde{\pi}_c$. First, we can obtain an improvement whenever $\tilde{q}^{X_S} > 0$ by lowering $\tilde{q}^{X_S}$–thus raising the probability of realization $\tilde{q}^{X_H}$–in a way that tampering incentives remain constant but this transformed experiment raises the designer's payoff by raising the probability of the favorable outcome $s = \tilde{q}^{X_H}$.

Second, we can improve upon $\pi = \{0, q\}$ whenever the unaudited posterior after $m = q$ exceeds $\underline{q}_H$. To see this, define $p \equiv \Pr[s = 0] = (q - \mu)/q$ and suppose that π induces a scale-up probability τ after an unaudited $m = q$. The expected gain from tampering is then $(1 - \lambda)\,\tau\,(v_H - v_S)$ and this establishes the tampering threshold $\bar{c}_S = (1 - \lambda)\,\tau\,(v_H - v_S)$. If $\tau > 0$, this requires that the principal's posterior after an unaudited $m = q$ must not fall below $\underline{q}_H$, so that Bayesian updating requires that

$$\frac{(1-p)\,q}{(1-p) + pF\left(\bar{c}_S\right)} \geq \underline{q}_H,$$

which, giving the Bayesian consistency constraint $(1-p)\,q = \mu$, leads to $p\bar{F}\left(\bar{c}_S\right)\underline{q}_H \geq \underline{q}_H - \mu$, and, using (1), can be expressed as

$$p\frac{\bar{F}\left(\bar{c}_S\right)}{p_S^C} \geq 1. \tag{A.3}$$

If this constraint is slack, then the unaudited posterior is strictly above $\underline{q}_H$ and the principal's sequential rationality implies that $\tau = 1$. But then, experiment $\{0, q - \varepsilon\}$ such that (A.3) is still slack (so that $\tau' = 1$) leads to the same tampering thresholds and decisions–implying that conditional on each realization the designer's expected utility has not changed–but the favorable outcome $s = q - \varepsilon$ is now more likely. Note that every status-quo experiment satisfies (A.3) with equality–this is also represented in (3). Therefore, any $\pi = \{0, q\}$ can be weakly improved upon by some status-quo experiment.

(ii) The proof of part i shows that there is always a status-quo experiment with a unique equilibrium that gives the designer a (weakly) higher payoff than any other experiment. This establishes $V^* = V_S(\mu; \lambda, k)$. ■

Proof of Lemma 1: For experiment $\pi = \{0, q\} \in \Pi_S$ with $p \equiv \Pr[s = 0]$, the designer's utility can be written as

$$\begin{aligned}
v_S(\tau(\bar{c}), \mu; \lambda, k) &= \lambda [(1-p) v_H + p v_S] \\
&\quad + (1-\lambda) \left[p\bar{F}(\bar{c}) v_S + \left(1 - p\bar{F}(\bar{c})\right)(\tau v_H + (1-\tau) v_S)\right]\} \\
&\quad - \mathbb{I}_{\{k=\mathcal{I}\}} p \int_0^{\bar{c}} \bar{F}(c)\, dc \\
&= v_H - (1-\lambda)(1-\tau)\Delta_S - p\left\{\lambda\Delta_S + \bar{c}\bar{F}(\bar{c}) - \mathbb{I}_{\{k=\mathcal{I}\}} \int_0^{\bar{c}} \bar{F}(c)\, dc\right\} \\
&= v_H - (1-\lambda)\Delta_S + \bar{c} - p\bar{F}(\bar{c})\left\{\frac{\lambda\Delta_S}{\bar{F}(\bar{c})} + \bar{c}\bar{F}(\bar{c}) - \mathbb{I}_{\{k=\mathcal{I}\}}\eta(\bar{c})\right\},
\end{aligned}$$

in which we have used (4) for $\bar{c}$ and the definition of $\eta(c)$ in (6). Noting from (3) that $p_S^C = \frac{\underline{q}_H - \mu}{\underline{q}_H}$ and that the Bayesian updating constraint (A.3) is satisfied with equality gives (7-8). ■

Proof of Proposition 3: Suppose that F admits a density so that $\eta(c)$ is differentiable. Using (7) and (8), the difference in the designer's marginal payoff from a higher tampering threshold $\bar{c}$ when moving from integration to separation is

$$\frac{\partial\left(v_S(\tau(\bar{c}), \mu; \mathcal{I}) - v_S(\tau(\bar{c}), \mu; \mathcal{S})\right)}{\partial \bar{c}} = \frac{\partial\left(m_S(\bar{c}; \mathcal{I}) - m_S(\bar{c}; \mathcal{S})\right)}{\partial \bar{c}}\left(\underline{q}_H - \mu\right) = -p_S^C \eta'(\bar{c}) \leq 0.$$

Therefore, the optimal tampering threshold under integration is lower than under separation, $\bar{c}_S^*(\lambda, \mathcal{I}) \leq \bar{c}_S^*(\lambda, \mathcal{S})$, which implies $q_S^*(\lambda, \mathcal{S}) \geq q_S^*(\lambda, \mathcal{I})$. ■

Proof of Proposition 4: (i) Consider experiment $\{0, q(\bar{c})\}$ inducing tampering threshold $\bar{c}$. Using (7) and (8), we have

$$\frac{\partial m_S(\bar{c}; \lambda, k)}{\partial \lambda} = \frac{\Delta_S}{\overline{F}(\bar{c})}\left(\frac{\mu - \underline{q}_H}{\underline{q}_H}\right) = -\frac{p_S^C \Delta_S}{\overline{F}(\bar{c})}, \; k \in \{\mathcal{S}, \mathcal{I}\},$$

implying

$$\frac{\partial v_S(\tau(\bar{c}), \mu; \lambda, k)}{\partial \lambda} = \Delta_S - p_S^C \frac{\Delta_i}{\overline{F}(\bar{c})} = \Delta_S \Pr[s = q(\bar{c})],$$

which is non-increasing in $\bar{c}$. Therefore, $\partial^2 v_S / \partial(-\lambda)\, \partial \bar{c} \geq 0$.

(ii) Define the feasible set of tampering thresholds

$$\mathcal{C}_S \equiv [0, (1-\lambda)\Delta_S] \cap [0, \overline{F}^{-1}(p_S^C / p_S^{FI})], \tag{A.4}$$

with $p_S^{FI} = \Pr[s = 1]$ for the fully informative experiment $\{0, 1\}$. To understand $\mathcal{C}_S$, note that $\bar{c} \in \mathcal{C}_S$ must satisfy two conditions. First, it must correspond to some scale-up probability

$\tau \in [0,1]$–from (4) this implies that $\bar{c} \in [0,(1-\lambda)\Delta_S]$. Second, the experiment $\{0, q(\bar{c})\}$ must be feasible–i.e., $q(\bar{c}) \leq 1$–which requires $\bar{F}(\bar{c})/p_S^C \geq \frac{1}{1-\mu} = 1/p_S^{FI}$ –see (3).

We can write the designer's problem in terms of selecting $\bar{c}^*$ that solves

$$\max_{\bar{c}} v_S(\tau(\bar{c}), \mu; \lambda, k), \; s.t. \; \bar{c} \in \mathcal{C}_S. \tag{A.5}$$

The feasible set $\mathcal{C}_S$ is increasing in the strong set order with respect to $-\lambda$ and, from part (i), $v_S(\tau(\bar{c}), \mu; \lambda, k)$ is supermodular in $(\bar{c}, -\lambda)$. Theorem 4' in Milgrom-Shannon (1994) then implies that the set of maximizers of (A.5) increases in the strong set order sense with $-\lambda$. From (3), for a fixed threshold $\bar{c}$ the experiment $\{0, q(\bar{c})\}$ is independent of λ, so the set of optimal experiments $q_S^*(\lambda, k)$ decreases in the strong set order sense with λ. ■

Proof of Proposition 5: Applying (11), the principal's equilibrium expected utility when organizing to innovate is

$$U(\lambda, k) = \underline{q}_H + \Pr[s = q_S^*(\lambda, k)]\, \lambda \left(q_S^*(\lambda, k) - \underline{q}_H\right).$$

Proposition 3 shows that $q_S^*(\lambda, \mathcal{S}) \geq q_S^*(\lambda, I)$ implying $U(\lambda, \mathcal{S}) \geq U(\lambda, \mathcal{I})$.

Note that for $\lambda = 1$ the designer always selects the commitment experiment, thus, regardless of the task-allocation, $q_S^*(1, k) = \underline{q}_H$. By the definition of designer's responsiveness to auditing, there exists $0 < \lambda < 1$ with $q_S^*(\lambda, k) > \underline{q}_H$ and $U(\lambda, k) > U(1, k)$. Therefore, $\lambda^* < 1$. Conversely, if $\lambda^* < 1$ then, for some $k-$allocation, $U(\lambda^*, k) > U(1, k)$ which implies $q_S^*(\lambda^*, k) > \underline{q}_H$ meaning that the designer is responsive to auditing.

We now show that if $f(0) > 0$ then the designer is responsive to auditing under separation. If $f(0) > 0$, then whenever $\lambda < 1$ the principal never approves without a conclusive audit if the designer selects $\{0, \underline{q}_H\}$. In other words, $\tau = 0$ for experiment $\{0, \underline{q}_H\}$ and

$$v_S(0, \mu; \lambda, k) = v_H - (1-\lambda)\Delta_S + \bar{c}_S(0) - m_S(\bar{c}_S(0); \lambda, k)\left(\underline{q}_H - \mu\right),$$

with $\bar{c}_S(0) = 0$. We now study conditions such that (a) there exists an experiment that leads to a positive scale-up probability, and (b) the designer's incremental payoff from an experiment that approves with positive probability is positive. These conditions ensure that the designer is responsive to auditing.

Consider first (a). The infimum tampering probability among experiments with $\tau > 0$ is $F[\bar{c}_i(0)] = F[0]$. The experiment that induces the highest posterior if unaudited is $\{0, 1\}$, and, for this experiment, $\Pr[s=0] = 1-\mu = p_S^{FI}$. Therefore, there exists an experiment with a positive scale-up probability, iff

$$\frac{p_S^C}{\overline{F}(0)} < p_S^{FI} \Longleftrightarrow \overline{F}(0) > \frac{p_S^C}{p_S^{FI}} \, (<1) \, .$$

If F admits a density at zero, this condition is always satisfied for any $\lambda \in [0,1]$.

Consider now (b). Noting from (4) that $\bar{c}_S(\tau) = \tau (1-\lambda) \Delta_S$, we can differentiate (7)—taking into account (8) and the definition of p_S^C in (1)—to obtain

$$\left.\frac{\partial v_S(\tau,\mu;\lambda,k)}{\partial \tau}\right|_{\tau=0} = (1-\lambda)\Delta_S \left[\left(1-p_S^C\right) - p_S^C \frac{f\left(\bar{c}_S(0)\right)}{\left(\overline{F}(\bar{c}_S(0))\right)^2} \lambda \Delta_S \right].$$

Since $\bar{c}_S(0) = 0$, the condition $\partial v_S(\tau,\mu;\lambda,k)/\partial\tau|_{\tau=0} > 0$ translates to

$$\frac{1-p_S^C}{p_S^C \Delta_S} > \lambda \frac{f(0)}{\left(\overline{F}(0)\right)^2},$$

and there is always a $0<\lambda<1$ that satisfies this condition. ■

Proof of Lemma 2: Setting $k = \mathcal{S}$ in (7) and (8) we have

$$\begin{aligned} v_S(\tau(\bar{c}),\mu;\lambda,\mathcal{S}) &= v_H - (1-\lambda)\Delta_S + \bar{c} - \Delta_S \frac{q_H - \mu}{\underline{q}_H - \underline{q}_S} \left(\frac{\lambda}{\overline{F}(\bar{c})} + \frac{\bar{c}}{\Delta_S} \right) \\ &= v_H - (1-\lambda)\Delta_S + \left(1-p_S^C\right)\bar{c} - \lambda \frac{p_S^C \Delta_S}{\overline{F}(\bar{c})}. \end{aligned}$$

Program (A.5) defines the designer's problem and the feasible set of tampering thresholds $\mathcal{C}_S = [0,(1-\lambda)\Delta_S] \cap [0, \overline{F}^{-1}(p_S^C/p_S^{FI})]$ is defined in (A.4). Whenever it exists, the marginal payoff from increased tampering is

$$\frac{\partial v_S(\tau(\bar{c}),\mu;\lambda,\mathcal{S})}{\partial \bar{c}} = \left(1-p_S^C\right) - \lambda p_S^C \Delta_S \frac{f(\bar{c})}{\left(\overline{F}(\bar{c})\right)^2} = \lambda p_S^C \Delta_S \left(\frac{\phi_S}{\lambda} - L(\bar{c}) \right).$$

The single-crossing condition implies that $v_S(\tau(\bar{c}),\mu;\lambda,\mathcal{S})$ is quasiconcave in $\bar{c}$. Suppose first that $\partial v_S(\tau(0),\mu;\lambda,\mathcal{S})/\partial\bar{c} = \lambda p_S^C \Delta_S ((\phi_S/\lambda) - L(0)) \leq 0$, implying $\partial v_S(\tau(\bar{c}),\mu;\lambda,\mathcal{S})/\partial\bar{c} \leq 0$, for $\bar{c} \geq 0$. In this case, we have $\bar{c}^* = 0$, and the designer selects the commitment experiment $\left\{0, \underline{q}_H\right\}$. Suppose now that $\lambda p_S^C \Delta_S ((\phi_S/\lambda) - L(0)) > 0$, and let $\bar{c}_{crit}$ be the minimum threshold that satisfies $\partial v_S(\tau(\bar{c}_{crit}),\mu;\lambda,\mathcal{S})/\partial\bar{c} = 0$ (and set $\bar{c}_{crit} = \infty$ if no such threshold exists). Then, the solution to the designer's problem satisfies

$$\bar{c}^*(\lambda) = \min\left[\bar{c}_{crit}(\lambda), (1-\lambda)\Delta_S, \bar{c}_{FI}\right].$$

■

Proof of Proposition 6: (i) For a fixed F and λ, suppose that the designer under separation selects $\pi = \{0, q\} \in \Pi_S$, with $p = \Pr[s = 0] = (q - \mu)/q$.[33] We first show that there is an upper bound on λ that does not depend on F:

$$\lambda \leq \frac{1 - p_S^C}{1 - p_S^C + p - p_S^C} \equiv \widetilde{\lambda}_S(p). \tag{A.6}$$

To see this, we express $v_S(\tau(p), \mu; \lambda, \mathcal{S})$ as a function of p: using (3) we have $\bar{F}(\overline{c}_S(p)) = p_S^C/p$ and we can write (7-8) for $k = \mathcal{S}$ as

$$\begin{aligned} v_S(\tau(p), \mu; \lambda, \mathcal{S}) &= v_H - (1 - \lambda)\Delta_S + \overline{c}_S(p) - p_S^C \left(\frac{\lambda \Delta_S}{\overline{F}(\overline{c}_S(p))} + \overline{c}_S(p) \right) \\ &= v_S + \lambda \Delta_S + \overline{c}_S(p) - \lambda \frac{p_S^C \Delta_S}{\overline{F}(\overline{c}_S(p))} - p_S^C \overline{c}_S(p) \\ &= v_S + \lambda \Delta_S (1 - p) + \left(1 - p_S^C\right) \overline{F}^{-1}(p_S^C/p). \end{aligned}$$

Designer's optimality of π_S requires

$$\begin{aligned} \left(\overline{F}(\overline{c}_S(p)) =\right) \frac{p_S^C}{p} &\leq \overline{F}((1 - \lambda)\Delta_S), \\ v_S(\tau(p'), \mu; \lambda, \mathcal{S}) &\leq v_S(\tau(p), \mu; \lambda, \mathcal{S}) \text{ for } p' \in \left[p_S^C, p\right]. \end{aligned}$$

The first condition follows from $\overline{c}_S(p) \leq (1 - \lambda)\Delta_S$, as the gain from tampering is bounded by $(1 - \lambda)\Delta_S$, while the second is the designer's incentive compatibility constraint when comparing π to status-quo experiments that are less informative than π.[34] Setting $p' = p_S^C$ above, and obviating the common term v_S, incentive compatibility implies

$$\lambda \Delta_S \left(1 - p_S^C\right) \leq \lambda \Delta_S (1 - p) + \left(1 - p_S^C\right) \overline{F}^{-1}(p_S^C/p) \leq \lambda \Delta_S (1 - p) + \left(1 - p_S^C\right)(1 - \lambda)\Delta_S,$$

from which we obtain (A.6). Inverting (A.6), any experiment that can be implemented with auditing intensity $\lambda > 1/(2 - \underline{q}_H)$ must satisfy

$$p \leq 2p_S^C - 1 + \frac{1 - p_S^C}{\lambda} \equiv p(\lambda)$$

[33] Recall that, regardless of the cost distribution, the principal prefers to separate tasks when organizing to innovate–see Proposition 3.

[34] If $\tau(p) = 1$, the designer cannot improve scale-up probability by switching to a status-quo experiment that is more informative than $\pi = \{0, q\}$ so that trivially $v_S(\tau(p'), \mu; \lambda, \mathcal{S}) \leq v_S(\tau(p), \mu; \lambda, \mathcal{S})$ for $p' > p$. When implementing $\pi(p)$ with auditing intensity $\widetilde{\lambda}_S(p)$ we will look at cost distributions for which $\tau(p) = 1$.

Taking into account $q = \mu/(1-p)$, and after some calculations, this implies that the most informative experiment implementable with auditing λ is

$$q(\lambda) = \frac{\lambda}{2\lambda - 1}\underline{q}_H.$$

(ii) We now derive the cost distributions that would lead the designer to select $\pi_S(\lambda) = \{0, q(\lambda)\}$, with $\Pr[s=0] = p(\lambda)$, when auditing is λ. Suppose that experiment $\pi_S(\lambda)$ leads to the principal's rubberstamping, $\tau = 1$, so that $\bar{c}_S = \overline{F}^{-1}(p_S^C/p(\lambda)) = (1-\lambda)\Delta_S$. Incentive compatibility requires that for any $p' \in \left[p_S^C, p(\lambda)\right]$,

$$\left(1-p_S^C\right)\overline{F}^{-1}(p_S^C/p') \leq \lambda\Delta_S\left(p'-p\right) + \left(1-p_S^C\right)\left(1-\lambda\right)\Delta_S.$$

Using $\overline{F}^{-1}(p_S^C/p') = F^{-1}(\left(p'-p_S^C\right)/p')$ and simplifying we have

$$\frac{p'-p_S^C}{p'} \leq F\left(\lambda\Delta_S\frac{p'-p_S^C}{1-p_S^C}\right).$$

Alternatively, letting $c = \lambda\Delta_S\left(p'-p_S^C\right)/\left(1-p_S^C\right)$, we have

$$F\left(c\right) \geq \frac{c}{c+\lambda\frac{p_S^C\Delta_S}{1-p_S^C}} = \frac{c}{c+\frac{p_S^C\Delta_S}{1-2p_S^C+p(\lambda)}} \text{ for } c \leq \lambda\Delta_S\frac{p(\lambda)-p_S^C}{1-p_S^C}. \tag{A.7}$$

That is, the likelihood of low tampering costs must be sufficiently high to allow the principal to approve with low probability if the experiment is not very informative. Note that our argument didn't require the distribution to be smooth or to have a density. One distribution that satisfies (A.7) is supported only on two cost realizations, 0 and $\lambda\Delta_S\frac{p(\lambda)-p_S^C}{1-p_S^C}$, with

$$\Pr\left[c=0\right] = \left(p(\lambda)-p_S^C\right)/p(\lambda), \tag{A.8}$$

and, in equilibrium, the agent only tampers if $c = 0$ so that expected tampering costs are zero. ■

Proof of Proposition 7: From (11), for each $\pi_S(\lambda) = \{0, q(\lambda)\}$, with $p(\lambda) = \Pr[s=0]$, auditing λ, and cost distribution satisfying (A.7), the principal's utility is

$$U\left(\pi_S(q)\right) = \underline{q}_H + (1-p)\,\lambda\left(q(\lambda)-\underline{q}_H\right) = \underline{q}_H + \frac{\left(q(\lambda)-\underline{q}_H\right)\mu}{2q-\underline{q}_H},$$

which is increasing in $q(\lambda)$. Thus, the principal optimally sets $q(\lambda_{opt}) = 1$ which requires $\lambda_{opt} = 1/(2-\underline{q}_H)$. Setting $p(\lambda) = (1-\mu)$ in (A.8), we obtain $\Pr\left[c=0\right] = \mu(1-$

$\underline{q}_H)/(\underline{q}_H(1-\mu))$ for the cost distribution supported on 0 and $\widetilde{c} = \lambda_{opt}\Delta_S \frac{1-\mu-p_S^C}{1-\mu-p_S^C} = (1-\underline{q}_H)/(2-\underline{q}_H)$; this distribution minimizes tampering costs among all those inducing a fully informative experiment. As this distribution induces zero costs on the agent, the designer under integration and separation would select the same experiment. ■

Proof of Proposition 8: (i) Let Π^Σ be the set of experiments with outcomes $\left\{\underline{q}_L, \underline{q}_S, q\right\}$, indexed by (p_L, p_S, p) with $p_L, p_S, p \geq 0$ and $p_L + p_S = 1$, and defined as follows:

$$\Pi^\Sigma \equiv \left\{\left\{\underline{q}_L, \underline{q}_S, q\right\} : q \geq \underline{q}_H \; and \Pr[q] = 1-p, \Pr\left[\underline{q}_L\right] = p * p_L, \Pr\left[\underline{q}_S\right] = p * p_S\right\},$$

and equilibrium decision making

$$d_U\left(\underline{q}_i\right) = d_I\left(m, \underline{q}_i\right) = d_i; d_I(m,q) = d_H; d_U(q) = \tau d_H + (1-\tau)d_S,$$

and note that $\Pi^R \subset \Pi^\Sigma$.

Consider an arbitrary finite experiment $\tilde{\pi} = \{q, \Pr[q]\}_{q \in S(\tilde{\pi})}$ and suppose that players follow a PBE of the communication subgame in which the principal, if indifferent after a conclusive audit–which only applies if either $\underline{q}_S \in S(\tilde{\pi})$ or $\underline{q}_H \in S(\tilde{\pi})$–always selects the agents' preferred decision–i.e., $d_I^*(m, \underline{q}_i) = d_i \; i = \{S, H\}$. We show that there exists $\tilde{\pi}^R \in \Pi^R$ that (weakly) improves the designer's payoff relative to $\tilde{\pi}$. Therefore, if $\pi^* \in \Pi^R$ maximizes the designer's payoff when restricted to Π^R, then selecting π^* is a PBE of the design subgame, as the designer's expected utility cannot be improved by any alternative $\tilde{\pi}$.

We proceed in two steps. In step 1 we derive an experiment $\tilde{\pi}^\Sigma \in \Pi^\Sigma$ that improves the designer's payoff relative to $\tilde{\pi}$. In step 2, we show that the designer's maximum expected payoff in Π^Σ is achieved by an experiment in Π^R.

Step 1: Let $S_T(\tilde{\pi})$ be the set of tampering types as in (A.1) and recall that, from Proposition 1, $M_T(\tilde{\pi})$ is the set of tampered outcomes. Proposition 1 shows that, if $d_I^*(m, \underline{q}_i) = d_i$ whenever $\underline{q}_i \in S(\pi), i = \{S, H\}$, then $S_T(\tilde{\pi}) \cap M_T(\tilde{\pi}) = \emptyset$.

Partition $S(\tilde{\pi})$ by defining $X_L(\tilde{\pi}) = S(\tilde{\pi}) \cap (\underline{q}_L, \underline{q}_S)$, $X_S(\tilde{\pi}) = S(\tilde{\pi}) \cap \left(\underline{q}_S, \underline{q}_H\right)$ and $X_H(\tilde{\pi}) = S(\tilde{\pi}) \cap [\underline{q}_H, 1]$. Following the same logic as in the proof of Proposition 2 we can show that tampering types correspond to low realizations while tampered outcomes are associated with high realizations of the experiment, i.e.,

$$q_{S_T} \equiv \max\{q : q \in S_T(\tilde{\pi})\} < \min\{q : q \in M_T(\tilde{\pi})\} \equiv q_{M_T}. \quad \text{(A.9)}$$

and that (A.9) implies that all messages in $X_i(\tilde{\pi})$ lead to the same unaudited (mixture over) decision(s).

We now construct $\tilde{\pi}_c$ that has an equilibrium that gives the designer the same expected utility as the equilibrium of $\tilde{\pi}$. We do so by replacing all realizations in $X_i(\tilde{\pi})$, $i = L, S, H$, with a realization $s = \tilde{q}^{X_i}$ that is its conditional expectation, i.e.,

$$\tilde{q}^{X_i} = \frac{\sum_{q \in X_i(\tilde{\pi})} \Pr[q]\, q}{\sum_{q \in X_i(\tilde{\pi})} \Pr[q]}, \;\; \Pr\left[\tilde{q}^{X_i}\right] = \sum_{q \in X_i(\tilde{\pi})} \Pr[q],$$

and adjusting the equilibrium (mixture over) messages to

$$m_c(\tilde{q}^{X_i}, c) = \frac{\sum_{q \in X_i(\tilde{\pi})} \Pr[q] \sum_{j=\{L,S,H\}} \sum_{q' \in X_j(\tilde{\pi})} \Pr[m(q,c) = q']\, \tilde{q}^{X_j}}{\sum_{q \in X_i(\tilde{\pi})} \Pr[q]}.$$

We complement the definition by having threshold types $\underline{q}_i$ send message $m = \tilde{q}^{X_j}$ whenever they were sending a message $m \in X_j(\tilde{\pi})$. As all messages in $X_i(\tilde{\pi})$ led to the same unaudited decision, the same decision must now be optimal for the principal with experiment $\tilde{\pi}_c$, as the tampering threshold corresponding to $\tilde{q}^{X_i}$ is the same as the threshold for $q \in X_i$. Thus, the designer's expected payoff from $\tilde{\pi}$ and $\tilde{\pi}_c$ coincide.

To obtain an improvement within the set Π^{Σ}, suppose that $X_L(\tilde{\pi})$ or $X_S(\tilde{\pi})$ is non-empty–otherwise $\tilde{\pi}_c \in \Pi^{\Sigma}$. By either lowering $\tilde{q}^{X_L} > \underline{q}_L$ or $\tilde{q}^{X_S} > \underline{q}_S$, we can raise the probability of realization $\tilde{q}^{X_H}$ in a way that tampering incentives remain constant but this transformed experiment raises the designer's payoff by raising the probability of the favorable outcome $s = \tilde{q}^{X_H}$. Therefore, if $X_L(\tilde{\pi})$ or $X_S(\tilde{\pi})$ is non-empty, there is an experiment in Π^{Σ} that gives the designer a higher payoff.

Step 2: Let $\tilde{\pi}^{\Sigma} \in \Pi^{\Sigma}$, described by (p_L, p_S, p) with associated scale-up probability τ after an unaudited $m = q$. As the principal will select the agents' preferred decision after a conclusive audit reveals $s = \underline{q}_i$, the expected gain from tampering is $(1-\lambda)(\tau v_H + (1-\tau) v_S - v_i)$ and this establishes the tampering threshold

$$\bar{c}_i \equiv \bar{c}(\underline{q}_i) = (1-\lambda)(\tau v_H + (1-\tau) v_S - v_i) = (1-\lambda)\left[\tau \Delta_S + (\Delta_i - \Delta_S)\right]. \qquad \text{(A.10)}$$

Suppose first that $\tau > 0$. This requires that the principal's posterior after an unaudited $m = q$ must not fall below $\underline{q}_H$, so that Bayesian updating requires that

$$\frac{(1-p)\, q + p \sum_{i=L,S} p_i F(\bar{c}_i)\, q_i}{(1-p) + p \sum_{i=L,S} p_i F(\bar{c}_i)} \geq \underline{q}_H,$$

which, giving the Bayesian consistency constraint $(1-p)\,q=\mu-p\sum_{i=L,S}p_iq_i$, leads to

$$\begin{aligned}
\mu-p\sum_{i=L,S}p_iq_i+p\sum_{i=L,S}p_iF\left(\bar{c}_i\right)q_i &\geq \underline{q}_H\left((1-p)+p\sum_{i=L,S}p_iF\left(\bar{c}_i\right)\right)\\
p\left(\underline{q}_H\left(1-\sum_{i=L,S}p_iF\left(\bar{c}_i\right)\right)-\sum_{i=L,S}p_i\left(1-F\left(\bar{c}_i\right)\right)q_i\right) &\geq \underline{q}_H-\mu,\\
p\left(\sum_{i=L,S}p_i\bar{F}\left(\bar{c}_i\right)\left(\underline{q}_H-q_i\right)\right) &\geq \underline{q}_H-\mu,
\end{aligned}$$

which, using (1) can be expressed as

$$p\sum_{i=L,S}p_i\frac{\bar{F}\left(\bar{c}_i\right)}{p_i^C}\geq 1. \tag{A.11}$$

Note that if this constraint is slack, then the unaudited posterior is strictly above $\underline{q}_H$ and the principal's sequential rationality implies that $\tau=1$. But then, experiment $\tilde{\pi}^\Sigma$ cannot be optimal for the designer. Indeed, consider $\pi'\in\Pi^\Sigma$, described by (p'_L,p'_S,p'), that differs from $\tilde{\pi}^\Sigma$ only in that $p'<p$, while $p'_i=p_i$, but such that the constraint (A.11) is still slack (so that $\tau'=1$). As tampering thresholds and decisions have not changed, conditional on each realization the designer's expected utility has not changed, but the favorable outcome $s=q$ is now more likely, thus reaching a contradiction. Thus, in a designer's optimal experiment in Π^Σ, (A.11) must hold with equality.

We now show that the maximum expected utility of the designer is attained on the boundary of Π^Σ, i.e., by setting $p_i=0$ or $p_i=1$. But these are precisely the experiments in Π.

Let

$$W_i(\bar{c}_i;\lambda,k)=\lambda\Delta_i+\bar{c}_i\bar{F}\left(\bar{c}_i\right)-\mathbb{I}_{\{k=\mathcal{I}\}}\int_0^{\bar{c}_i}\bar{F}\left(c\right)dc. \tag{A.12}$$

For experiment $\tilde{\pi}^{\Sigma}$, the designer's utility can be written as

$$\begin{aligned}
V\left(\tilde{\pi}^{\Sigma};\lambda,k\right) &= \sum_{i=L,S} p_i\{\lambda\left[(1-p)\,v_H+pv_i\right] \\
&\quad +(1-\lambda)\left[p\bar{F}\left(\bar{c}_i\right)v_i+\left(1-p\bar{F}\left(\bar{c}_i\right)\right)\left(\tau v_H+(1-\tau)\,v_S\right)\right]\} \\
&\quad -\mathbb{I}_{\{k=\mathcal{I}\}}p\sum_{i=L,S}p_i\int_0^{\bar{c}_i}\bar{F}\left(c\right)dc \\
&= \sum_{i=L,S}p_i\left\{v_i+\lambda(1-p)\Delta_i+\left(1-p\bar{F}\left(\bar{c}_i\right)\right)\bar{c}_i-\mathbb{I}_{\{k=\mathcal{I}\}}p\int_0^{\bar{c}_i}\bar{F}\left(c\right)dc\right\} \\
&= v_H-(1-\lambda)\,(1-\tau)\,\Delta_S-p\sum_{i=L,S}p_i\left\{\lambda\Delta_i+\bar{c}_i\bar{F}\left(\bar{c}_i\right)-\mathbb{I}_{\{k=\mathcal{I}\}}\int_0^{\bar{c}_i}\bar{F}\left(c\right)dc\right\} \\
&= v_H-(1-\lambda)\,(1-\tau)\,\Delta_S-p\sum_{i=L,S}p_iW_i(\bar{c}_i;\lambda,k). \qquad \text{(A.13)}
\end{aligned}$$

in which we have used (A.10) for $\bar{c}_i$. Fix an scale-up probability $\tau>1$–which also determines the tampering thresholds $\bar{c}_i$, see (A.10)–and consider the optimal $\hat{\pi}\in\Pi^{\Sigma}$ that maximizes $V\left(\pi;\lambda,k\right)$ among experiments in Π^{Σ} with scale-up probability τ. Then, replacing p with the binding Bayesian updating constraint (A.11), experiment $\hat{\pi}$ solves

$$\max_{(p_L,p_S)}\frac{\sum_{i=L,S}p_iW_i(\bar{c}_i;\lambda,k)}{\sum_{i=L,S}p_i\frac{\bar{F}(\bar{c}_i)}{p_i^C}},\text{s.t. }p_L+p_S=1,$$

which is a quasiconcave program as the objective function is quasilinear—since it is the ratio of linear functionals—and the constraint set is convex (see Boyd and Vandenberghe 2004). Then, there is always an extreme point of the simplex that solves this program. In other words, there is always a robust experiment that maximizes the designer's payoff when selecting experiments in Π^{Σ}. This concludes the proof of part i of the Proposition.

(ii) Note that the payoff from any experiment in Π^R can be approximated by experiments of the form $\left\{\underline{q}_i+\epsilon,q\right\}$ which have a unique communication equilibrium–i.e., we can express $\bar{V}(\mu;\lambda,k)$ as $\bar{V}(\mu;\lambda,k)=\sup V\left(\pi;\lambda,k\right),\text{s.t.},\pi=\left\{\underline{q}_i+\epsilon,q\right\},\epsilon>0,i\in\{L,S\}$. Thus, we must have $V^*\geq\bar{V}(\mu;\lambda,k)$. The proof of part 1 showed that for any experiment $\tilde{\pi}$ such that the principal is never indifferent after a conclusive audit there is a robust experiment that gives the designer a (weakly) higher payoff. Since $V^*>\bar{V}(\mu;\lambda,k)$ implies that the principal must be indifferent after a conclusive audit, then we must have $\underline{q}_S\in S(\pi^*)$ for any equilibrium experiment π^* that yields a payoff $V^*>\bar{V}(\mu;\lambda,k)$. ■

Proof of Lemma 3: Setting $p_S = 0$ and $p_L = 0$ in (A.13) and using (A.12) we obtain $v_i(\tau(\bar{c}), \mu; \lambda, k)$ for $i \in \{L, S\}$. Noting from (19) that $\Pr\left[s = \underline{q}_i\right] = \frac{q-\mu}{q-\underline{q}_i}$ gives (21-22). ■

Proof of Lemma 4: Note first that if the designer selects an "up-or-down" experiment when $\lambda = 1$, so that $p_L^C \Delta_L < p_S^C \Delta_S$, then we must have $\phi_L > \phi_S$. Indeed, since $1 - p_S^C < 1 - p_L^C$ then

$$p_L^C \Delta_L < p_S^C \Delta_S \Rightarrow \frac{p_S^C \Delta_S}{p_L^C \Delta_L} > 1 > \frac{1 - p_S^C}{1 - p_L^C} \Rightarrow \phi_L > \phi_S.$$

We now show that if $L(c) - (\phi_i/\lambda)$ is single-crossing in $[0, \Delta_i]$ and $\phi_L > \phi_S$ then $\bar{c}_S^*(\lambda, \mathcal{S}) \leq \bar{c}_L^*(\lambda, \mathcal{S})$; that is, the tampering threshold under separation is larger when the designer's experiment is restricted to Π_L^R rather than restricted to Π_S. Then, (24) implies that the principal's is weakly better-off when the designer is restricted to Π_L^R rather than Π_S^R.

Maximizing over $\bar{c}$ for $k = \mathcal{S}$ in (21-22) for each $i \in \{L, S\}$ and recalling that $\bar{F}^{-1}\left(p_i^C / p_i^{FI}\right)$ is the threshold induced by experiment $\{\underline{q}_i, 1\}$ we can write–see lemma 2,

$$\bar{c}_i^*(\lambda, \mathcal{S}) = \min\left[L^{-1}(\phi_i/\lambda), (1 - \lambda)\Delta_i, \bar{F}^{-1}\left(p_i^C / p_i^{FI}\right)\right].$$

Since $p_i^C / p_i^{FI} = \frac{q_H - \mu}{q_H - \underline{q}_i} \frac{1 - \underline{q}_i}{1 - \mu}$ increases in $\underline{q}_i$, we have $\bar{F}^{-1}\left(p_S^C / (1 - \mu)\right) < \bar{F}^{-1}\left(p_L^C / (1 - \mu)\right)$. Second, if $L(c) - (\phi_i/\lambda)$ is single crossing and $\phi_L > \phi_S$, we must have $L^{-1}(\phi_S/\lambda) < L^{-1}(\phi_L/\lambda)$. Combining both observations with $\Delta_S < \Delta_L$, we must have $\bar{c}_S^*(\lambda, \mathcal{S}) \leq \bar{c}_L^*(\lambda, \mathcal{S})$. ■

Proof of Proposition 9: (i) We can adapt the proof of Proposition 3 to show that the principal prefers to separate tasks if task allocation does not change the optimal class of the designer's experiment (i.e., if $i^*(\mathcal{S}, \lambda) = i^*(\mathcal{I}, \lambda)$). Therefore, for integration to be optimal, we must have an adverse switch so that $i^*(\mathcal{S}, \lambda) = S$ and $i^*(\mathcal{I}, \lambda) = L$. We show that if $W(\Delta_S, (1 - \lambda)\Delta_S) > W(\Delta_L, (1 - \lambda)\Delta_L)$ then we can find a range of parameters so that the principal strictly prefers to integrate tasks. To economize on notation, let

$$T(\Delta_i, \tau) \equiv \frac{\lambda \Delta_i}{\overline{F}(\bar{c}_i(\tau))} + \bar{c}_i(\tau),$$

so that $W(\Delta_i, \bar{c})$ defined in (25) simplifies to $W(\Delta_i, \bar{c}) = \eta(\bar{c}) / T(\Delta_i, \tau(\bar{c}))$.

We first derive a sufficient condition for adverse switches when restricted to high scale-up probabilities, i.e., such that for all $\tau \geq \tilde{\tau}$ we have

$$v_S(\tau, \mu; \lambda, \mathcal{S}) \geq v_L(\tau, \mu; \lambda, \mathcal{S}) \text{ and } v_S(\tau, \mu; \lambda, \mathcal{I}) \leq v_L(\tau, \mu; \lambda, \mathcal{I}).$$

Using (21), (22), and the definition of $\bar{c}_i(\tau)$ in (20) these two conditions translate to

$$\begin{aligned} p_S^C T(\Delta_S,\tau) &\leq p_L^C T(\Delta_L,\tau), \text{ and} \\ p_S^C \left(T(\Delta_S,\tau)+\eta\left(\bar{c}_S(\tau)\right)\right) &\geq p_L^C \left(T(\Delta_L,\tau)+\eta\left(\bar{c}_L(\tau)\right)\right), \end{aligned}$$

which simplifies to

$$\frac{T(\Delta_S,\tau)+\eta\left(\bar{c}_S(\tau)\right)}{T(\Delta_L,\tau)+\eta\left(\bar{c}_L(\tau)\right)} \geq \frac{p_L^C}{p_S^C} \geq \frac{T(\Delta_S,\tau)}{T(\Delta_L,\tau)}. \tag{A.14}$$

Note that a necessary condition for (A.14) is that $T(\Delta_L,\tau)\,\eta\left(\bar{c}_S(\tau)\right) \geq T(\Delta_S,\tau)\,\eta\left(\bar{c}_L(\tau)\right)$, which is implied by

$$W\left(\Delta_S,\bar{c}_S(\tau)\right) \geq W\left(\Delta_L,\bar{c}_L(\tau)\right). \tag{A.15}$$

The condition $W(\Delta_S,(1-\lambda)\Delta_S) > W(\Delta_L,(1-\lambda)\Delta_L)$ is equivalent to (A.15) setting $\tau = 1$. Continuity of W and $\bar{c}_i(\tau)$ implies that there is $\tilde{\tau} < 1$, so that (A.15) is satisfied for $\tau > \tilde{\tau}$. In fact, since $p_L^C/p_S^C = \left(\underline{q}_H - \underline{q}_S\right)/\left(\underline{q}_H - \underline{q}_L\right)$ and W does not depend on the prior μ, then we can find $\underline{q}_S$ with $\underline{q}_L < \underline{q}_S < \underline{q}_H$ so that p_L^C/p_S^C is closed to 1 and (A.14) holds for $\tau > \tilde{\tau}$.

Finally, we have that both under separation and integration, the designer's optimal experiment tends to the maximum scale-up probability as $\mu \to \underline{q}_H$, i.e., $\tau\left(\bar{c}_i^*(\lambda,k)\right) \to 1$ as $\mu \to \underline{q}_H$. Then we can find μ, with $\underline{q}_S < \mu < \underline{q}_H$ so that the principal's optimal experiment satisfies $\tau\left(\bar{c}_i^*(\lambda,k)\right) > \tilde{\tau}$, $k = \mathcal{S},\mathcal{I}$. This implies that we have an adverse switch for λ: the designer under separation would select $i^*(\mathcal{S},\lambda) = S$ but under integration he would select $i^*(\mathcal{I},\lambda) = L$. For large α_L, so that "up-or-down" experiments are more valuable to the principal, then she would optimally integrate tasks.

(iii) If $p_L^C > (\Delta_S/\Delta_L)\,p_L^C$, then setting $\lambda = 1$ leads to experiment $\{\underline{q}_S, \underline{q}_H\}$ regardless of task allocation and the principal derives no surplus. Then, Proposition 5 shows that if $f(0) > 0$, then $\lambda^* < 1$.[35] If $p_L^C < (\Delta_S/\Delta_L)\,p_L^C$, however, the designer selects $\left\{\underline{q}_L, \underline{q}_H\right\}$ if $\lambda = 1$ and setting $\lambda < 1$ may trigger an adverse switch. If, however, the designer under a $k-$allocation is sensitive to auditing for some λ' and $i^*(k,\lambda') = L$, then

$$\lambda' \frac{F\left(\bar{c}_S^*(\lambda';i^*)\right)}{\bar{F}\left(\bar{c}_S^*(\lambda';i^*)\right)} > 0 \left(= \frac{F(0)}{\bar{F}(0)}\right)$$

and (24) shows that the principal's utility increases when reducing λ from 1 to λ'. ■

[35] As shown in the Online Appendix B, the designer will always select a status-quo experiment if he does so when $\lambda = 1$. Therefore, the principal can only gain from an imperfect audit if the designer selects a more informative status-quo experiment.

Proof of Corollary 1: Suppose that the designer is responsive to auditing but $\lambda^* = 1$. Then, from Proposition 9-ii we must have that $p_L^C < (\Delta_S/\Delta_L)\, p_S^C$ but for each λ so that he prefers some experiment $\{\underline{q}_L, q\}$, $q > \underline{q}_H$ to $\{\underline{q}_L, \underline{q}_H\}$ it must be that $i^*(k, \lambda) = S$. Since $p_L^C < (\Delta_S/\Delta_L)\, p_S^C$, then Lemma 4 implies that for any λ the designer's choice under separation improves the principal's utility, and she cannot be worse-off by committing to rule out the status-quo. But in this case, the principal's problem converts to a situation in which she organizes to innovate (as she selects from two decisions, d_L and d_H) and Proposition 9-ii implies $\lambda^* < 1$. ■

Proof of Proposition 10: Suppose first that $p_L^C < (\Delta_S/\Delta_L)\, p_S^C$, so that the designer selects $\{\underline{q}_L, \underline{q}_H\} = \{0, \underline{q}_H\}$ if $\lambda = 1$. We prove the optimality of (17) and (18) in two steps. First, we show that the cost distributions (16) satisfy the designer's incentive compatibility for experiments $\{0, q(\lambda)\} \in \Pi_L^R$ which now must also account for the possibility of switching to an status-quo experiment. Second, after deriving the optimal auditing and cost distribution, we show that the principal cannot do better by instead inducing the designer to select a status-quo experiment.

Consider $\pi_L(p) = \{\underline{q}_L, q\}\ (= \{0, q\})$ with $p = \Pr\left[s = \underline{q}_L\right] = (q - \mu)/(q - \underline{q}_L)$. Then, by a similar reasoning as in the proof of Proposition 6, if the designer under separation selects π_L when auditing is λ then we must have

$$\lambda \leq \frac{1 - p_L^C}{1 - p_L^C + p - p_L^C} \equiv \widetilde{\lambda}_L(p).$$

Consider a cost distribution supported on 0 and $\bar{c}^* \equiv \widetilde{\lambda}_L(p)\Delta_L \frac{p - p_L^C}{1 - p_L^C}$, with

$$\Pr\left[c = 0\right] = \left(p - p_L^C\right)/p. \tag{A.16}$$

The argument in the proof of Proposition 6 showed that if the auditing intensity is $\widetilde{\lambda}_L(p)$, the designer, when restricted to Π_L^R, selects experiment $\pi_L(p)$. We now show that the designer does not wish to switch and select instead $\{\underline{q}_S, q'\} \in \Pi_S^R$. Using (A.16) we first note that any $\{\underline{q}_S, q'\}$ that leads to a positive scale-up probability must satisfy

$$\Pr\left[s = \underline{q}_S\right] \geq \frac{p_S^C}{1 - \Pr\left[c = 0\right]} = p\frac{p_S^C}{p_L^C},$$

as the agent always tampers when $c = 0$. Moreover, all robust experiments satisfying this condition lead to the principal rubberstamping the agent's recommendation. Recalling that

$v_L = 0$, experiments in Π_S are dominated by $\pi_L(p)$ iff

$$\begin{aligned} & v_S + \lambda(1 - \Pr[s = \underline{q}_S])\Delta_S + (1-\lambda)\left(1 - p_S^C\right)\Delta_S \\ \leq\; & \lambda(1-p)\Delta_L + (1-\lambda)\left(1 - p_L^C\right)\Delta_L. \end{aligned}$$

Rearranging and noting that $v_S + \Delta_S - \Delta_L = v_L = 0$, then incentive compatibility requires

$$0 \leq \lambda \Pr[s = \underline{q}_S]\Delta_S - \lambda p \Delta_L + (1-\lambda)\, p_L^C \Delta_L - (1-\lambda)\, p_L^C \Delta_L.$$

Since we assumed that $p_L^C \Delta_L < p_S^C \Delta_S$, then

$$\begin{aligned} & \lambda \Pr\left[s = \underline{q}_S\right]\Delta_S - \lambda p \Delta_L + (1-\lambda)\, p_S^C \Delta_S - (1-\lambda)\, p_L^C \Delta_L \\ \geq\; & \lambda p \frac{p_S^C}{p_L^C}\Delta_S - \lambda p \Delta_L + (1-\lambda)\, p_L^C \Delta_L - (1-\lambda)\, p_L^C \Delta_L \\ =\; & \left(\frac{\lambda p}{p_L^C} + (1-\lambda)\right)\left(p_S^C \Delta_S - p_L^C \Delta_L\right) \geq 0. \end{aligned}$$

Therefore, the designer subject to auditing $\widetilde{\lambda}_L(p)$ and the two-point cost distribution above would optimally select $\pi_L(p)$ with $\Pr[s = \underline{q}_L] = p$.

From (24), the principal's utility is

$$\begin{aligned} U\left(\pi_L(p)\right) &= \underline{q}_H + \left(\underline{q}_H - \mu\right)\left(\widetilde{\lambda}_L(p)\frac{p - p_L^C}{p_L^C} + \left(1 + \widetilde{\lambda}_L(p)\frac{p - p_L^C}{p_L^C}\right)\left(\frac{\alpha_L - \underline{q}_H}{\underline{q}_H - \underline{q}_L}\right)\right) \\ &= \underline{q}_H + \left(\underline{q}_H - \mu\right)\left(\frac{1 - p_L^C}{p_L^C}\frac{p - p_L^C}{1 - 2p_L^C + p} + \left(1 + \frac{1 - p_L^C}{p_L^C}\frac{p - p_L^C}{1 - 2p_L^C + p}\right)\left(\frac{\alpha_L - \underline{q}_H}{\underline{q}_H - \underline{q}_L}\right)\right) \end{aligned}$$

where we have used $F\left(\bar{c}^*\right)/\bar{F}\left(\bar{c}^*\right) = \Pr[c = 0]/\left(1 - \Pr[c = 0]\right) = \left(p - p_L^C\right)/p_L^C$. This expression is increasing in p, so that the principal sets $p = 1 - \mu$ —and the designer selects in response a fully informative experiment—implying that $\lambda_{opt} = \widetilde{\lambda}_L(1-\mu) = 1/\left(2 - \underline{q}_H\right)$.

To end the case when $p_L^C < \left(\Delta_S/\Delta_L\right) p_S^C$, we show that the principal cannot improve her payoff by instead inducing the designer to select an experiment in Π_S^R. To see this, suppose that the designer were restricted to select experiments in Π_S^R. Then, using $\widetilde{\lambda}_S(p)$ in (A.6) expressed as a function of q,

$$\widetilde{\lambda}_S(q) = \frac{q - \underline{q}_S}{q - \underline{q}_S + q - \underline{q}_H}.$$

Replacing this auditing intensity in (23), we see that the principal's maximum expected utility is obtained from (23) by setting $q = 1$. But then we have

$$U\left(\pi_S(q=1)\right) = \underline{q}_H + \frac{\left(1-\underline{q}_H\right)\left(\mu - \underline{q}_S\right)}{2-\underline{q}_S-\underline{q}_H} \leq \underline{q}_H + \frac{\left(1-\underline{q}_H\right)\left(\mu - \underline{q}_L\right)}{2-\underline{q}_L-\underline{q}_H} \leq U\left(\pi_L(q=1)\right). \tag{A.17}$$

Suppose now that $p_L^C > (\Delta_S/\Delta_L)\, p_S^C$, so that the designer selects a status-quo robust experiment if $\lambda = 1$. Equation (A.17) shows that the principal's maximum payoff when the designer is restricted to "up-or-down" experiments is greater than her maximum payoff if the designer is restricted to a status-quo experiment. Therefore, the principal would rule out the status-quo decision d_S, and (17) and (18) would again be an optimal data governance. ■

Proof of Lemma 5: Suppose that the principal organizes for scale and consider experiment $\{\underline{q}_S, q\}$, $q > \underline{q}_H$. The principal is indifferent between decisions d_L and d_S both after a conclusive audit determines $s = \underline{q}_S$ and after an unaudited message $m = \underline{q}_S$. Let $\tau_I(m)$ be the probability of choosing d_S after message m and a conclusive audit finds $s = \underline{q}_S$ and $\tau_U(m)$ be (i) the probability of selecting d_S after an unaudited $m = \underline{q}_S$, and (ii) the probability of selecting d_H after an unaudited $m = q$. Then,

$$v_I(\underline{q}_S, \underline{q}_S) = v_L + \tau_I(\underline{q}_S)(v_S - v_L)$$
$$v_U(q) = v_S + \tau_U(q)(v_H - v_S)$$

are, respectively, the payoff from truthtelling after a conclusive audit if $s = \underline{q}_S$ and the payoff after an inconclusive audit if $m = q$. Note that the gain from tampering after $s = \underline{q}_S$ is

$$\widetilde{c} \equiv (1-\lambda)[(1-\tau_U(\underline{q}_S))(\Delta_L - \Delta_S) + \tau_U(q)\Delta_S] + \lambda(\tau_I(q) - \tau_I(\underline{q}_S))(\Delta_L - \Delta_S), \tag{A.18}$$

so that the probability of truthful communication conditional on $s = \underline{q}_S$ is $\bar{F}(\widetilde{c})$. The designer's equilibrium payoff from a non-robust status-quo experiment V_S^{NR} can then be written as

$$V_S^{NR} \equiv \Pr[s = \underline{q}_S]\left(v_I(\underline{q}_S, \underline{q}_S) + F(\widetilde{c})\widetilde{c}\right) + \Pr[s = q]\left(\lambda v_H + (1-\lambda) v_U(q)\right). \tag{A.19}$$

Recall that $\bar{c}_S$, defined in (3), is the tampering gain that leads the designer's posterior to $\underline{q}_H$ after an unaudited message $m = q$, and it satisfies $\bar{F}\left(\bar{c}_S\right) = p_S^C\left(\frac{q-\underline{q}_S}{q-\mu}\right)$. Therefore, principal's sequential rationality requires $\tau_U(q) = 1$ if $\widetilde{c} < \bar{c}_S$ and $\tau_U(q) = 0$ if $\widetilde{c} > \bar{c}_S$.

To find the designer's maximum payoff, fix $\tau_I(\underline{q}_S)$, which pins down the payoff under an audited truthful report $v_I(\underline{q}_S, \underline{q}_S)$. Suppose that $\widetilde{c} < \overline{c}_S$. Then, $\tau_U(q) = 1$ and $v_U(q) = v_H$, so that $(A.19)$ increases in $\widetilde{c}$. Likewise, if $\widetilde{c} > \overline{c}_S$ then $\tau_U(q) = 0$ and $v_U(q) = v_S$, so that $(A.19)$ again increases with $\widetilde{c}$. In other words, the designer's payoff increases with the tampering gain whenever it is different from the threshold for a robust experiment. Suppose now that $\widetilde{c} = \overline{c}_S$. Then, to maintain (A.18) constant we must decrease scale-up probability $\tau_U(q)$ whenever $\tau_I(q)$ increases, but this reduces $v_U(q)$ and so V_S^{NR} decreases with $\tau_I(q)$–see (A.19). Therefore, for fixed $\tau_I(\underline{q}_S)$ the maximum payoff to the designer is either achieved in an equilibrium in which $\widetilde{c} = \overline{c}_S$ or an equilibrium which maximizes $\widetilde{c}$ –which, if $\max \widetilde{c} > \overline{c}_S$ is attained by setting $\tau_I(q) = 1$.

We now show that the maximum payoff to the designer is obtained when $\widetilde{c} = \overline{c}_S$ and $v_I(\underline{q}_S, \underline{q}_S) = v_U(\underline{q}_S) = v_S$ so that the principal rewards truthtelling. First, note that for $v_I(\underline{q}_S, \underline{q}_S) = v_U(\underline{q}_S) = v_S$ there is always an equilibrium in which the principal scales-up with positive probability after an unaudited $m = q$. This follows as setting $\tau_U(q)$ at the level of a robust experiment and adjusting $\tau_I(q) < 1$ would give $\widetilde{c} \equiv (1-\lambda)[\tau_U(q)\Delta_S] + \lambda(\tau_I(q) - 1)(\Delta_L - \Delta_S) < \overline{c}$. This also implies that the designer's payoff exceeds $v_S + \lambda p v_H$ as the principal scales-up also when the audit is inconclusive. However, if $\widetilde{c} > \overline{c}_S$, then the maximum payoff to the designer cannot be above $v_S + \lambda p v_H$, as the principal scales-up only if a conclusive audit yields $s = q$. It follows that the designer's optimal is achieved when $\widetilde{c} = \overline{c}_S$. Finally, suppose that the optimal is achieved for $v_I(\underline{q}_S, \underline{q}_S) < v_S$ or $v_U(\underline{q}_S) < v_S$ so that the principal (partially) punishes truthtelling when $m = \underline{q}_S$. Increasing either value while maintaining $\widetilde{c} = \overline{c}_S$ raises the payoff $v_U(q)$ and thus increasing V_S^{NR}. Therefore at the maximum we must have $v_I(\underline{q}_S, \underline{q}_S) = v_U(\underline{q}_S) = v_S$

In summary, the designer's maximum payoff from experiment $\{\underline{q}_S, q\}$ is achieved by choosing d_S after the agent truthfully sends $m = \underline{q}_S$ but punishing tampering with the lowest $\tau_I(q)$ consistent with a tampering gain of $\widetilde{c} = \overline{c}_S$. Note that if the principal rubberstamps the agent's recommendation for a robust experiment, then the lowest $\tau_I(q)$ consistent with $\widetilde{c} = \overline{c}_S$ is precisely $\tau_I(q) = 1$. Therefore, a robust status-quo experiment $\{\underline{q}_S, q\}$ achieves the designer's maximum payoff in any PBE following $\{\underline{q}_S, q\}$ if the principal rubberstamps the agent's recommendation. ■

Proof of Proposition 11: In the text. ■