Organizing Data Analytics*

RICARDO ALONSO†

London School of Economics and CEPR

ODILON CÂMARA‡

University of Southern California

January 16, 2023

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 analysis.

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, Marco Ottaviani, 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, evidence shows that companies are struggling to capture value from analytics (McKinsey and Co, 2018a). A key friction is that organizations still *need to rely on people* to design the experiments, analyze the data and report the results, and their preferences might not be aligned with the goals of the organization. The information that reaches decision makers is then 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 study optimal organizational practices for managing data analytics in the face of these frictions.

To illustrate our main insights, consider the following principal-agent scenario. A business unit manager (agent) is proposing an "innovation" that he created. That is, he developed a new design for a product or a new production process and would like the firm to adopt it. A high-level executive in the firm's headquarters (principal) must choose whether to adopt the innovation (decision d_H) or to toss it out and retain the status quo (decision d_S). There is uncertainty regarding the true consequences of adopting the innovation: compared to the status quo product or process, the innovation has a higher quality ("high" state $\theta = 1$) with probability μ, and a lower quality ("low" state $\theta = 0$) with probability $1 - \mu$. The principal

[1]Brynjolfsson et al. (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.

would like to adopt the innovation if its quality is "high" and toss it out if it is "low". The key conflict of interest is that the agent is biased as he prefers the firm to adopt his innovation and, consequently, he is inclined to overstate its quality.[2] How can the principal acquire more information to make a better decision?

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.[3] In a frictionless world, the principal would be able to directly design and run experiments to learn about the quality of the innovation and then decide whether to adopt it.

As a motivating example, the principal could run an A/B test contrasting the new and old products. Differences in test design change what can be learned about the product: the product could be tested by only using a sample of urban consumers, only using a sample of rural consumers, or using a proportional sample of urban and rural consumers. Many other characteristics of the test could be adjusted, such as the size of the sample and the set of control variables used. The principal would like to implement the "most informative" test, where we say that a test π_A is more informative than a test π_B if observing test π_A allows the principal to make better decisions (achieve a higher expected payoff).[4]

The first practical problem is that, typically, high-level executives cannot design and run the experiments themselves: they must delegate experimentation to an agent (business unit manager) with more time and expertise but whose preferences may not be fully aligned with them.[5] Delegation creates a second problem: even if a very informative experiment is feasible, the agent might strategically select a less informative experiment in order to increase

[2]Business managers may favor the implementation of their own innovations because of a preference for empire building, because their human capital is tied to this decision, or because this would increase their visibility and improve their outside opportunities.

[3]Most data analytics is used for process or product improvement or is related to other types of innovation—see Wu et al. (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.

[4]Therefore, our definition of "more informative" follows Blackwell and Girshick (1954).

[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.

the probability of obtaining an outcome that would convince the principal to adopt his innovation. In our example, the agent might choose to run the A/B test only using consumers from urban areas because the new product has a better chance of success within this demographic, while the principal would prefer a more informative test using a proportional mix of urban and rural consumers. Finally, a third problem is that the agent still needs to report the outcome of the experiment to the principal. In this communication stage, the agent might be tempted to tamper with the experiment and report a false outcome. With these frictions in mind, we want to understand how the organization of data analytics can provide incentives for agents to implement the most informative experiments and truthfully report their outcomes.

We start with an overview of the main elements and timing of our model, which is illustrated by Figure 1. The principal first chooses the company's *data governance*, comprising two types of policies: a *tampering prevention* policy and a *tampering detection* policy. Tampering prevention is captured by a cost distribution $F(c)$, which defines how costly will be for the agent to tamper with the result of the experiment. The principal can change the distribution $F(c)$ by making it easier or harder to tamper with the experiment.[6] Tampering detection is captured by the variable $\lambda \in [0, 1]$, which represents the firm's auditing intensity - the probability that the principal also observes the actual outcome of the experiment.

We see *data analytics* as comprising two tasks: the *experimental design* task specifies which data to collect and how to process it, while the *analysis* task is the actual implementation of the experiment according to the specified design and the reporting of its outcome. The principal must allocate these tasks: she can assign both tasks to a single agent (integration) or assign them to two different agents (separation).

After the principal announces the data governance policies and task allocation, the agent(s) perform the two data analytics tasks. The agent responsible for the experimental design (the "designer") strategically designs an experiment π that reveals information about the payoff-relevant state — as in Kamenica and Gentzkow (2011), KG henceforth.[7] In our

[6]For instance, through a by-law that defines the punishment for tampering, or through data security measures that make tampering more or less costly.

[7]While our experimental design task is similar to KG, our novel features are the agent's possibility of paying a cost to tamper and the principal's ability to audit and design organizational features.

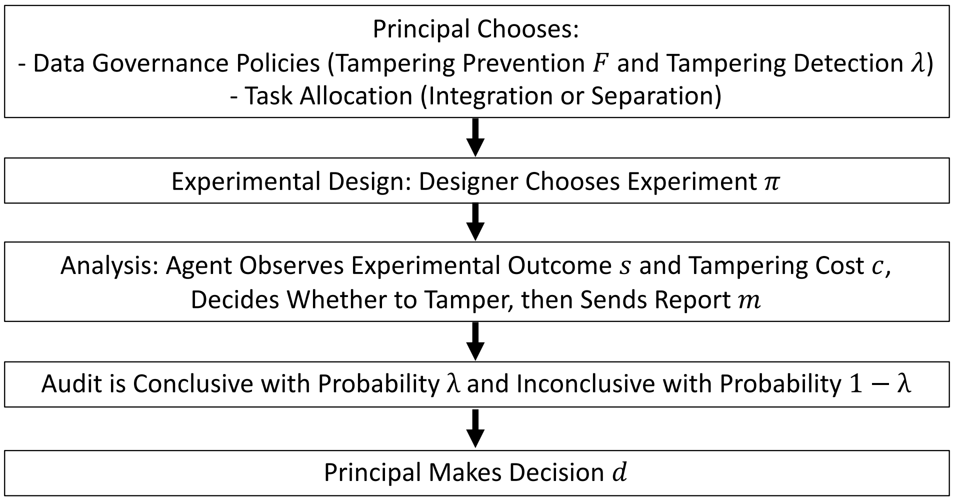


Figure 1: Timing of Decisions

motivating example, the agent can design all the details of the protocol that will be used to run the A/B test. Then, the agent responsible for the analysis (the same agent under integration or a different agent under separation) runs experiment π and privately observes its outcome s. At this point, the agent observes how costly it would be to tamper with the experiment: he observes the cost c drawn from $F(c)$. The agent can truthfully report the outcome of the experiment (send a message $m = s$) or can pay the cost c to tamper with the outcome and send a false report (send a message $m \neq s$). The audit takes place — with probability λ the audit is conclusive and the principal observes both the message m and the true outcome s; with probability $1 - \lambda$ the audit is inconclusive and the principal only observes the message m. Finally, the principal updates her belief and decides whether to adopt the innovation or retain the status quo.

We first show that we can greatly simplify the designer's problem: there exists an optimal experiment taking the form of a pass/fail test with no type II error and an optimal rate α of type I error. If the innovation has a high quality, then the test always results in a "pass" outcome, interpreted as a recommendation to adopt it. If the innovation has a low quality, then with probability α the result will be a "pass" (a false positive), while with probability $1 - \alpha$ the result will be a "fail," interpreted as a recommendation to retain the status quo. The fundamental conflict of interest is that the principal prefers tests with a lower α (these tests are more informative) while the agent prefers tests with a higher α (they imply a higher probability of a pass result). In our A/B test example, a test that correctly balances the sample lowers the probability of a false positive, while strategically selecting a more favorable sample increases the probability of a false positive. In equilibrium, if the audit is not perfect (if $\lambda < 1$), then the agent will tamper (i.e., report a "pass" outcome when the actual test outcome was "fail") as long as his tampering cost c is sufficiently low.

First, consider what happens under integration (single agent). To understand equilibrium strategies, we start by isolating the roles of tampering prevention and detection. Fix an imperfect prevention policy[8] and consider the principal's choice of an optimal tampering detection policy. If the experiment's design was exogenously fixed — if α were exogenous — then the principal would choose a perfect audit $\lambda = 1$ to eliminate any tampering. However, this is not true when the design is endogenously chosen by the agent. Our first result highlights the role of tampering detection: increasing auditing intensity reduces the informativeness of the agent's experiment (the endogenous α increases with λ). To wit, agents 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 which provides no valuable information to the principal — the chosen α is so high that when the experiment recommends adoption, the principal is just indifferent between the two decisions and hence gains no surplus from the experiment. With an imperfect audit, the agent will be tempted to tamper after observing a negative result. The principal takes into account this tampering and discounts adoption recommendations after an inconclusive audit. To compensate for the tampering and partially offset the principal's mistrust, the agent must then select a more informative experiment (must lower the probability of a false positive). Thus, by committing to an imperfect audit ($\lambda < 1$), the principal optimally trades-off the increased tampering with the increased informativeness of the experiment.

Now fix an imperfect audit and consider the principal's choice of an optimal tampering prevention policy. Our second result shows that the principal prefers an imperfect prevention policy that results in low tampering costs being sufficiently likely. This incentivizes tampering and provides decision makers with commitment power to reject self-serving recommendations. The principal's credible threat of rejecting an adoption recommendation leads the

[8]Prevention policy $F(c)$ is imperfect if the agent tampers with a positive probability.

agent to again design a more informative experiment (lower α).

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 "credible skepticism" in the organization: the principal can refrain from following the agents' adoption recommendations issued with weak supporting evidence (high α), forcing the latter to provide stronger evidence backing them. 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 agent always selects a fully informative experiment* ($\alpha = 0$) but tampers with positive probability. That is, organizations in our model 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 (see Section 5.3 for details).

So far we have focused on the case of integration: the same agent is responsible for the tasks of *experimental design* and *analysis.* However, in many cases, these tasks can be delegated to different workers. If the principal could delegate the experimental design role to an unbiased data scientist—while the biased manager runs the experiment and possibly tampers the result—then the principal would prefer to foster a culture of trust in analytics. The trusted data scientist would design a fully informative experiment and the principal would implement 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).[9] However, even if she could hire unbiased data scientists, empirical evidence shows that firms cannot simply delegate experimental design to them—lack of visibility and knowledge of the business unit forces firms to rely on the participation of business managers when designing experiments.[10] To circumvent this problem, analytics centers typically em-

[9]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).

[10] 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, effective delegation of the design of experiments requires the involvement of employees with knowledge of the business unit. McKinsey recommends that the design of

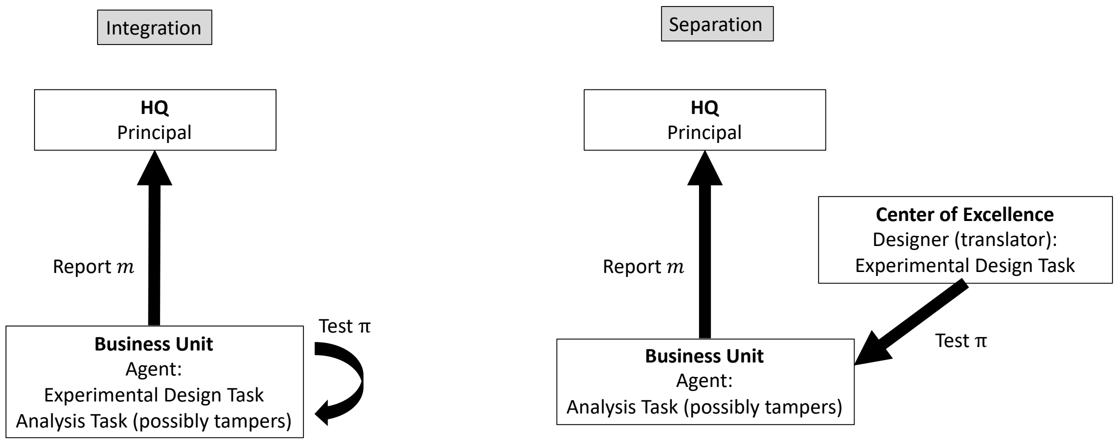


Figure 2: Integration versus Separation

ploy "translators"—employees sourced from the business units with business knowledge—to work with the data scientists on the design of experiments (McKinsey 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. Given that we observe firms using translators from business units, we want to contrast the integration case — the same agent designs and analyses the experiment — and the separation case — the design task is allocated to a designer (translator) outside the business unit but with the same preference as the agent; see Figure 2. Does the principal indeed benefit from separation?

Our fourth result shows that the principal prefers to separate the design task from the analysis task, even when the designer has the same bias as the agent. When tasks are integrated, the agent incurs the tampering costs whenever he misrepresents negative results; consequently, he prefers to design a less informative experiment, which reduces his temptation to tamper and allows him to economize on tampering costs. In contrast, under separation, the designer is willing to implement a more informative experiment since the agent will be the one incurring the tampering costs. Moreover, in line with our previous results, imperfect tampering prevention/detection policies continue to be optimal in the case of separation. Our insights resonate with organizations that centralize the design of experiments in a Center of Excellence (CoE) while coming short of implementing watertight auditing measures.[11]

analytics solutions needs to have business participation from the start.

[11]For instance, McKinsey and Co. McKinsey and Co. (2018b). reports on several firms centralizing data

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. Section 6 extends the model to allow for tampering costs to be incurred only if audited and also studies the case of multiple decisions. 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 involved in data analytics in a parsimonious way, 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 a decision.
Preferences and Prior Beliefs: Players are expected utility maximizers. To facilitate the analysis, we consider a binary state-space with typical realization $\theta \in \Theta = \{0, 1\}$; players hold a common prior $\mu = \Pr[\theta = 1]$. The principal selects d from $\{d_S, d_H\}$, and has preferences characterized by $u(d, \theta)$,

$$u(d, \theta) = \begin{cases} \underline{q}_H & \text{for } d = d_S, \\ \theta & \text{for } d = d_H, \end{cases}$$

with $0 < \underline{q}_H < 1$. In words, the principal either keeps the status quo d_S, or "approves" the innovation d_H, selecting d_H only when her posterior belief q does not fall below $\underline{q}_H$.

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 $0 = v_S < v_H$, so that they benefit from persuading the principal to approve d_H. We focus on the more interesting case where $\mu < \underline{q}_H$–so that the principal retains the status quo in the absence of any experimentation–and we let $\Delta \equiv v_H - v_S$ be the designer/agent's gain from inducing their preferred decision d_H.

Strategic Experimentation, Reporting and Tampering: All players process information according to Bayes' rule. We consider three stages.

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

analytics around a CoE tasked with homogenizing data analytics and supporting the different business units.

$S(\pi)$ and a family of likelihood functions over $S(\pi)$, $\{\pi(\cdot|\theta)\}_{\theta\in\Theta}$, with $\pi(\cdot|\theta) \in \Delta(S(\pi))$, where $\Delta(S(\pi))$ represents the set of probability distributions over $S(\pi)$. We say that the designer "experiments more" when he selects a Blackwell-more informative experiment (see Blackwell and Girshick (1954)).

We make two assumptions regarding experimental design that are consistent with our initial motivation—i.e., the drastic reduction in the costs of data gathering and storage (see Goldfarb and Tucker, 2019). First, 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 perfectly 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/reporting stage. The agent privately observes the outcome $s \in S(\pi)$ and submits a report/message $m \in S(\pi)$ to the principal, which is 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.[12] 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 by the agent when he tampers. In Section 6.1, we show that our main insights 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 experimental outcome 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 chosen experiment and the agent's report. Key in our model is the principal's ability to

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

evaluate the truthfulness of this report, and undo the effect of any misrepresentation, by auditing the experiment. We assume that with probability λ the audit is conclusive and the principal observes s, 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.[13] To lighten the exposition, we say that "the message/recommendation is (un)audited" when the audit is (in)conclusive. Thus, when the message is audited the principal selects (a possibly mixed) $d_A(m, s)$ which depends on the message m and the outcome s; if unaudited, she selects $d_U(m)$.

Organizational Design: Agents 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. In terms of tampering prevention, the principal can enact data encryption and authentication systems to preserve data integrity, or security measures to make access to data storage systems costly. Effectively, through these policies the principal can specify any distribution of tampering costs F, including making tampering arbitrarily costly.

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.[14] 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

[13] See, for instance, Kolotilin (2018), Kolotilin et al. (2017), and Guo and Shmaya (2019) for information-design models where the receiver is privately informed.

[14] 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.

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.[15]

Commitment Experiment: An important assumption is that the principal can commit to an imperfect audit, i.e., to $\lambda < 1$. To understand what happens if the audit were perfect ($\lambda = 1$), and for future reference, consider binary (α, β)-tests π with $S(\pi) = \{S, H\}$, $\Pr[s = H|\theta = 1] = 1-\beta$ and $\Pr[s = H|\theta = 0] = \alpha$. These are tests that recommend a decision (d_S if $s = S$ and d_H if $s = H$) and incur rates of type I and type II errors of α and β, respectively. If $\lambda = 1$, then the designer selects the experiment that minimizes the probability of retaining the status quo, $\Pr[s = S] = \mu\beta + (1-\mu)(1-\alpha)$, subject to the principal approving after $s = H$, i.e., subject to $\Pr[\theta = 1|s = H] \geq \underline{q}_H$. As shown in KG, the optimal experiment π_C, which we will refer to as the *"commitment experiment,"* can be found by optimizing in the smaller set of (α, β) experiments. In fact, π_C sets $\Pr[\theta = 1|s = H] = \underline{q}_H$ with $\beta = 0$ and

$$\alpha = \frac{\mu/(1-\mu)}{\underline{q}_H(1-\underline{q}_H)}(\equiv \alpha_H), \tag{1}$$

which leads to a probability of retaining the status quo of

$$p_S^C \equiv \Pr[s = S] = \frac{\underline{q}_H - \mu}{\underline{q}_H} = (1-\mu)(1-\alpha_H). \tag{2}$$

As the principal is indifferent between approving and retaining the status quo when $s = H$, the experiment does not provide any valuable information regarding decision d_H. If the organization hopes to induce more experimentation, it must be able to limit the success rate of an audit to $\lambda < 1$. 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. Note the importance of the principal's commitment to an imperfect audit: absent such commitment, and once the designer selects

[15] 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, 2016a for details). In an example of task-integration, the BIT would both design and conduct small-scale experiments for the UK Government.

an experiment, the principal can completely eliminate the effect of tampering by setting $\lambda = 1$. In anticipation of a perfect audit, however, the designer would select an experiment that provides no value to the principal.

Timing: The timing of the game is illustrated in Figure 1. The principal publicly selects a task allocation k and data governance (λ, F). Given (k, λ, F), the designer publicly selects π with outcomes $S(\pi)$. The communication subgame follows: Nature draws θ and the agent privately observes $s \in S(\pi)$ and the cost c, generated according to π and F, and selects $m \in S(\pi)$. 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 for each (k, λ, F).[16]

3 Equilibrium Experimentation and Reporting

We start the organizational-design analysis by studying equilibria in the *designer-agent-principal* game corresponding to a fixed organizational structure. We work backwards and first characterize equilibria in the communication subgame for any experiment π, which will determine both expected tampering and the distribution over the principal's decisions. We then turn to the designer's optimal choice by introducing the set of *status-quo experiments*—a set of pass/fail tests which always contains a solution to the designer's problem.

3.1 Equilibrium Tampering

The agent decides whether to tamper by comparing the gain from misrepresenting the experimental outcome to the realized tampering cost. Note that $(1 - \lambda)\,\Delta$ is the maximum gain from tampering in any equilibrium (both on- and off- the equilibrium path). Throughout the paper, we consider data policies (λ, F) that guarantee after every experimental outcome the existence of tampering costs that make tampering unprofitable:

[16]Specifically, we look for profiles of strategies that constitute a weak Perfect Bayesian Equilibria for every subgame.

Assumption 1 (All messages on-path) *The tampering prevention policy satisfies*

$$\bar{F}\left((1-\lambda)\,\Delta\right) > 0. \tag{3}$$

Assumption 1 ensures a positive probability of truthful reporting for every potential experiment chosen by the designer and limits the scope of the principal to discipline tampering by holding "optimistic or pessimistic" beliefs after an off-the-equilibrium-path message.

The agent that tampers sends a message that leads to the largest approval probability after an inconclusive audit. Thus, if after different experimental outcomes the agent tampers by sending different messages, then they all must induce the same unaudited decision (or mixtures over decisions). This observation helps characterize tampering behavior in the communication subgame.

Proposition 1. *Let $m^*(s,c)$ be the agent's equilibrium reporting in the communication subgame following the choice of π with experimental outcomes $S(\pi)$. Then,*
(i) For each $s \in S(\pi)$, there exists $\bar{c}(s)$, with $\bar{F}(\bar{c}(s)) > 0$, such that $m^*(s,c) = s$ if $c > \bar{c}(s)$ and $m^*(s,c) \neq s$ if $c < \bar{c}(s)$;
(ii) Let $M_T(\pi) = \{s \in S(\pi) : \exists (s_z, c),\ m^*(s_z, c) = s, s_z \neq s\}$ be the set of "tampered outcomes." Then for every $s, s' \in M_T(\pi)$, (a) $d_U(s) = d_U(s')$, and (b) $\bar{c}(s) = \bar{c}(s') = 0$.

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

3.2 Equilibrium Experimentation

To characterize the designer's equilibrium experiment, we introduce the set of "status-quo" experiments Π_S. These are binary experiments that recommend d_H (outcome $s = H$) or d_S

(outcome $s=S$), and such that $\Pr[s=H|\theta=1]=1$ but $\Pr[s=H|\theta=0]=\alpha$. Thus, status-quo experiments are "pass/fail" tests with no false negatives when following a d_S recommendation, but with the rate of false positives when following a d_H recommendation set to α.

Definition (Status-quo Experiments) *Define the set of status-quo experiments,* Π_S,

$$\Pi_S = \{\pi_S(\alpha) : s \in \{S, H\}, \Pr[s=H|\theta=1]=1 \textit{ and } \Pr[s=H|\theta=0]=\alpha; \alpha \in A\}, \quad (4)$$

where α *ranges in the set*

$$A = \left[max\left\{0, \frac{\alpha_H - F((1-\lambda)\Delta)}{\bar{F}((1-\lambda)\Delta)}\right\}, \alpha_H\right], \quad (5)$$

with α_H *the type I error incurred under the commitment experiment–see (1).*

Associated with each status-quo experiment $\pi_S(\alpha)$ is the unique continuation equilibrium of the ensuing communication subgame that pins down the designer's payoff if he selects $\pi_S(\alpha)$. This equilibrium is characterized by tampering thresholds and approval probability[17]

$$\bar{F}(\bar{c}) = \frac{1-\alpha_H}{1-\alpha}, \textit{ and } \tau = \frac{\bar{c}}{(1-\lambda)\Delta}, \quad (6)$$

alongside the principal's decision making

$$d_U(S) = d_A(m, S) = d_S; d_A(m, H) = d_H; d_U(H) = \tau d_H + (1-\tau)d_S. \quad (7)$$

A key property of status-quo experiments is that equilibrium tampering leads the principal's posterior to $\underline{q}_H$ after an inconclusive audit of an approval recommendation ($m=H$), so that she is indifferent between approving or retaining the status-quo.[18] In equilibrium, she selects d_H with probability τ, and d_S with probability $1-\tau$–see (7). This approval probability must be consistent with the incentives to tamper—as given by the tampering threshold (6)—which also dictate the probability $\bar{F}(\bar{c})$ that the agent reports truthfully after $s=S$.

[17] For a status-quo experiment, the agent tampers only if $s=S$. Abusing notation, we will refer to the tampering threshold $\bar{c}$ with the understanding that $\bar{c}(S)=\bar{c}$ and $\bar{c}(H)=0$.

[18] Note that it can never be optimal for the designer to select an experiment such that, in equilibrium, the principal has a posterior belief strictly above $\underline{q}_H$ after an unaudited $s=H$. Indeed, the designer could move to an experiment that after $\theta=0$ induces $s=H$ slightly more often. This experiment would still generate an unaudited posterior belief above $\underline{q}_H$–so that the principal still approves with probability one and the agent faces the same tampering incentives– but would raise the probability of a high outcome, thus increasing the designer's payoff.

Importantly, the set Π_S depends on the auditing intensity λ–this can be seen in (5) as the minimum type I error in Π_S increases with λ. For instance, if $\lambda = 1$ then the commitment experiment is the only status-quo experiment —i.e, $A = \{\alpha_H\}$. Finally, all experiments in Π_S are ordered according to their informativeness: trivially, an experiment with a lower type I error α (equivalently, higher $\bar{c}$ or higher τ) corresponds to a Blackwell-more informative experiment–see Footnote 4.

We now present our main equilibrium characterization of the *designer-agent-principal* game. Given data governance policies (λ, F) and task allocation $k \in \{\mathcal{S}, \mathcal{I}\}$, let $v_S(\alpha)$ be the designer's equilibrium payoff in a communication subgame after he selects $\pi_S(\alpha) \in \Pi_S$, and let

$$V_S \equiv \max_{\alpha \in A} v_S(\alpha), \tag{8}$$

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

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

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 selects an experiment in Π_S. We prove this claim in the appendix by constructing from an arbitrary experiment π' a status-quo experiment that gives the designer a (weakly) higher payoff. For instance, if for π' the principal's posterior is strictly above $\underline{q}_H$ when approving an unaudited recommendation, then the designer can increase the likelihood of this recommendation without sacrificing approval probability. Second, all equilibria give the designer the same expected payoff; thus, to find this payoff we can restrict attention to status-quo experiments.

For the remainder, let $\pi_S(\alpha^*) \in \Pi_S$ be the designer's optimal experiment, with α^* the associated type I error and $\bar{c}^*$ satisfying (6) for $\alpha = \alpha^*$ the induced tampering threshold, where we omit the explicit dependence on the organizational design parameters $\{k, \lambda, F\}$. Also, we will abuse notation and use $\alpha^*(x)$ to denote the designer's optimal choice of α for an x organizational lever (task allocation, auditing or tampering prevention).

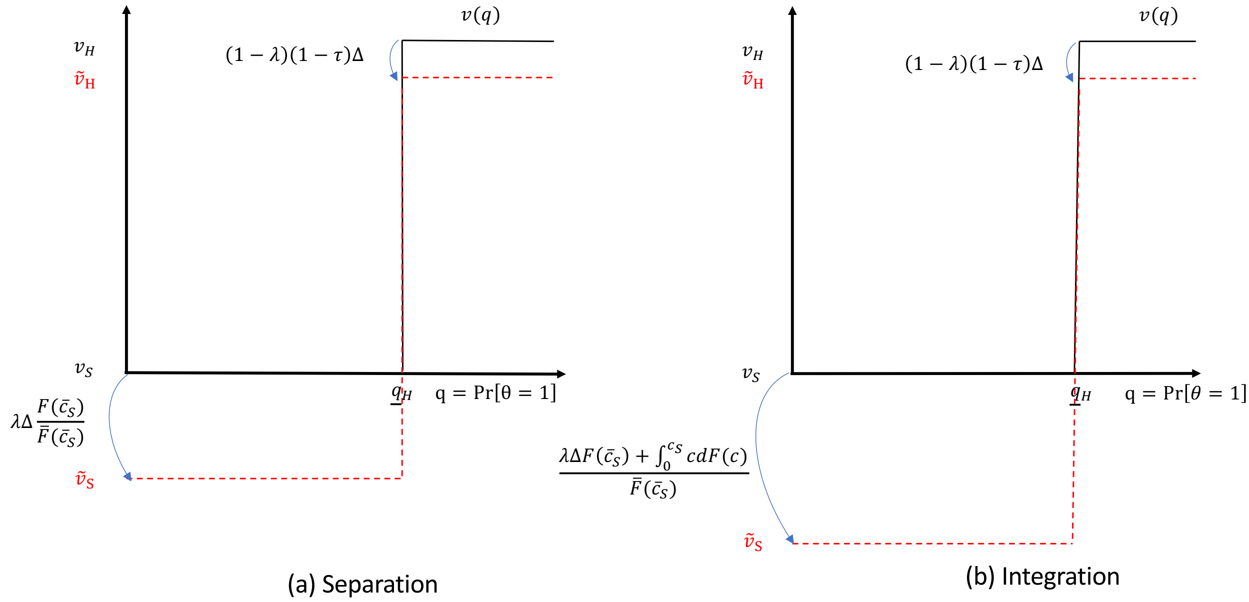


Figure 3: Payoffs for the original game (solid line) and the equivalent "full commitment" game (dotted line).

3.3 The Impact of Tampering on the Designer's Payoffs

If the principal perfectly audits the experiment ($\lambda = 1$), then tampering is inconsequential but the designer selects the commitment experiment; the designer's payoff is then $\left(1 - p_S^C\right)\Delta$ where $1-p_S^C$ is the probability that the experiment recommends d_H–see (2). A key property of equilibria is that for any status-quo experiment $\pi_S(\alpha)$ that generates $s = H$ with probability $1 - p(\alpha)$, the agent nevertheless reports $m = H$ with probability $1 - p_S^C$.[19] This observation helps us write the designer's payoff from an experiment that induces threshold $\bar{c}$ as[20]

$$v_S(\bar{c}) = (1 - p(\alpha(\bar{c})))\,\lambda\Delta + (1 - p_S^C)(1-\lambda)\Delta\tau(\bar{c}) - \mathbb{I}_{\{k=\mathcal{I}\}}p(\alpha(\bar{c}))\int_0^{\bar{c}} cdF(c). \quad (9)$$

The first term in (9) is the designer's payoff when the audit is conclusive–in which case, and regardless of tampering, the innovation is approved iff $s = H$–while the second term is the payoff when the audit is inconclusive and the last term captures the tampering costs when tasks are integrated.

[19]This follows from the fact that any unaudited d_H recommendation leads to the same posterior $\underline{q}_H$ as in the commitment experiment.

[20]$\alpha(\bar{c})$ denotes the type I error satisfying (6) for a given threshold $\bar{c}$.

How does imperfect auditing and tampering affect the designer's equilibrium payoff from $\pi_S(\alpha)$? For comparisons with the case $\lambda = 1$, we can express (9) as

$$v_S(\bar{c}) = \left(1 - p_S^C\right) \tilde{v}_H + p_S^C \tilde{v}_S, \tag{10}$$

with

$$\tilde{v}_H \equiv v_H - (1-\lambda)(1-\tau(\bar{c}))\Delta, \tag{11}$$

$$\tilde{v}_S \equiv v_S - \lambda\Delta\frac{F(\bar{c})}{\bar{F}(\bar{c})} - \mathbb{I}_{\{k=\mathcal{I}\}}\frac{\int_0^{\bar{c}} c dF(c)}{\overline{F}(\bar{c})}. \tag{12}$$

The principal keeps the status quo with probability $(1-\lambda)(1-\tau(\bar{c}))$ after $m = H$ –this explains (11)– but also when the agent reports $s = S$ or when he tampers and the report is audited, which happens with probability $\lambda F(\bar{c}) p_S^C / \bar{F}(\bar{c})$—see (4)—this explains (12) for $k = \mathcal{S}$. Finally, tampering costs are incurred only if $s = S$ – which occurs with probability $p_S^C / \bar{F}(\bar{c})$ – in which case the agent tampers if $c \le \bar{c}$. This explains (12) for $k = \mathcal{I}$.

Figure 3 compares the payoffs under perfect commitment to the equivalent payoffs in the imperfect commitment using representation (11-12). These expressions capture the trade-off that the designer faces: a higher approval probability increases the payoff after an unaudited message $m = H$, $\tilde{v}_H$, but a higher approval probability (higher τ) can only result from a higher tampering threshold (higher $\bar{c}$) and, thus, a higher likelihood of tampering. To keep the principal's posterior from falling below the approval threshold $\underline{q}_H$, the designer must select a more informative experiment (lower α), thus, increasing the likelihood of observing an unfavorable outcome $s = S$ and lowering $\tilde{v}_S$.

4 Organizational Design: Task Allocation

We now turn to the organizational design problem. To understand the trade-offs that the principal faces, consider her expected utility from an status-quo experiment that leads to posterior q after an unaudited approval recommendation :

$$U(\lambda, k) = \underline{q}_H + \Pr[s = H] \lambda (q - \underline{q}_H). \tag{13}$$

The principal benefits from this experiment only after auditing an approval 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 tampering behavior (6), her expected utility increases with both the auditing intensity and with the odds of tampering,

$$U(\lambda, k) = \underline{q}_H + \left(\underline{q}_H - \mu\right) \lambda \frac{F(\bar{c}^*)}{\bar{F}(\bar{c}^*)}. \tag{14}$$

This expression showcases our main insight: fostering experimentation while discouraging tampering are conflicting goals. The principal can always eliminate frictions in communication by perfectly auditing the experiment, or by making tampering sufficiently costly—she would certainly set $\lambda = 1$ if the experiment were exogenous (fixed α) and auditing costless. However, setting $\lambda = 1$ would reduce the information she obtains because the designer would then resort to an experiment with a higher type I error knowing that the agent would not tamper with the outcome. In fact, a fallible data governance allows her to credibly withhold approval if the evidence in favor of d_H is 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 $\lambda < 1$.

4.1 Optimal Task Allocation

Expression (14) clarifies that, given equilibrium behavior, the principal is always willing to trade-off distortions in communication for more informative experimentation.[21] How do the different organizational levers help her motivate experimentation? First, for a fix data governance policy, the principal benefits from task separation as this leads the designer to experiment more.

Proposition 3. *The designer selects a more informative experiment if tasks are separated; i.e., we have* $\alpha^*(\mathcal{S}) \leq \alpha^*(\mathcal{I})$.

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; reducing tampering costs can only be achieved through a reduction in approval probability, leading to a less informative experiment.[22]

[21] Indeed, for a fixed auditing intensity, $U(\lambda, k)$ increases with the odds of tampering $F(\bar{c}^*)/\bar{F}(\bar{c}^*)$.

[22] Decreasing approval probability lowers both the equilibrium tampering threshold and the likelihood that a tampering outcome occurs, leading to lower expected tampering costs.

In other words, when tasks are integrated, the agent incurs the tampering cost whenever he misrepresents negative results; consequently, he prefers to design a less informative experiment, which reduces his temptation to tamper. In contrast, under separation, the designer is willing to implement a more informative experiment since the agent will be the one incurring the tampering costs.

5 Organizational Design: Tampering Prevention and Detection

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

5.1 Tampering Detection

Consider a fixed tampering prevention policy 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. Once experimental design is delegated, however, 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–see (5)–and (ii) (weakly) reduces the principal's approval probability as the agent's incentives to tamper increase when the audit is less likely to uncover it–see (6). Both effects lead the designer to experiment more in response to reductions in audit intensity.

Proposition 4. *Fix a task allocation $k \in \{\mathcal{S}, \mathcal{I}\}$. Then, type I error $\alpha^*(\lambda)$ is non-decreasing in λ, implying that lowering auditing intensity leads the designer to select a more informative experiment.*

To understand Proposition 4, consider the effect on the designer's payoff in (9) of increasing λ. We can rewrite (9) in terms of the likelihood of a high outcome, $\Pr[s = H] = 1 - p(\alpha)$, for an experiment inducing threshold $\bar{c}$

$$v_S(\bar{c}) = \Pr[s = H]\lambda\Delta + (1 - p_S^C)\bar{c} - \mathbb{I}_{\{k=\mathcal{I}\}} \Pr[s = H] \int_0^{\bar{c}} c dF(c).$$

For a fixed experiment, approval probability must increase if the principal audits more often to keep the same tampering incentives–thus the second and third term above remain unchanged and the marginal effect of a higher λ is equal to $\Delta \Pr[s = H]$. Intuitively, the designer benefits from more intense auditing as, facing the same tampering incentives, the principal approves the innovation more often when the outcome is $s = H$. As this gain is proportional to $\Pr[s = H]$, which is larger for less informative experiments, the designer gains more from increased auditing when the experiment is less informative. Put differently, the complementarity between informativeness and auditing implies that the incentives to experiment are lower for higher λ. Moreover, auditing also reduces the set of status-quo experiments. Proposition 4 shows that the combined effect of a higher λ unambiguously discourages experimentation.

5.1.1 The Optimality of Lax Auditing

From (14), setting $\lambda = 1$ leads to an experiment from which the principal derives no value. Therefore, the principal gains from data analytics only if lax auditing leads to more experimentation. This motivates the notion of the designer's responsiveness to auditing.

Definition (Responsiveness) *The designer is responsive to auditing if for some* $k \in (\mathcal{I}, \mathcal{S})$, $\alpha < \alpha_H$*, and* $0 < \lambda < 1$*, he strictly prefers experiment* $\pi(\alpha)$ *to* $\pi(\alpha_H)$.

We thus reach one of our main results: if auditing is costless and the designer is responsive to auditing, then the principal optimally 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 optimal choice. Then,* $\lambda^* < 1$ *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: if the tampering prevention policy allows for a positive probability of low tampering costs, the optimal audit should be imperfect.

5.1.2 Optimal Auditing

How much should the principal audit the agent's report if she separates tasks? To derive her optimal audit, 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 $\bar{F}\left(\bar{c}_{FI}\right) = 1-\alpha_H$, so that $\bar{c}_{FI}$ is the threshold induced by a fully informative experiment.

Lemma 1. *Fix $\lambda \in (0,1)$ and suppose that $L(c) - (\phi_S/\lambda)$ is single-crossing in $[0,\Delta]$, from negative to positive, with $\phi_S \equiv (1-(1-\mu)(1-\alpha_H))/((1-\mu)(1-\alpha_H)\Delta)$. Then, the designer under separation selects $\alpha = \alpha_H$ (i.e., $\bar{c}^* = 0$) if $L(0) \geq \phi_S/\lambda$. Otherwise, he selects an experiment that induces tampering threshold*

$$\bar{c}^* = \min\left[L^{-1}(\phi_S/\lambda), (1-\lambda)\Delta, \bar{c}_{FI}\right]. \tag{15}$$

Consistent with Proposition 4, as auditing intensifies the designer's optimal experiment leads to less tampering, but is less informative—the equilibrium tampering threshold (15) 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 $\bar{c}_{FI}$ corresponds to a fully informative experiment; $(1-\lambda)\Delta$ corresponds to the case that the principal always rubber-stamps (i.e., automatically approves) the agent's unaudited recommendation; and the first term in (15) 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 $\bar{c}^* = 0$, i.e., he selects the commitment experiment. 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 (14), the principal's problem for a fixed prevention policy is

$$\lambda^* \in \arg\max_{\lambda \in [0,1]} \lambda \frac{F\left(\bar{c}^*\right)}{\bar{F}\left(\bar{c}^*\right)}, \text{ with } \bar{c}^* \text{ given by (15).} \tag{16}$$

The optimal audit will, in general, be sensitive to the firm's prevention policy F and the preferences of agents. To illustrate (16), we study a case where F leads to tampering costs that are uniformly distributed.

Example: Uniform Distribution. Let $\Delta = 1$ and suppose that c is uniformly distributed in $[0,1]$, so that $\bar{F}\left[(1-\lambda)\Delta\right] = \lambda$. From (9), the designer's utility under separation when

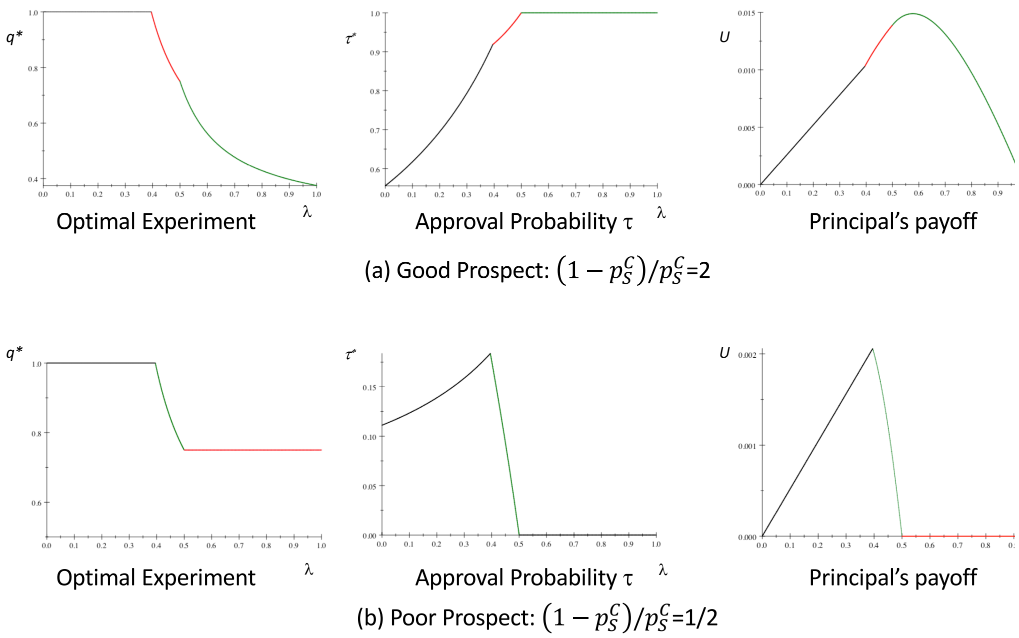


Figure 4: Equilibrium experimentation as a function of auditing for (a) a good prospect, and (b) a poor prospect.

the experiment induces $\overline{c} \in [0, 1-\lambda]$ is

$$v_S(\alpha(\overline{c}), \mu) = \lambda + \overline{c} - m_S(\overline{c}) \left(\underline{q}_H - \mu \right) = \lambda + \overline{c} - (1-\mu)(1-\alpha_H) \left(\overline{c} + \frac{\lambda}{1-\overline{c}} \right),$$

which is concave in $\overline{c}$. Denote by $\overline{c}_{crit} \equiv 1 - \sqrt{\lambda \frac{(1-\mu)(1-\alpha_H)}{1-(1-\mu)(1-\alpha_H)}} = 1 - \sqrt{\lambda \frac{p_S^C}{1-p_S^C}}$–see (1) and (2)–its unconstrained maximum. Then, mirroring (15), the designer's optimal experiment leads to a tampering threshold

$$\overline{c}^* = \min \left\{ \max \left\{ 0, \overline{c}_{crit} \right\}, 1-\lambda, \alpha_H \right\}.$$

We can now fully characterize the equilibrium experiment–in terms of the induced threshold–as a function of λ. Note that $\left(1-p_S^C\right)/p_S^C$ are the approval odds when $\lambda = 1$. If $\lambda \geq \left(1-p_S^C\right)/p_S^C$, then $\overline{c}_{crit} \leq 0$ and the designer selects $\overline{c}^* = 0$, i.e., selects the commitment experiment. If $\lambda \leq p_S^C / \left(1-p_S^C\right)$, then $\overline{c}_{crit} \geq 1-\lambda$ 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 always rubber-stamps an (audited or unaudited) approval recommendation. Finally, if $\left(1-p_S^C\right)/p_S^C \leq \lambda \leq p_S^C / \left(1-p_S^C\right)$, then $\overline{c}^* = \min \left[\overline{c}_{crit}, \alpha_H\right]$ and the designer limits the informativeness of the experiment, leading to intermediate approval probabilities after an inconclusive audit.

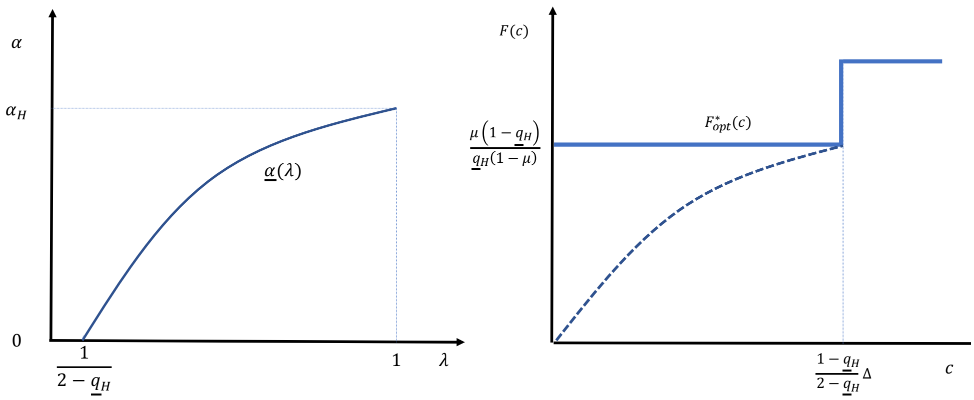
Figure 4 describes two cases, with $\left(1-p_S^C\right)/p_S^C$ taking values 2 and 1/2.[23] If $\left(1-p_S^C\right)/p_S^C = 2$, then the innovation idea is a *good prospect*: it is likely to be perceived after experimentation as a profitable alternative to the current status quo. Then, the designer 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 4-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 rubber-stamp the agent's recommendations.

If $\left(1-p_S^C\right)/p_S^C = 1/2$, 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 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 4-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, as the probability of approving an unaudited recommendation is small.

5.2 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 makes low tampering costs sufficiently likely.

[23]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 $\frac{1-p_S^C}{p_S^C} = 1/2$,



(a) Minimum type I error $\underline{\alpha}(\lambda)$.

(b) Minimum cost distribution to induce $\pi(\underline{\alpha}(\lambda))$ for $\lambda = \frac{1}{2-\underline{q}_H}$.

Figure 5: Minimum Type I error 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

$$\alpha^*(\lambda) \geq \underline{\alpha}(\lambda) \textit{ with } \underline{\alpha}(\lambda) \equiv \frac{2\lambda - 1}{\lambda}\alpha_H - \frac{\mu(1-\lambda)}{(1-\mu)\,\lambda}. \tag{17}$$

(ii) There is a multiplicity of prevention policies that achieve the bound in (17), all of them satisfying the following inequality

$$F\left(c\right) \geq \frac{c}{c + \lambda \frac{p_S^C}{1-p_S^C}\Delta} \textit{ for } c \leq (1-\lambda)\Delta. \tag{18}$$

We prove this proposition by solving an auxiliary problem: to find the maximum auditing intensity that induces the selection of experiment $\pi(\alpha)$ for some cost distribution. The solution is $\widetilde{\lambda}(\alpha) = 1/(1 + (\alpha_H - \alpha)/((\mu 1 - \mu)) - \alpha))$ which is obtained by ensuring that switching to the commitment experiment is never profitable for the designer. Inverting this relation, we then obtain the most informative experiment consistent with auditing intensity λ—depicted in Figure 5(a).

There are many different distributions that would lead the designer to select the experiment with the lowest type I error $\underline{\alpha}(\lambda)$. In all of them, $F(c)$ must exceed some lower bound—

see the dotted line in Figure 5-b. That is, tampering for low cost realizations must be sufficiently likely for the principal to 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 (18) is supported only on two cost realizations, 0 and $(1-\lambda)\Delta$—see the solid-line F_0^* in Figure 5-b—and the agent only tampers if $c = 0$; thus, expected tampering costs are zero.

5.3 Optimal Data Governance

Tampering prevention and detection are perfect substitutes when it comes to dissuading tampering–the principal could achieve zero tampering by implementing a perfect detection $\lambda = 1$ or a perfect prevention $\Pr[c > \Delta] = 1$. 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 the optimal join design of* $\{k^*_{opt}, \lambda^*_{opt}, F^*_{opt}\}$.
(i) The principal sets a prior-independent auditing intensity

$$\lambda^*_{opt} = \frac{1}{2 - \underline{q}_H}. \tag{19}$$

(ii) The designer selects a fully informative experiment.
(iii) There is a multiplicity of optimal tampering prevention policies but, among them, the following minimizes expected tampering costs,

$$F^*_{opt}(c) = \begin{cases} \alpha_H \text{ for } c \in [0, \frac{1-\underline{q}_H}{2-\underline{q}_H}), \\ 1 \text{ for } c \geq \frac{1-\underline{q}_H}{2-\underline{q}_H}. \end{cases} \tag{20}$$

For this cost distribution, the principal is indifferent between separating and integrating tasks.
(iv) The principal can implement (20) 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.*

Our results show that 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 (19) is always lower than 1, but higher than 1/2, and increases with the principal's approval threshold $\underline{q}_H$.

We prove this proposition by appealing to Proposition 6 and optimizing over λ. We also obtain that $\alpha^*(\lambda^*_{opt}) = 0$—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 rubber-stamps any approval recommendation after an inconclusive audit. If the organization wants to minimize the costs imposed upon agents—say because of concerns with hiring costs—then the optimal prevention policy makes tampering either costless or completely deters tampering (see Figure 5). This policy also makes task allocation irrelevant, as expected tampering costs are always zero.

The optimal organization $\{k^*_{opt}, \lambda^*_{opt}, F^*_{opt}\}$ that satisfies (20) can be afforded an intuitive implementation: tampering is costless but the agent's 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 would adjust the internal auditing intensity to their prior belief–i.e., conducting more intense audits for poor prospects that are unlikely to be approved 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 instead known to the principal, the designer in anticipation would then select the commitment experiment. Therefore, the presence of internal screens or firewalls between auditing teams is key for the efficacy of this implementation. 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 (2022) 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.

Finally, our findings relate to insights in the 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 et al., 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).

6 Extensions

In this section we test the robustness of our insights by considering two extensions: tampering costs are only incurred if the agent is caught tampering, and the principal may face a more complex decision problem with a larger choice set.

6.1 Tampering costs incurred only if audit is conclusive.

In some scenarios, a tampering prevention policy may only affect the agent's payoffs if tampering is uncovered. For instance, if it takes the form of punishments after an outcome misrepresentation is exposed–that is, when tampering, the agent only incurs the cost c if the audit is conclusive, which happens with probability λ. We can study this case by adjusting our analysis and 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 $\lambda < 1$, Propositions 1 and 2 still apply, albeit with different threshold values. To see how this affects the designer's payoff, consider a status-quo experiment that induces tampering threshold $\overline{c}$. Then, it is still true that $\bar{F}\left(\overline{c}\right)\left(1-\alpha\right) = \left(1-\alpha_H\right)$–that is, equilibrium tampering must be such that the probability of sending message $s = S$ is equal to p_S^C, see (6)–but now approval probability must be lower to account for the probability that tampering goes undetected:

$$\tau = \frac{\lambda\overline{c}}{\left(1-\lambda\right)\Delta}.$$

Nevertheless, our main results are still robust to this variation. For instance, separating tasks or decreasing auditing intensity always increases experimentation. Moreover, 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. Finally, if the principal can freely shape $\{k, \lambda, F\}$, then the same organizational design as in Proposition 7 remains optimal-see online Appendix B.

6.2 Data Tampering Policies in Complex Environments

A key finding of our binary-decision setup is that enacting policies that allow for some tampering to go undetected leads to more informative experimentation. We show that a qualified version of this finding is still true when the principal has more options available, the qualification being needed as the designer may now select an experiment that recommends different decisions when tampering prevention and detection policies are imperfect.

We provide a full analysis of this case in the online Appendix B; here we briefly outline the main insights. To fix ideas, suppose that the principal now selects d from $\{d_L, d_S, d_H\}$, and has preferences characterized by $u(d, \theta)$, [24]

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

with $0 = \underline{q}_L \leq \underline{q}_S < \underline{q}_H < 1$. That is, the principal either keeps the status quo d_S, scales-up operations by choosing d_H, or scales-down by choosing d_L, with $\underline{q}_i$ the minimum posterior belief for which the principal still selects d_i. Agents have state-independent preferences $v(d_i, \theta) = v_i$ with $0 = v_L < v_S < v_H$.

With more decisions, the designer can still opt for a status-quo experiment that recommends d_H or d_S, but he could now select an "up-or-down" experiment that recommends d_H or d_L. For instance, if $\lambda = 1$, he selects an "up-or-down" experiment whenever $p_L^C < ((v_H - v_S)/(v_H - v_L))\, p_S^C$–see KG. In this case, a reduction in auditing intensity may lead the agent to switch from an "up-or-down" experiment to a (in this case less informative) status-quo experiment. That is, under certain conditions, reducing auditing intensity may reduce experimentation. The principal can however replicate the environment with binary decisions if she could rule-out the intermediate option (in this case the status-quo). In fact,

[24] $\mathbb{I}_A$ 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 $x \notin A$.

the principal benefits from ruling out decisions–thus reducing her discretion when responding to the agent's report–as now a lower auditing intensity always improves experimentation. If she can commit to ruling out certain decisions, then the optimal data governance in proposition 7 still holds for this more complex environment.

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 et al., 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).[25] 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

[25]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.

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. In our model, the principal can audit the agent's report but she cannot directly control the experiment. In contrast, Friedman et al. (2022) consider a regulator who could mandate full disclosure of a certain type of information but, under some conditions, chooses to mandate a less informative report to avoid crowding out firm's voluntarily disclosure of other information.

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 et al., 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 et al., 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 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 et al. (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.[26] 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

[26]Our communication stage is also related to models of communication with lying costs—e.g., Kartik et al. (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.

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.[27]

A main insight in these papers is that frictions in communication can be used to discipline agents if they underinvest in information acquisition.[28] 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.

Theoretical literature on Bayesian persuasion.

Our paper contributes to the growing literature on Bayesian persuasion following KG. Recent papers have applied versions of the Bayesian persuasion framework to study financial disclosure (Szydlowski, 2021), marketing and sales (Drakopoulos et al., 2021, Boleslavsky et al., 2017), the strategic disclosure of health-related information (Schweizer and Szech, 2018, Alizamir et al., 2020, de Véricourt et al., 2021), among many other topics.

Our model is most closely related to papers that relax the sender-commitment assumption in KG.[29] Papers in this recent literature differ on the modeling of imperfect commitment. For instance, Guo and Shmaya (2021) 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

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

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

[29]See also the literature on strategic sample selection, e.g., Tillio et al. (2017), Tillio et al. (2021),Hoffmann et al. (2020), Adda et al. (2020), Felgenhauer and Loerke (2017) and Libgober (Forthcoming).

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.[30] Instead, in our paper tampering is a generic equilibrium phenomenon resulting from the principal's choice of data governance. 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 et al. (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 et al. (2022) and Nguyen and Tan (2021). Lipnowski et al. (2022) consider an information design setup with output-tampering described by a (possibly state-dependent) credibility function specifying the likelihood that the sender can tamper at no cost and provide an elegant geometric characterization of the sender's value of persuasion. Similar to our result on credible skepticism, they show that the receiver can benefit from a sender with limited credibility–see their discussion on "productive mistrust." The key difference between our setups is the nature of this credibility function: it is exogenously specified in Lipnowski et al. (2022)[31] while in our setting it endogenously arises from the equilibrium incentives of the reporting agent. That is, in our setup the agent's tampering/credibility must be consistent with the principal's response to his messages. Nguyen and Tan (2021) also study public experimentation 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 (Con-

[30]Tampering-proof equilibria are the focus of Min (2020), while Guo and Shmaya (2021) 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".

[31]The sender can invest to select a more favorable state-dependent credibility function–see the working paper version Lipnowski et al. (2018)–albeit, once chosen, the likelihood of tampering is independent of the receiver's decision making.

dition 1 in Nguyen and Tan, 2021). 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.

Finally, one of our main contributions is to study how the principal (receiver) can optimally adapt certain organizational practices (namely, task allocation and data governance policies) to incentivize the sender to experiment more. Similarly, Alonso and Câmara (2016b) study how a committee (group of receivers) can optimally select certain practices (voting rules) to induce the sender to select more informative experiments. We believe that the effect of different organizational practices on endogenous experimentation is a promising research area.

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.

Alizamir, S., de Véricourt, F., & Wang, S. (2020). Warning against recurring risks: An information design approach. *Management Science*, *66*(10), 4612–4629.

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

Alonso, R., & Câmara, O. (2016b). Persuading Voters. *American Economic Review*, *106*(11), 3590–3605.

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.

Blackwell, D. A., & Girshick, M. A. (1954). *Theory of Games and Statistical Decisions*. John Wiley & Sons, Hoboken, NJ.

Blume, A., Board, O. J., & Kawamura, K. (2007). Noisy talk (2007-12-13). *Theoretical Economics*, *2*(4), 395–440.

Boleslavsky, R., Cotton, C. S., & Gurnani, H. (2017). Demonstrations and price competition in new product release. *Management Science*, *63*(6), 2016–2026.

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

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance? *https://ssrn.com/abstract=1819486*.

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.

de Véricourt, F., Gurkan, H., & Wang, S. (2021). Informing the public about a pandemic. *Management Science*, *67*(10), 6350–6357.

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.

Drakopoulos, K., Jain, S., & Randhawa, R. (2021). Persuading customers to buy early: The value of personalized information provisioning. *Management Science*, *67*(2), 828–853.

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* (tech. rep.). Working Paper. National Bureau of Economic Research.

Friedman, H. L., Hughes, J. S., & Michaeli, B. (2022). A rationale for imperfect reporting standards. *Management Science*, *68*(3), 2028–2046.

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). Princeton University Press.

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

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

Guo, Y., & Shmaya, E. (2021). Costly miscalibration. *Theoretical Economics*, *16*(2), 477–506.

Hoffmann, F., Inderst, R., & Ottaviani, M. (2020). Persuasion through selective disclosure: Implications for marketing, campaigning, and privacy regulation. *Management Science*, *66*(11), 4958–4979.

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* (tech. rep.). KPMG International Data Analytics.

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

Lennox, C., & Wu, X. (2022). Mandatory internal control audits, audit adjustments, and financial reporting quality: Evidence from china. *The Accounting Review*, *97*(1), 341–364.

Lewis, T. R., & Sappington, D. E. 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*.

Lipnowski, E., Ravid, D., & Shishkin, D. (2022). Persuasion via weak institutions. *Journal of Political Economy*, *130*(10), 2705–2730.

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* (tech. rep.). McKinsey Analytics.

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

McKinsey & Co. (2018b). *Building an effective analytics organization* (tech. rep.). 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. (2021). Bayesian persuasion with costly messages. *Journal of Economic Theory*, *193*, 105212.

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.

Schweizer, N., & Szech, N. (2018). Optimal revelation of life-changing information. *Management Science*, *64*(11), 5250–5262.

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

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

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.

Szydlowski, M. (2021). Optimal financing and disclosure. *Management Science*, *67*(1), 436–454.

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 $\pi = \{q, \Pr[q]\}_{q \in S(\pi)}$, with type space $S(\pi)$ which induces posterior $\Pr[\theta = 1 | s = q] = q$, consider a PBE of the communication subgame where the agent's reporting strategy is $m^*(s, c)$, and which leads to decisions $d^*_A(m, s)$ 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 an agent finds it profitable to send $s = q_z \neq q$ instead of $s = q$ after observing (q, c), he will strictly prefer to tamper after (q, c') if $c' < c$.

For part (ii), consider the set of tampered outcomes $M_T(\pi)$ defined in the proposition. Suppose that $q, q' \in M_T(\pi)$ but the distributions $d^*_U(q)$ and $d^*_U(q')$ lead to different expected payoffs for the agent.[32] If $\underline{q}_H \notin S(\pi)$, 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 $d^*_A(m, \underline{q}_H)$ is independent of m whenever $\underline{q}_H \in S(\pi)$. 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 $m^*(s,c)$ cannot be part of an equilibrium. Therefore, we must have that $d^*_U(q) = d^*_U(q')$ for $q, q' \in M_T(\pi)$. Finally, suppose that $q \in M_T(\pi)$, $q \neq \underline{q}_H$. Then, the agent 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 $\tilde{\pi} = \{q, \Pr[q]\}_{q \in S(\tilde{\pi})}$ with type space $S(\pi)$ which induces posterior $\Pr[\theta = 1|s = q] = q$ and the equilibrium reporting $m^*(q,c)$. 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 part of a PBE, as the designer's expected utility cannot be improved by any alternative $\tilde{\pi}$.

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

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

and recall that, from Proposition 1, $M_T(\tilde{\pi})$ is the set of tampered outcomes. Thus, the agent after observing $s = q \in S_T(\tilde{\pi})$ will tamper with positive probability while reporting $s = q' \in M_T(\tilde{\pi})$ with positive probability. Since $d^*_A(m, 0) = d_S$ whenever $s = 0 \in S(\pi)$, 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

[32] As the principal only mixes after an inconclusive audit when her posterior is $\underline{q}_H$, the agent must obtain a different expected payoff after an inconclusive audit when reporting q and q' if these distributions are different.

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

Next, partition $S(\tilde{\pi})$ into two sets $X_S(\tilde{\pi}) = S(\tilde{\pi}) \cap [0, \underline{q}_H)$ 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 after any outcome in $X_i(\tilde{\pi})$ the agent faces the same gain from tampering and must therefore have the same tampering threshold. Proposition 1-ii(a) implies that this is true if all outcomes in $X_i(\tilde{\pi})$ are tampered outcomes, i.e., 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 $q' \in X_i(\tilde{\pi})$ with $q' \notin M_T(\tilde{\pi})$. 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 $m = q_{M_T} \in M_T(\tilde{\pi})$ must be strictly lower than q_{M_T}. But then we must have that $d^*_U(q_{M_T}) = d_i$, otherwise the agent tampering 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 outcomes in $X_i(\tilde{\pi})$, $i = S, H$, with a single outcome $s = \tilde{q}^{X_i}$ 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 the agent sends message $m = \tilde{q}^{X_j}$ after observing $s = \tilde{q}^{X_i}$ when his cost 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, the designer can improve upon the binary experiment π with $S(\pi) = \{0, q\}$ 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 an approval probability τ after an unaudited $m = q$. The expected gain from tampering is then $(1 - \lambda)\,\tau\Delta$ and this establishes the tampering threshold $\bar{c} = (1 - \lambda)\,\tau\Delta$. 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

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

which, giving the Bayesian consistency constraint $(1 - p)\,q = \mu$, implies $p\bar{F}(\bar{c})\,\underline{q}_H \geq \underline{q}_H - \mu$, and, using (2), can be expressed as

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

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 π' with $S(\pi') = \{0, q - \epsilon\}$ 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 (4). Therefore, any binary experiment π with $S(\pi) = \{0, q\}$ 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 $V^* = V_S$.

For future reference, we express the designer's expected payoff from a status-quo experiment in terms solely of the induced equilibrium threshold $\bar{c}$. Using (10) with (11) and (12), we can write

$$v_S(\bar{c}) = \left(1 - p_S^C\right)\left(v_H - (1-\lambda)\left(1-\tau\left(\bar{c}\right)\right)\Delta\right) + p_S^C\left(v_S - \lambda\Delta\frac{F\left(\bar{c}\right)}{\bar{F}\left(\bar{c}\right)} - \mathbb{I}_{\{k=\mathcal{I}\}}\frac{\int_0^{\bar{c}} cdF(c)}{\overline{F}(\bar{c})}\right).$$

$$= v_H - (1-\lambda)\Delta + \bar{c} - p_S^C\left(\frac{\lambda\Delta}{\bar{F}\left(\bar{c}\right)} + \bar{c} + \mathbb{I}_{\{k=\mathcal{I}\}}\frac{\int_0^{\bar{c}} cdF(c)}{\overline{F}(\bar{c})}\right). \tag{A.4}$$

■

Proof of Proposition 3: Suppose that F admits a density so that $\int_0^{\bar{c}} cdF(c)/\overline{F}(\bar{c})$ is differentiable. Then, from (A.4), the difference in the designer's marginal payoff from a higher tampering threshold when moving from separation to integration is

$$\frac{\partial\left(v_S(\bar{c};\mathcal{I}) - v_S(\bar{c};\mathcal{S})\right)}{\partial\bar{c}} = -p_S^C\frac{d}{d\bar{c}}\left(\frac{\int_0^{\bar{c}} cdF(c)}{\overline{F}(\bar{c})}\right) \leq 0.$$

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

Proof of Proposition 4: First, consider experiment $\pi(\bar{c}) \in \Pi_S$ inducing tampering threshold $\bar{c}$, with $S(\pi(\bar{c})) = \{0, q\left(\bar{c}\right)\}$. From (A.4), we have

$$\frac{\partial v_S(\bar{c})}{\partial\lambda} = \Delta - p_S^C\frac{\Delta}{\overline{F}(\bar{c})} = \Delta\Pr\left[s = q\left(\bar{c}\right)\right],$$

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

Second, define the feasible set of tampering thresholds

$$\mathcal{C}_S \equiv [0, (1-\lambda)\Delta] \cap [0, \overline{F}^{-1}(p_S^C/p_S^{FI})], \tag{A.5}$$

with $p_S^{FI} = \Pr[s=1]$ for the fully informative experiment. To understand $\mathcal{C}_S$, note that $\bar{c} \in \mathcal{C}_S$ must satisfy two conditions. First, it must correspond to some approval probability $\tau \in [0,1]$—from (6) this implies that $\bar{c} \in [0, (1-\lambda)\Delta]$. Second, the experiment $\pi(\bar{c})$ must be feasible—i.e., $q(\bar{c}) \leq 1$—which requires $\bar{F}\left(\bar{c}\right)/p_S^C \geq \frac{1}{1-\mu} = 1/p_S^{FI}$ —see (4).

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

$$\max_{\bar{c}} v_S(\bar{c};\lambda),\ s.t.\ \bar{c} \in \mathcal{C}_S. \tag{A.6}$$

The feasible set $\mathcal{C}_S$ is increasing in the strong set order with respect to $-\lambda$ and, from part (i), $v_S(\bar{c};\lambda)$ is supermodular in $(\bar{c},-\lambda)$. Theorem 4' in Milgrom-Shannon (1994) then implies that the set of maximizers of (A.6) increases in the strong set order sense with $-\lambda$. From (4), for a fixed threshold $\bar{c}$ the experiment $\pi(\bar{c})$ is independent of λ, so the set of optimal experiments $c^*(\lambda)$ decreases in the strong set order sense with λ, implying that the set of designer-optimal Type I errors $\alpha^*(\lambda)$ increases in the strong set order sense with λ. ■

Proof of Proposition 5: Applying (13), the principal's equilibrium expected utility is given by (14). Note that for $\lambda = 1$ the designer always selects the commitment experiment, thus, regardless of the task-allocation, $\bar{c}^* = 0$. By the definition of designer's responsiveness to auditing, there exists $0 < \lambda < 1$ with $\alpha^*(\lambda) < \alpha_H$–implying that $\bar{c}^* > 0$ 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 that $\alpha^*(\lambda^*) < \alpha_H$ and 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 $\lambda < 1$ and the designer selects the commitment experiment. In other words, $\tau = 0$ for experiment with $\alpha = \alpha_H$ and $\bar{c} = 0$. Using (A.4), the designer's payoff is then

$$v_S(0) = v_H - (1-\lambda)\Delta - p_S^C \frac{\lambda\Delta}{\overline{F}(0)}.$$

We now study conditions such that (a) there exists an experiment that leads to a positive approval probability, and (b) the designer's incremental payoff from such experiment 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 $F(0)$. The experiment that induces the highest posterior if unaudited is the fully informative experiment for which $\Pr[s=0] = 1-\mu \equiv p_S^{FI}$. Therefore, there exists an experiment with a positive approval 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 (6) that $\bar{c} = \tau(1-\lambda)\Delta$, if F admits a density we can differentiate (A.4) to obtain

$$\begin{aligned}\left.\frac{\partial v_S(\tau)}{\partial \tau}\right|_{\tau=0} &= \left.\frac{\partial v_S(\bar{c})}{\partial \bar{c}}\frac{\partial \bar{c}}{\partial \tau}\right|_{\tau=0} \\ &= \left|(1-\lambda)\,\Delta\left(\left(1-p_S^C\right)-p_S^C\frac{f\left(\bar{c}\right)}{\left(\overline{F}(\bar{c})\right)^2}\lambda\Delta\right)\right|_{\bar{c}=0}.\end{aligned}$$

Therefore, the condition $\partial v_S(\tau)/\partial\tau|_{\tau=0} > 0$ translates to $\left(1-p_S^C\right)/p_S^C\Delta > \lambda f\left(0\right)\left(\overline{F}(0)\right)^2$ and we can always find a $0 < \lambda < 1$ satisfying this condition. ■

Proof of Lemma 1: Program (A.6) defines the designer's problem and the feasible set of tampering thresholds $\mathcal{C}_S$ is defined in (A.5). Setting $k = \mathcal{S}$ in (A.4), then the marginal payoff from increased tampering, whenever it exists, satisfies

$$\frac{\partial v_S(\bar{c})}{\partial \bar{c}} = \left(1-p_S^C\right) - \lambda p_S^C\Delta\frac{f(\bar{c})}{\left(\overline{F}(\bar{c})\right)^2} = \lambda p_S^C\Delta\left(\frac{\phi_S}{\lambda} - L(\bar{c})\right).$$

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

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

■

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

$$\lambda \leq \frac{1-p_S^C}{1-p_S^C+p-p_S^C} \equiv \widetilde{\lambda}_S(p). \tag{A.7}$$

To see this, we express $v_S(\bar{c}\left(p\right);\mathcal{S})$ as a function of p: using (4) we have $\bar{F}\left(\bar{c}(p)\right) = p_S^C/p$

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

and we can write (A.4) for $k=\mathcal{S}$ as

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

Designer's optimality of π_S requires

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

The first condition follows from $\overline{c}(p)\leq(1-\lambda)\,\Delta$, as the gain from tampering is bounded by $(1-\lambda)\,\Delta$, 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

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

from which we obtain (A.7). Inverting (A.7), any experiment that is induced with auditing intensity $\lambda>1/(2-\underline{q}_H)$ must satisfy

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

Taking into account (6) and that (A.3) is satisfied with equality for status-quo experiments, we can express the Type I error α associated with an experiment that induces $\Pr[s=0]=p$ as

$$\alpha=1-(1-\alpha_H)\frac{p}{p_S^C}.$$

Using (2), the minimum achievable Type I error with an auditing of λ, $\underline{\alpha}(\lambda)$, is

$$\begin{aligned} \underline{\alpha}(\lambda) &= 1-(1-\alpha_H)\frac{p(\lambda)}{p_S^C}=1-\frac{p(\lambda)}{1-\mu}\\ &= \frac{2\lambda-1}{\lambda}\alpha_H-\frac{\mu(1-\lambda)}{(1-\mu)\,\lambda}. \end{aligned}$$

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

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

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

Using the identity $\overline{F}^{-1}(p_S^C/p')=F^{-1}(\left(p'-p_S^C\right)/p')$ and noting that $p(\lambda)$ satisfies $\lambda(p(\lambda)-p_S^C)=(1-p_S^C)(1-\lambda)$, we can simplify the previous expression to

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

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

$$F\left(c\right)\geq\frac{c}{c+\lambda\frac{p_S^C\Delta}{1-p_S^C}}\text{ for }c\leq\lambda\Delta\frac{p(\lambda)-p_S^C}{1-p_S^C}=(1-\lambda)\Delta. \tag{A.8}$$

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.8) is supported only on two cost realizations, 0 and $(1-\lambda)\Delta$, with

$$\Pr\left[c=0\right]=\left(p(\lambda)-p_S^C\right)/p(\lambda), \tag{A.9}$$

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

Proof of Proposition 7: From (13), for each $\pi_S(\lambda)$, with $S(\pi_S(\lambda))=\{0,q(\lambda)\}$ and $p(\lambda)=\Pr[s=0]$, auditing λ, and cost distribution satisfying (A.8), the principal's utility is

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

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.9), we obtain $\Pr\left[c=0\right]=\alpha_H$ for the cost distribution supported on 0 and $\widetilde{c}=(1-\lambda_{opt})\Delta=(1-\underline{q}_H)\Delta/(2-\underline{q}_H)$; 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. ■