CESifo
AREA CONFERENCES 2022

Energy & Climate Economics
4 – 5 March 2022

Compliance and Truthfulness: Leveraging Peer Information with Competitive Audit Mechanisms

Timo Goeschl, Marcel Oestreich, and Alice Soldá

CESifo

Compliance and Truthfulness: Leveraging Peer Information with Competitive Audit Mechanisms*

Timo Goeschl
Heidelberg University

Marcel Oestreich
Brock University

Alice Soldà
Ghent University

February 1, 2022

Abstract

How to design audit mechanisms that harness the benefits of self-reporting for achieving compliance with regulatory targets while limiting misreporting is a pressing question in many regulatory contexts, from climate policies to public health. We theoretically and experimentally study the relative performance of a competitive audit mechanism in regulating a socially undesirable activity when peer information about others' activity levels is present or absent. Using random auditing as a benchmark, the analytical framework predicts always more truthful self-reporting under the competitive mechanism and, in the presence of peer information, fully compliant activity levels. A quasi-laboratory experiment confirms the core theoretical findings: Under the competitive audit mechanism, reports are more truthful on average and, in the presence of peer information, participants' average activity levels, while not reaching full compliance, are closer to the social optimum. The experimental results highlight the considerable potential of competitive audit mechanisms for achieving regulatory targets.

Keywords: Regulation; compliance; tournament theory; quasi-laboratory experiment.

JEL classification: D62; H41; H83; L51; Q58.

*We are grateful to Tim Cason, Jay Shimshack, John Stranlund, seminar audiences at the Universities of Bayreuth, Heidelberg, Mannheim and ZEW-Mannheim, and audiences from AERE, EAERE and CEA conferences for helpful comments; to Tobias Pfrommer for excellent research assistance, and to the GATE Lab for participant pool access. This research has been funded by the Federal Ministry of Education and Science (grant no. 01LA1806A). Oestreich acknowledges financial support by the Social Science and Humanities Research Council of Canada (grant no. 435-2017-463).

1 Introduction

In many policy areas, regulated entities are obligated to self-report the level of a regulated activity to a regulator, broadly defined (Innes, 2017): Employees self-report task completion (Allen Jr and Bunn III, 2003) or workplace incidents (Probst and Estrada, 2010) to employers. Firms self-report emissions and release of regulated pollutants to environmental protection agencies (Malik, 1993; Helland, 1998). In Europe, members of the public self-reported border crossings and compliance with quarantining regimes to public health authorities during the COVID19-pandemic (Burns et al., 2020). Self-reporting acknowledges that the regulated entities tend to be better informed about their own activity levels than the regulator, who typically has to incur considerable monitoring costs in order to observe the activity (Harford, 1987). Yet, self-reporting is obviously no perfect remedy for the information asymmetry between regulator and regulated: Entities can strategically misreport their activities unless dissuaded to do so by the threat of a truth-revealing regulatory audit coupled with fines (De Marchi and Hamilton, 2006). How to design efficient regulatory systems that harness the benefits of self-reporting while limiting misreporting is therefore a question of interest to academics and to regulators charged with enforcing policies, but constrained both by tight budgets for conducting costly audits and by limited enforcement options (Harford, 1987; Friesen and Gangadharan, 2013).

Progress in designing such systems has been considerable. This is true both with respect to the primary objective of aligning activity levels with the regulatory target and with respect to the secondary objective of encouraging more truthful self-reporting. The context of environmental regulations is a case in point (Cason et al., 2020): There, the typical setting features the emission of regulated pollutants as the regulated activity, firms as the regulated entities, and self-reported emissions as the information to be communicated to the regulator, who then levies fees on reported emissions. In this context, both theoretical and experimental results have demonstrated the potential of harnessing self-reporting: Gilpatric et al. (2011)

were the first to show that compared to a random auditing mechanism (RAM), self-reporting is more truthful when regulators employ competitive audit mechanisms (CAMs) that condition audit probabilities on self-reported activity levels.[1] The authors model CAMs as rank order tournaments, wherein the regulator's expectation of firms' emissions is subject to error and the regulator will audit those firms for which the difference between expected and reported emissions is greatest. Gilpatric et al. (2015) extend this finding to a dynamic setting. Cason et al. (2016) experimentally affirm the prediction of more truthful self-reporting under CAMs. They also find a contribution of CAMs towards emissions reduction that is not predicted by their theoretical model. Such a theoretical prediction comes from Oestreich (2017) who finds that in contrast to RAMs, there are CAMs that can implement the socially optimal level of emissions when firms have perfect peer information about each others' emissions. This is an extension of Oestreich (2015) who shows that CAMs can lead to higher or lower activity levels than RAMs, depending on their exact specification. There is also emerging empirical evidence that CAMs can enhance regulatory performance in the field (Earnhart and Friesen, 2021). Moving to the context of tax compliance, Bayer and Cowell (2009) and Vossler and Gilpatric (2018) likewise find that CAMs can enhance tax compliance by firms and individuals. Jointly, these findings raise the possibility that CAMs help budget-restricted regulators to meet not just the secondary objective of truthful self-reporting, but also realize the primary target of socially desired activity levels as long as peers are well informed about each other.

The present paper uses a combination of theory and experiments to extend the literature in two directions. One direction is to focus on the impact of the mechanisms on the activity levels of regulated entities, rather than only on their self-reporting. This focus is motivated both by the importance of activity levels as the primary regulatory target and provocative theoretical arguments that a properly designed CAM is capable of inducing socially optimal activity levels among the regulated (Oestreich, 2017). Cason et al. (2020) point out that

[1] CAMs are also sometimes referred to as "relative conditional audit mechanisms" (Cason et al., 2020), a terminology that captures more directly their underlying logic.

experimental evidence regarding activity levels is particularly scarce. The second direction of extension is to explore explicitly the role of peer information structures in determining activity and reporting outcomes under RAMs and CAMs. Peer information structures matter not only for key results in the literature, but also differ substantially in seemingly similar settings: Self-reporting employees may work side by side on the factory floor – or shirk in isolation in their home offices. Self-reporting polluters may be farmers crop-spraying adjacent fields – or firms that have little insight into each other's operations. The self-reporting citizens crossing a border may be members of a sports team – or complete strangers. A meaningful comparison of regulatory outcomes under different information structures requires a unifying theoretical framework and an experimental design in which information structures can be exogenously manipulated in isolation.

In extending the literature in these directions, the paper makes two main contributions. First, it presents an analytical framework that is able to generate insights isomorph to the main theoretical results in Gilpatric et al. (2011), Cason et al. (2016), and Oestreich (2017). It does so with three aims in mind: to clarify the comparative performance of the CAM and the RAM in terms of both activity levels and self-reporting in the presence and absence of peer information; to facilitate the design of an economic experiment that is closely aligned with the theory; and to generate hypotheses that are testable in that experimental environment.

We demonstrate that the core findings of the literature can be replicated in this framework: Self-reporting is more truthful when regulators employ CAMs compared to RAMs regardless of the peer information structure as in Gilpatric et al. (2011). In the absence of peer information about activity levels, the CAM induces the same level of actual activity as the RAM, but leads to more truthful self-reporting as in Cason et al. (2016). In the presence of perfect information about each others' activity levels, the CAM leads to more truthful self-reporting than the RAM and induces the socially optimal activity level while RAM does not as in Oestreich (2017). The intuition for the latter result relies on the strategic interaction between an agent's self-reporting and others' activity levels, which is observable when peer information is present. In equilibrium, the direct effect of increasing the activity level is exactly counterbalanced by the strategic effect that the observable increase has on the self-reporting of others. These findings constitute the core testable hypotheses that predict the presence and direction of differences in activity levels and reporting patterns between the CAM and the RAM depending on whether peer information is present or absent.

The second contribution is the demonstration, in a quasi-laboratory experiment with 131 participants, that all but one of the core hypotheses survive testing in a controlled setting. In this setting, participants play seven rounds of a game that mimics the unifying theoretical model. In each round, participants are (re-)matched in groups of three and make two individual decisions: activity level and self-reported level. Each self-reported unit incurs a fixed fee; each non-reported unit of activity incurs a penalty if the participant is audited. After each round, exactly one of the three participants is audited. There are two treatment dimensions, the audit mechanism (RAM or CAM) and the participant's information about other group members' activity levels (No Information or Perfect Information). In the RAM condition, the individual audit probability is fixed and uniform. In the CAM condition, the individual audit probability depends on the participant's self-reported level relative to the reports of the other two participants in the group.

We show that in the treatment condition without peer information (NI), mean activity levels in the CAM and the RAM condition are statistically indistinguishable. The prediction that the CAM does not outperform the RAM in regulating activity levels is therefore borne out, but so is the prediction that reporting under the CAM is more truthful. There, players self-report significantly higher activity than under the RAM. In the treatment condition with peer information (PI), self-reports in the CAM condition are not only more truthful, but activity levels are also more aligned with regulatory targets relative to the RAM. To our understanding this is the first experimental evidence to confirm these theoretical predictions, and the experiment generates evidence to corroborate the underlying causal mechanism. Yet, we also find that activity levels under the CAM do not align perfectly with regulatory targets and explore candidate explanations for this deviation.

In the following section, the paper develops the analytical framework that culminates in the derivation of three main testable hypotheses. The experimental design and its parametrization are presented in section 3 and lead to the concrete experimental predictions. Section 4 reports the results of the experiment and of our statistical tests. Section 5 discusses the experimental evidence and section 6 concludes with a summary and research outlook.

2 Theoretical Framework

The theoretical framework presented in this section serves a dual purpose. One is integrating the existing insights from the literature in a parsimonious framework that allows activity levels to be endogenized and variations in peer information to be captured. The other is developing a framework that can bridge into the experimental laboratory by allowing testable hypotheses to be generated and by readily translating into an experimental design.

To serve this dual purpose, we present a parsimonious analytical framework based on a generalization of Oestreich (2017). The basic strategic interactions in this framework are those described by Gilpatric et al. (2011) and Gilpatric et al. (2015) in that regulated agents self-report their activity level and the audit decision by the authorities does (CAM) or does not (RAM) depend on the relative comparison of the firms' reports. In line with Cason et al. (2016) we implicitly analyze a simultaneous move game whereby players make their choices about the activity level and self-reporting at the same time (or without observing the others' actions) in one variant of our model.

While our analytical framework is rooted in the previous literature, we also extend it in several ways. Building on Gilpatric et al. (2011) and Gilpatric et al. (2015) we allow not only for endogenous self-reporting, but also for endogenous activity choices. Furthermore, our analytical framework captures two variations in the peer information structure: one wherein firms have perfect peer-information about each other's activity levels as in Oestreich (2017) and another wherein firms have no peer-information as in Cason et al. (2016). To accommodate both information structures, the solution concept in our paper is the equilibrium of a sequential move game whereby players observe the activity levels of their competitors first before self-reporting their own activity. This coincides with the simultaneous-move equilibrium when peer information is absent.

Setting. Consider a setting in which a regulator is charged with enforcing a fee-based regulation for n regulated agents. The n agents choose a privately beneficial, but externality-generating activity whose level is denoted by e_i. The risk-neutral agents accrue benefits from the activity captured by the benefit function $g(e_i)$. All agents have the same benefit function which is motivated by the common practice of enforcement agencies to group agents according to observable characteristics such as risk, size and industry before allocating audit resources as described in Telle (2015). The benefit function is assumed to be strictly concave with a maximum at e^0. Hence, in the absence of regulation, agents choose the maximum beneficial activity level, i.e. $e_i = e^0$, for all $i \in n$ where the marginal benefits are zero, that is $g'(e^0) = 0$.

Given their activity level e_i, agents self-report activity r_i and pay a linear fee t for every self-reported unit. The fee level t is exogenously set by a higher authority. A candidate for the fee level would be one that follows Pigovian principles and equates the fee level with the social marginal cost. Accordingly, we expect the socially efficient actvity level to be e^t, which is implicitely defined by $g'(e^t) = t$. Activity levels may be under-reported by the agents to reduce fee payments. If so, $e_i - r_i$ is the amount of under-reported activity level by agent i.

Regulator. The regulatory agency is charged with enforcing the fee system. It can only observe the chosen activity level by agents after conducting a costly audit. Its operating budget is fixed including the resources allotted to conducting audits. Let K be the number of agents which the regulator can afford to audit, where $K \leq n$. Let $k \equiv K/n$ define the audit rate. If the regulator decides to increase the audit probability for one agent, it has to decrease the audit probability of at least one other agent in order to keep its budget at balance. Specifically, we have at all times that the assigned audit probabilities add up to the number of total audits: $\sum_{i=1}^{n} p_i = K$.

Let $\mathbf{r} = (r_1, ..., r_n)$ denote the vector of reported activity levels for the n agents. The regulator's problem is to decide on how to spread a limited auditing budget across the n agents. Accordingly, we define an audit mechanism as a strategy for the regulator to assign an audit probability p_i to every regulated agent i. The announced mechanism is represented by the function $p_i : (r_1, ..., r_n) \to [0, 1]^n \; \forall i \in n$ that maps the vector of agents' activity-reports into agents' audit probabilities. We assume that the regulator knows the unregulated activity level e^0, possibly from the time without the regulation, and can use it as reference value when designing the audit mechanism $p_i(\mathbf{r})$.[2] After an audit, the regulator can perfectly observe the actual activity level chosen by the audited agent and levy a linear penalty θ per unit of under-reported activity level, where $\theta > t$.

Information structure. How much agents know about each others' activity levels varies considerably in the field. Here, we consider the two limit cases of perfect peer information (Assumption PI) and of no peer information (Assumption NI) and separately analyze their effects on agents' activity levels and self-reporting.

Assumption NI Agents have *no information* about each others activity levels.

Assumption PI Agents have *perfect information* about each others activity levels.

Timing. The multistage game between regulator and agents consists of the following four stages:

- In the *first stage*, the regulator announces an audit mechanism $p_i : (r_1, ..., r_n) \to [0, 1]^n$ which maps activity-reports into audit probabilities for each agent upon receiving the

[2] This assumption is without loss of generality. Specifically, the audit mechanism requires the regulator to use some arbitrary high emission level as reference value. The unregulated activity level e^0 seems to be a natural candidate for this reference value, but other values would also work.

reports.

- In the *second stage*, agents choose the activity level e_i. Agents are not informed (Assumption NI) or perfectly informed (Assumption PI) about the activities of the other agents.

- In the *third stage*, agents choose activity reports r_i that are submitted to the regulator.

- In the *fourth stage*, some of the agents are audited according to the announced audit mechanism. A linear fine θ is levied for every unit of under-reported activity levels $[e_i - r_i]$ detected in an audit.

Agent's problem. Agent i chooses activity level e_i and activity report r_i in order to maximize expected profits:[3]

$$\max_{e_i \geq 0,\ r_i \leq e_i} \mathbb{E}\Pi_i(\mathbf{e}, \mathbf{r}(\mathbf{e})) = g(e_i) - tr_i - p_i(\mathbf{r}(\mathbf{e}))\theta[e_i - r_i] \ \forall i \in n, \tag{1}$$

where $\mathbf{e}$ denotes the vector of activity levels and $\mathbf{r}$ denotes the vector of reports chosen by all agents. Activity levels provide benefits to the agent through $g(e_i)$ and their expected cost is determined endogenously by the fee t on activity report r_i, the agent's individual audit probability p_i and penalty θ for potentially under-reported activity levels $[e_i - r_i]$.

Random audit mechanism. The random audit mechanism (RAM) is the common benchmark in the literature. The RAM allocates equal audit probabilities among all agents regardless of reports such that $p_i = k \ \forall i \in n$. It is well known that in theory, the RAM can fully enforce fees on the regulated activity as long as the expected marginal cost of under-reporting, $k\theta$ is larger than or equal to the fee rate, t. In that case, agents have no beneficial alternative

[3]In line with regulatory practice, over-reporting is not rewarded. Therefore, rational agents never over-report, that is $r_i \leq e_i$. Hence, without loss of generality, we can set $\max\{\theta(e_i - r_i), 0\} = \theta(e_i - r_i)$, and restrict the set of reported activity levels to be $r_i \leq e_i$.

but to report truthfully. Knowing it is going to pay fees on all of its activity, an agent chooses socially efficient activity levels e^t. Thus, for sufficiently large expected fines, $k\theta \geq t$, the regulator can fully enforce truthful reporting where $r_i = e_i$ and implement the socially efficient activity level $e_i = e^t \; \forall i \in n$.

Regulatory reality is typically characterized by limited auditing budgets and capped fines. Therefore, much of the literature focuses on the case in which the expected fine is below the fee, $k\theta < t$. In this case, the RAM induces neither truthful reporting nor socially efficient activity levels because it is cheaper for the regulated agents to under-report activity levels (evade fees t) and instead face the expected penalty $k\theta$.

Proposition 1 (Activity and reporting under RAM) *If $k\theta < t$, the random audit mechanism induces zero activity-reporting, i.e.: $r_i = 0 \; \forall i \in n$ and per-agent activity level $e_i = e^{k\theta}$, which is implicitly defined by:*

$$g'(e^{k\theta}) = k\theta \quad \textit{for} \;\; \forall i \in n. \tag{2}$$

Activity levels and self-reports are independent of the information structure.

Proof. See Appendix A.1.

The predictions of zero activity-reporting and excessive activity levels under capped fines and insufficient auditing budgets under the RAM motivate the search for more sophisticated audit mechanisms such as CAMs that can harness strategic interactions among the regulated agents to gain auditing leverage.

Competitive audit mechanism. The literature on CAMs is steadily growing focusing on the following set of features that CAMs tend to have in common:

- CAMs are applied in settings in which regulated agents self-report activity levels to a regulator.

- CAMs *decrease* the audit probability of an agent if the agent *increases* its reported activity levels: $\partial p_i(\mathbf{r})/\partial r_i < 0 \ \forall i \in n$,
- CAMs *increase* the audit probability of an agent if another agent *increases* its reported activity level: $\partial p_i(\mathbf{r})/\partial r_j > 0 \ \forall j \neq i \in n$,
- CAMs keep the regulator budget balanced: If the regulator increases the audit probability for one agent, it has to decrease the audit probability of at least one other agent: $\sum_{i=1}^{n} \partial p_i(\mathbf{r})/\partial r_j = 0 \ \forall j \neq i \in n$,

We adopt the CAM proposed by Oestreich (2017) that shares these features and is capable of inducing the socially optimal activity level in equilibrium such that $e_i = e^t \ \forall i \in n$. According to this CAM, the regulator allocates the audit probabilities according to the following audit mechanism:[4]

$$p_i(\mathbf{r}) = k + \lambda \ln\left(\frac{(R_i)^{n-1}}{\prod_{j\neq i}^{n}(R_j)}\right), \tag{3}$$

where $R_i = e^0 - r_i$ and e^0 serves as a reference value for the regulator to compare reports against.[5] Parameter λ determines the *degree of competitiveness* induced by the CAM. Specifically, the audit probability of agent i changes by λ given a one percent increase in the report of agant i. If $\lambda = 0$, random auditing results where $p_i = k \ \forall i \in n$. If $\lambda > 0$, the audit mechanism is competitive in that higher reports relative to other agents result in lower assigned audit probabilities. The CAM adopted here relies on the special case where $\lambda = (t/\theta - k)/((n-1)(2-N))$, and $N = \left(n - 2 + \sqrt{n^2 + 4n - 4}\right)/(2\,(n-1))$. This specific functional form for $p_i(\mathbf{r})$ can induce the optimal activity level e^t for all agents as shown below.

One interesting feature of this CAM is the inverse relationship between the size of relative audit budget of the regulator (measured by the difference between t/θ and k) and the degree

[4]As usual, the audit probabilities are bound by zero and one. Formally, the audit probability is $\min[\max[p_i, 0], 1]$.

[5]This specific contest success function adds noise through the probabilistic selection of the audited firms similar than a standard Tullock contest. Please refer to Oestreich (2015) for the formal comparison of the performance for enforcement of the Tullock and the more competitive contest format, namely the all-pay auction.

of competitiveness induced by the CAM. This implies that for audit budgets of sufficient size ($t/\theta \leq k$), the CAM coincides with a fully enforcing RAM that implements the optimal activity level and truthful reporting.

Illustrative example. To aid intuition, figure 1 illustrates the allocation of audit probabilities under the adopted CAM for a simple example. There are two agents ($n = 2$) and a regulatory budget sufficient for a single audit ($K = 1$). In that case, the interior part of the audit function for agent 1 simplifies to:

$$p_1(r_1, r_2) = \frac{1}{2} + \lambda \ln(\frac{R_1}{R_2}),$$

with $R_1 = e^0 - r_1$, $R_2 = e^0 - r_2$ and $\lambda = (t/\theta - 1/2)/((2 - \sqrt{2}))$. The lines in Figure 1 trace out the audit rates p_1 and p_2 as a function of reports r_1 and r_2. The report of agent 2 is fixed at the equilibrium value $r_2 = r_2^*$ and the report of agent 1, r_1, varies along the horizontal axis. When the reports coincide ($r_1 = r_2$), the audit probabilities also coincide ($p_1 = p_2 = 1/2$). If r_1 is increased, p_1 decreases and p_2 increases.

Equilibrium concept. Agents are symmetric. We therefore conjecture that there is a symmetric equilibrium in pure strategies, where $e_i = e_j$, $r_i = r_j$ and $p_i = k\ \forall j \neq i \in n$. For a discussion about the existence of the symmetric equilibrium in pure strategies, we refer to Oestreich (2017). We show in Appendix B that the symmetric equilibrium in pure strategies exists for the case of perfect and no information between agents using the specific set of parameters of the subsequent experiment. As usual, the game is solved by way of backwards induction.

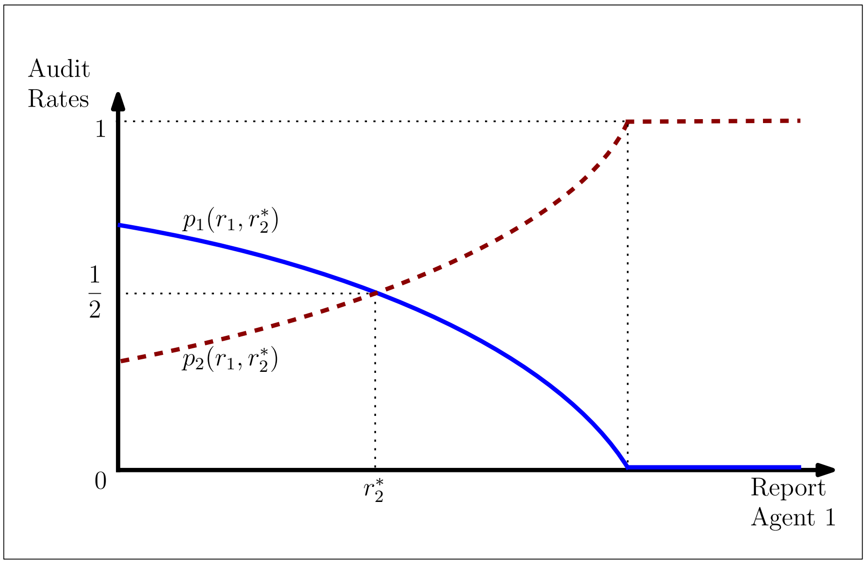


Figure 1: Optimal audit rates p_1 and p_2 for $n = 2$ and $K = 1$ as a function of report r_1 by agent 1, given equilibrium reporting $r_2 = r_2^*$ by agent 2.

2.1 Stage 4: Audits and enforcement

The implementation of the audit mechanism and the imposition of fines is automatic. There are no choices to be made by either regulator or agents.

2.2 Stage 3: Reporting equilibrium

In this stage, agents simultaneously choose reports in order to minimize the total cost of their chosen activity level given the announced audit mechanism, their own activity level and the other agents' activity reports. Agents' reporting choices take into account both the fees on reported activity levels and the potential for audit and enforcement in stage 4.

Differentiating profit function (1) with respect to r_i yields the first-order condition (FOC) for an interior reporting solution $(\partial \mathbb{E}\Pi_i(\mathbf{e}, \mathbf{r}(\mathbf{e}))/\partial r_i = 0)$ – denoted by r_1^*. This FOC can be re-written as:

$$\underbrace{p_i\theta}_{\substack{\text{direct}\\ \text{MB}}} \; \underbrace{+\lambda(n-1)\theta(\frac{e_i - r_i^*}{e^0 - r_i^*})}_{\substack{\text{indirect}\\ \text{MB}}} = \underbrace{t}_{\text{MC}}, \; at \; r_i = r_i^* \in [0, e_i]. \tag{4}$$

For interior solutions ($0 < r_i^* < e_i$), the first-order condition (4) has an intuitive interpretation: The marginal cost (MC) of reporting another unit of activity is the unit fee t. The marginal benefit (MB) of reporting has a *direct* and an *indirect* component. The direct benefit of increasing reported activity by one unit is that it reduces under-reporting by one unit, thus decreasing the expected fine by $p_i\theta$. Its indirect benefit is that, holding other agents' reports constant, the agent is less likely to be audited when reporting higher activity level. This, in turn, lowers the expected fine for the remaining under-reported activity levels by $-(\partial p_i/\partial r_i)\theta(e_i - r_i)$, with $\partial p_i/\partial r_i = -\lambda(n-1)/(e^0 - r_i)$ under the CAM in (3). This indirect effect is responsible to inducing agents to report some of their activity level while they would report zero under the RAM, i.e. when $\partial p_i/\partial r_i = 0$.

Proposition 2 (Reporting under CAM) *The competitive audit mechanism (3) induces a symmetric reporting equilibrium given by:*

$$r_i^*(\mathbf{e}) = \frac{e_i^* - e^0(2-N)}{N-1} \; \forall i \in n, \tag{5}$$

where $N = \left(n - 2 + \sqrt{n^2 + 4n - 4}\right)/(2\,(n-1))$ *and* e_i^* *is the equilibrium activity level. As long as* $e_i^* - e^0(2-N) > 0$*, the reporting equilibrium under the CAM is positive and thus larger than the reporting equilibrium under the RAM.*

Self-reports are independent of the information structure.

Proof. See Appendix A.2.

A sufficient condition for non-zero reporting in equilibrium is that the marginal benefits of activity ($g'(e)$) are high enough and/or the unit fee on activity (t) low enough such that regulation does not suppress the equilibrium activity level (e_i^*) too far below the unregulated activity level (e^0). How far regulation can suppress equilibrium activity without inducing zero reporting depends, through N, on the number of agents n: The more agents the regulator has to regulate, the higher the marginal benefit of activity (or the lower the unit fee) needs to be in order to ensure that agents still report any positive activity level $r_i^* > 0$. In practice, n is likely to be small, given that only firms of sufficient comparability can be subject to the CAM.

2.3 Stage 2: Activity equilibrium

In this stage, agents simultaneously choose activity levels while considering how their choices translate into the reporting equilibrium at stage 3 given the audit mechanism $p_i(.)$ and the other agents' activity levels. It is in this stage that assumptions about the presence or absence of information about peers' activity levels can affect agents' choices, requiring an analysis of both cases.

For ease of exposition, we start with the case of perfect peer information (Assumption PI). As in Tirole (1988), the determination of optimal activity levels relies on the total derivative of expected profits $\mathbb{E}\Pi_i(\mathbf{e}, \mathbf{r}(\mathbf{e}))$ with respect to e_i and an application of the envelope theorem. From the optimization at the reporting stage we know that $\partial\mathbb{E}\Pi_i/\partial r_i = 0$. The effect of e_i on $\mathbb{E}\Pi_i$ through the agent's own reporting choice is therefore irrelevant for the determination of optimum activity levels.

Under the conjecture of a symmetric pure-strategy equilibrium, the starting point of any deviation from equilibrium play is $e_1 = ... = e_n$. Here we consider a deviation of agent i such that $e_i \neq e_1 = .. = e_{i-1} = e_{i+1} = .. = e_n$. In this case, it must be true that the strategic effects of all other agents are identical. Thus, the total derivative is:

$$\frac{d\mathbb{E}\Pi_i}{de_i} = \underbrace{\frac{\partial E\Pi_i}{\partial e_i}}_{\substack{\text{direct}\\ \text{effect}}} + \underbrace{(n-1)(\frac{\partial E\Pi_i}{\partial r_j}\frac{\partial r_j}{\partial e_i})}_{\substack{\text{strategic}\\ \text{effects}}},\ \forall j = k \neq i \in n. \tag{6}$$

Using (6) and the particular profit function in (1), the FOC for a profit maximum is:

$$\underbrace{g'(e_i) - p_i\theta}_{\substack{\text{direct}\\ \text{effect}}} - \underbrace{(n-1)(\frac{\partial p_i}{\partial r_j}\frac{\partial r_j}{\partial e_i}\theta(e_i - r_i^*))}_{\substack{\text{strategic}\\ \text{effects}}} = 0,\ \forall j = k \neq i \in n. \tag{7}$$

The LHS of expression (7) contains two main parts, the *direct effect* and the *strategic effect* of varying activity levels on profit. The direct effect of changing e_i consist of the marginal benefit of varying activity levels ($g'(e_i)$) net of expected fines ($p_i\theta$). The strategic effect depends on the peer information structure. Since agents observe others' activity levels (Assumption PI), a change in e_i not only changes the agent's own reporting behavior, but also the other agents' reporting behavior (via $(\partial r_j/\partial e_i)(n-1)$). The change in the other agents' reporting behavior affects the audit probability of agent i, p_i, which in turn affects agent i's expected fine of unreported activity levels (in proportion to $(\partial p_i/\partial r_j)\theta(e_i - r_i^*)$). The total effect of e_i on $E\Pi_i$ is the sum of the direct and strategic effects.

At the point of symmetry ($e_i = e_j$ and $r_i = r_j$, $p_i = k$, $\frac{\partial p_i}{\partial r_i} = \frac{\partial p_j}{\partial r_j}$ and $\frac{\partial r_j}{\partial e_i} = \frac{\partial r_k}{\partial e_i}$), we can re-write the FOC as:

$$g'(e_i) = k\theta + \frac{\partial r_j}{\partial e_i}(t - k\theta),\ \text{at } \forall j = k \neq i \in n. \tag{8}$$

We can learn from (8) that if $\partial r_j/\partial e_i = 1$ at $e_i = e_j$, then as a result we get $g'(e_i) = t$, i.e. a necessary condition for socially efficient activity levels in equilibrium holds. In other words, if an audit mechanism induces all other agents j to increase their report by one unit when agent i increases her activity level by one unit, then this audit mechanism may implement efficient activity levels among all agents. This is precisely what the CAM in (3) achieves under Assumption PI.

Proposition 3 (Activity under CAM - perfect information) *Given that agents have perfect information about each other's activity levels (Assumption PI), the competitive audit mechanism (3) induces socially efficient activity levels for all agents, i.e.*

$$e_i = e^t \qquad \forall i \in n.$$

Proof. See Appendix A.3.

The other limiting case of peer information structures is when agents have no information about each others' activity levels (Assumption NI). There, the report of one agent cannot react to a change in the activity levels of other agents. Thus $\partial r_j / \partial e_i = 0$, and the strategic effect falls away. As a result, $g'(e_i) = k\theta < t$ in equilibrium, i.e. the agent equalizes marginal benefits from activity levels $g'(e_i)$ to marginal cost, $k\theta$. By concavity of the benefit function $g(e_i)$, this activity level $e^{k\theta}$ is higher than that under perfect information e^t. In fact, it coincides with the activity level under the RAM ($p_i = k$).

Proposition 4 (Activity under CAM - no information) *Given that agents have no information about each other's activity level (Assumption NI), competitive auditing leads to the same per-agent activity level as random auditing, $e^{k\theta}$, implicitly defined by:*

$$g'(e^{k\theta}) = k\theta < t.$$

Activity levels are higher than the socially optimal activity level.

Propositions 2, 3, and 4 constitute the core deliverables of an analytical framework that aims to integrate previous theoretical contributions into a setting that can bridge into the laboratory. Propositions 2 and 4, in particular, are isomorph to the main theoretical findings by Gilpatric et al. (2011) and Cason et al. (2016), respectively. Proposition 3 is isomorph to the main finding by Oestreich (2017). Jointly, they provide a body of results that can be subjected to experimental testing.

Collusion. One remaining concern in the CAM centers on the possibility of collusion among regulated agents. Collusion can occur in equilibrium in repeat interactions and with sufficiently patient agents. When colluding, agents subject to a CAM reduce their reported activity level to zero. This equalizes the audit probability across agents. As a result, all agents rationally adjust their activity level to the RAM level, namely $e = e^{\alpha\theta}$. Both the primary objective of reducing activity levels and secondary objective of truthful self-reporting would be undermined. Since agents would be better off in this case compared to the Nash equilibrium under the CAM, the threat of collusion to the CAM is real and a possible outcome in an experimental implementation.

While regulators need to take the threat of collusion seriously, its impact should not be overstated. Even optimal collusive behavior by regulated firms would lead to outcomes that are not worse from the regulator's point of view than the RAM. So while collusion undermines the leverage gained from competitive auditing, activity levels are not higher and self-reporting not lower than with random auditing.

3 Hypotheses, Experimental Design, and Procedures

3.1 Testable Hypotheses

A natural structure for a test of the relative performance of CAM and RAM in the presence (Assumption PI) or absence (Assumption NI) of peer information is a 2x2 design. In this design, the experimenter manipulates, along one dimension, the audit mechanism (RAM vs.

CAM) and along the other, the information agents have about each other's level of activity (NI vs. PI). Reorganizing propositions 1 to 4 to fit a 2x2 design leads us to three main hypotheses derived from the theoretical framework and testable in its experimental implementation.

The first hypothesis considers the relative performance of CAM and RAM by combining the insights of propositions 1, 2, and 4 about activity and reporting in the absence of peer information.

Hypothesis 1. In a setting in which participants have *no information* about each other's activity level (NI), participants making choices under the competitive audit mechanism (CAM) will exhibit, on average, a) the same level of activity and b) a higher level of reported activity than participants making choices under the random audit mechanism (RAM).

The second hypothesis considers the relative performance of CAM and RAM by combining the insights of propositions 1, 2, and 3 about activity and reporting in the presence of peer information.

Hypothesis 2. In a setting in which participants have *perfect information* about each other's activity level (PI), participants making choices under the competitive audit mechanism (CAM) will exhibit, on average, a) a lower level of activity and b) a higher level of reported activity than participants making choices under the random audit mechanism (RAM).

On activity levels, Hypothesis 2 relies on the fact that the strategic interaction among agents under the CAM creates an additional marginal cost from producing an extra unit of activity that is only present if (i) agents can observe each others' activity levels and (ii) if the audit mechanism creates a reporting competition among the agents. The experimental CAM therefore needs to be "calibrated" to induce the optimal activity level this way (see section 3.2). On reporting, both Hypothesis 1 and Hypothesis 2 rely on the direct and indirect components of reporting one more unit of activity on the expected fine. The indirect effect, in particular, is present whether or not the other agents can observe each others' activity levels.

The properly calibrated CAM should, in theory, not just support better regulatory performance of the CAM compared to the RAM, as captured in Hypotheses 1 and 2. It should also implement social efficient activity levels. This leads to the final hypothesis about experimental outcomes.

Hypothesis 3. In a setting in which participants have *perfect information* about each others' activity level (PI), average activity levels under the competitive audit mechanism (CAM) do not differ from the socially efficient level of activity.

3.2 Experimental Design

In the experimental implementation, participants play seven rounds of a game that mimics the theoretical framework of section 2 under a unique audit mechanism (CAM or RAM) and a unique information structure (PI or NI).[6] In each round, participants are matched in groups of three and make two decisions individually: first, their activity level and second, their self-reported activity level. Each unit reported incurs a fee t. Each unit of activity not reported costs θ if an audit reveals that the participant under-reported. Audits take place at the end of each round. Per round, exactly one participant in each group is audited according to its treatment condition (RAM or CAM). To ensure independence between each round, participants are re-matched every round. Due to a perfect stranger matching procedure, a participant never encounters the same group member more than once.

The unfolding of a particular round is displayed in Figure 2: Each round is composed of 4 stages: a production stage, an information stage, a report stage and an audit stage.

[6]Screenshots of the experimental program and an English translation of the instructions are provided in Appendix E.

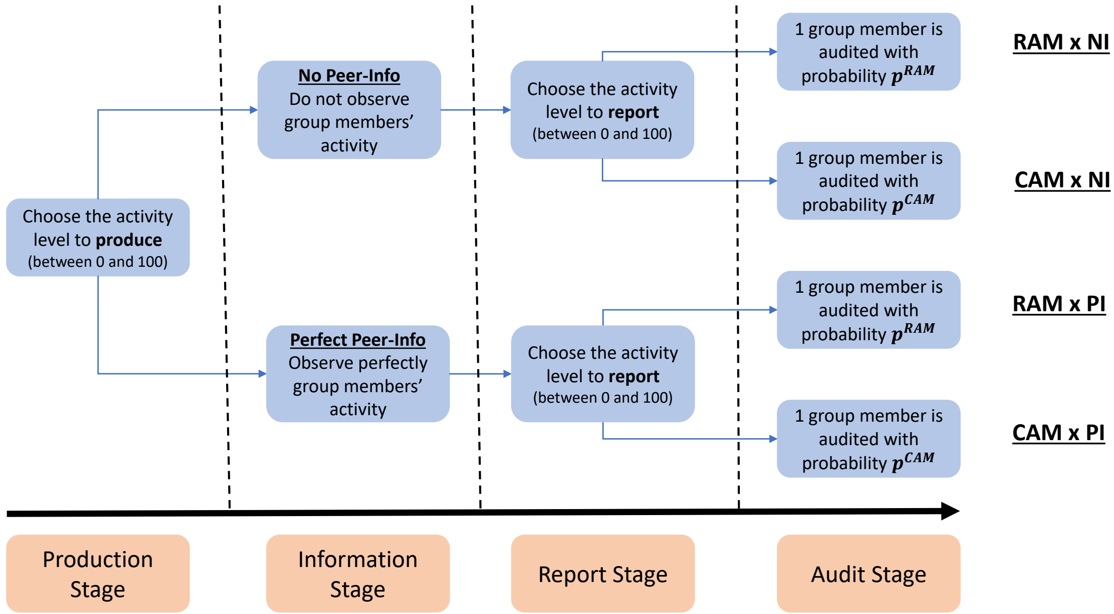


Figure 2: The four stages of an experimental round.

In the production stage, participants choose the level of activity to produce on a slider. In the information stage, participants in the PI treatment are reminded of their chosen activity level and are informed about the chosen activity level of their fellow group members. In the NI treatment, participants do not receive any information about their fellow group members and are only reminded of their own activity level. In the report stage, participants choose the activity level they wish to report on a slider. In the audit stage, at the end of each round, one participant per group is selected for an audit according to the assigned audit probabilities by the audit mechanism. In the RAM treatment, each participant has a fixed probability of 1/3 of being audited regardless of their decisions or the decisions of their fellow group members. In the CAM treatment, the audit probability for a participant depends on her report relative to the reports of her fellow group members. The *higher* a participant's report relative to the reports of her fellow group members, the *lower* her probability of being audited. Conversely, the *lower* a participant's report relative to the reports of her fellow group members, the *higher* her probability of being audited. The exact audit probabilities are calculated according to the CAM algorithm presented in equation (3). Participants can see their final audit probability and conditional payoffs. In addition, information about the actual and reported activity level of every group member are displayed on the screen. This information is the same in all treatments.[7] By pressing a button, participants see whether they have been audited and their earnings for this round. Before moving to the next round, participants are asked to record the information provided on the screen on their personal record sheet.

Post-experimental Questionnaire. At the end of the session, participants are asked to report their age, gender and whether French is their native language. In addition, we elicit risk attitudes following Dohmen et al. (2005) by asking participants to indicate how willing they are to take risks in general on a scale from 0 (not willing at all) to 10 (extremely willing).

Parametrization. Table 1 presents the functional forms and parameters chosen for each variable in the experiment. With these parameter values, a symmetric equilibrium in pure strategies exists under the CAM.[8]

[7]This feature of the design ensures that treatment differences are not driven by differences in learning.

[8]We show in Appendix B that there is no profitable deviation for either agent from the symmetric equilibrium and that the reporting equilibrium is positive. Specifically, we show the evolution of profits if one of the players deviates from the socially optimal activity level e^t and the resulting changes to the reporting equilibrium. We show that there is a global profit maximum at $e_i = e^t$ for $i = 1, 2, 3$.

Table 1: Parameters of the experiment

Notation/ Functional form	Definition	Parameters
N	Number of participants per group	3
K	Number of audits per round	1
p^{RAM}	Random audit probability	0.33
e	Activity level	[0, 100]
r	Reported activity level	[0, 100]
$g'(e) = 10 - 0.1e$	Marginal benefit from activity	
t	Fee on reported activity level	2.5
θ	Penalty on under-reported activity	3
e^0	Unregulated activity level	100
e^t	Optimal activity level	75

3.3 Implementation

The experimental design, hypotheses and procedure were pre-registered on the AEA RCT Registry.[9]

Participants. A total of 131 participants completed the experiment. Participants were recruited via Hroot (Bock et al., 2014) from a large pool of students, mainly from local engineering, business, and medical schools, who had previously registered to be potential participants in economics experiments at GATE-lab (Ecully, France). Overall, 57% of the participants were female and the average age was 23 years (SD = 3.99).

Procedure. The experiment was programmed using oTree (Chen et al., 2016) and con-

[9]RCT ID: AEARCTR-0004996

ducted online in a highly controlled environment that mimics the conditions of the laboratory ("quasi-lab experiment").[10] The experiment was carried out over a series of eight sessions varying between 15 and 21 participants during fall 2020. Digital copies of the instructions were provided to the participants, which were read aloud by the experimenter. To facilitate learning, participants were asked to answer questions about two hypothetical scenarios related to the experiment.[11] In order to help participants to make informed decisions, we provided them with a profit calculator (Healy, 2006; Requate and Waichman, 2011; Cason and Gangadharan, 2013) that allowed them to simulate how their choices would affect their expected payoffs, both in the production and the report stage.[12] The experiment took an average of 1.75 hours. We pre-registered a sample size of 30 independent observations per treatment. With that sample size, the minimum detectable effect size with statistical power at the recommended .80 level is Cohen's d=0.74 for mean comparisons between treatments (Cohen, 2013).

Payment. Participants were paid the sum of their earnings for four randomly selected rounds in addition to a €2 show-up fee and an additional €3 for completing the experiment. The average payoff was €24.71 (SD = 7.02). Participant earnings were denominated in ECU (experimental currency), which was exchanged for euros at the end of the session.[13] At the end of the session, participants were sent a link to retrieve their payment electronically via a third-party platform.

[10] See Appendix C for more details about our online setting.

[11] The comprehension questionnaire is available in Appendix E.2.

[12] See Appendix E.3 for more details about the profit calculator.

[13] In order to avoid large variations in payoffs between treatments, we use an exchange rate of 20 ECU equals €1 for the CAM treatments and 30 ECU equals €1 in the RAM treatments.

4 Results

We first provide descriptive statistics about our participant pool and give an overview of the raw data. This is followed by the core part with tests of the main hypotheses about treatment differences in both actual and reported activity levels and a comparison between observed activity levels and the Nash predictions. We conclude with exploratory analyses that exploit some of the experimental evidence to inform future research.

4.1 Descriptive Statistics

Balance Check. Table 2 summarizes participants demographics by treatments. Using one-way ANOVAs, we find no significant difference between treatments in the percentage of female (F(3,130)=0.43, p=0.733), age (F(3,130)=1.67, p=0.177) and the percentage of participants who indicated French as their native language (F(3,130)=0.29, p=0.831).

Table 2: Summary of participants demographics, by treatments.

Treatments	Cluster	% female	mean age	% French
RAM-NI	33	57.58%	23.73	90.91%
		(3.259)	(0.318)	(1.896)
RAM-PI	33	54.55%	23.88	84.85%
		(3.283)	(0.303)	(2.364)
CAM-NI	30	65.63%	23.5	84.38%
		(3.181)	(0.253)	(2.431)
CAM-PI	35	52.78%	22	88.89%
		(3.151)	(0.119)	(1.983)

Note: Table 2 displays the number of participants, the percentage of female, the mean age, and the percentage of French native speaker, by treatments. Standard errors in parentheses.

Data Overview. Figure 3 provides an overview of our experimental data. Starting with participants' activity levels, the upper panel shows their distribution aggregated across rounds at the participants level for both RAM (in red) and CAM (in blue), with the NI treatment at the top and the PI treatment below. The lower panel displays, on a different scale, the same information for participants' self-reported activity levels.

Inspecting the upper panel ('Activity'), most observed activity levels lie in the interval of 70 to 100, with means between 83 and 91 and therefore above the social optimum at 75. In the absence of peer information (NI), the CAM (blue) and the RAM (red) produce visually similar distributions of activity levels (top). In the presence of peer information (PI), by contrast, the distributions visually diverge (bottom): Activity levels under the CAM shift perceptibly to the left of those under the RAM. This is a first indication that activity levels are lower under CAM with PI. However, the mean of activity levels under CAM–PI is clearly situated to the right of the optimum at 75.

Turning to the lower panel ('Reported Activity'), observed reports fall into a wide range between 0 and 90, but show clear patterns across treatments. Starting with the RAM treatment in both information conditions PI and NI, first recall the Nash equilibrium reporting prediction of zero (Proposition 1). While some participants indeed choose zero or near zero reporting levels, most of the observations are well away from zero and therefore more truthful than predicted. A glance at the distributions also shows that the reporting levels under the CAM treatments sit discernably to the right of those under the RAM treatments, irrespective of information structure (PI and NI). This divergence of distributions is suggestive of higher reporting levels under the CAM regardless of the information structure. Taken together, the experimental data exhibit considerable heterogeneity in participants' activity and reporting choices, but also reveal patterns that are in line with the theoretical results and invite formal testing.

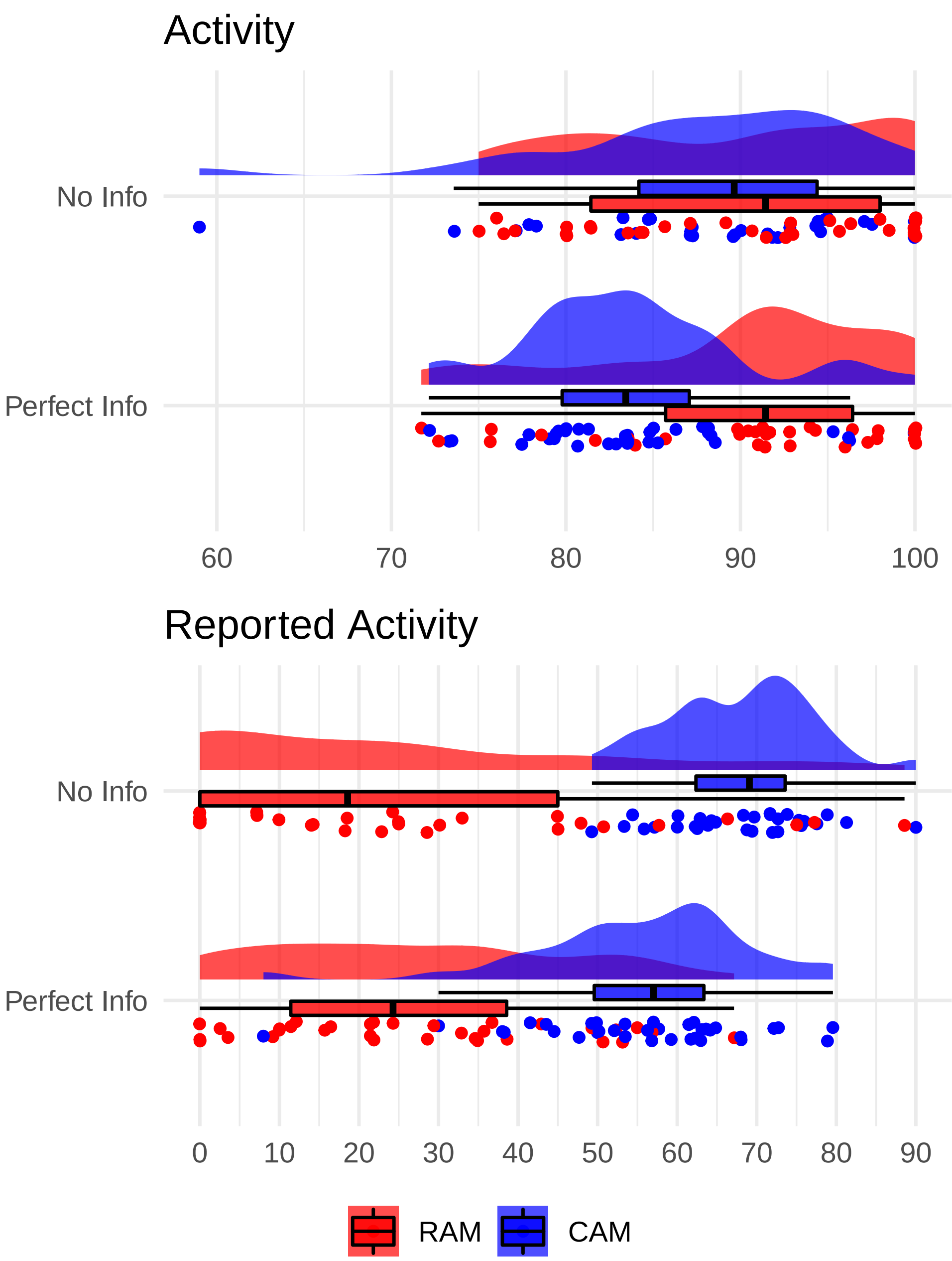


Figure 3: Distribution of activity levels (upper panel) and self-reported activity (lower panel) in the NI (top) and PI (bottom) condition under the RAM (red) and the CAM (blue) aggregated across rounds.

4.2 Tests of Main Hypotheses

Treatment differences. Our main results are displayed in the four panels of Figure 4. They show the evolution of the average levels of actual (top panels) and reported (bottom) activity levels in each round, under both the RAM (in red) and the CAM (in blue) in the NI (left) and the PI condition (right). The theoretical predictions are visible as dotted lines in matching colors.

Consistent with Hypotheses 1 and 2, the top-right panel of Figure 4 shows that the average activity level under the CAM in each round is below the average activity level under the RAM when peer information is present (PI). At the same time, activity levels under the CAM are above those predicted for the treatment. The top-left panel, on the other hand, shows that in the absence of peer-information, activity levels under the CAM and RAM accord with the theoretical predictions and are statistically indistinguishable.

Results on self-reporting are equally consistent with Hypothesese 1 and 2: The bottom panels of Figure 4 show that in every single round, the average reporting levels under the CAM are higher than those under the RAM. Reporting is above theoretical predictions in both information conditions.

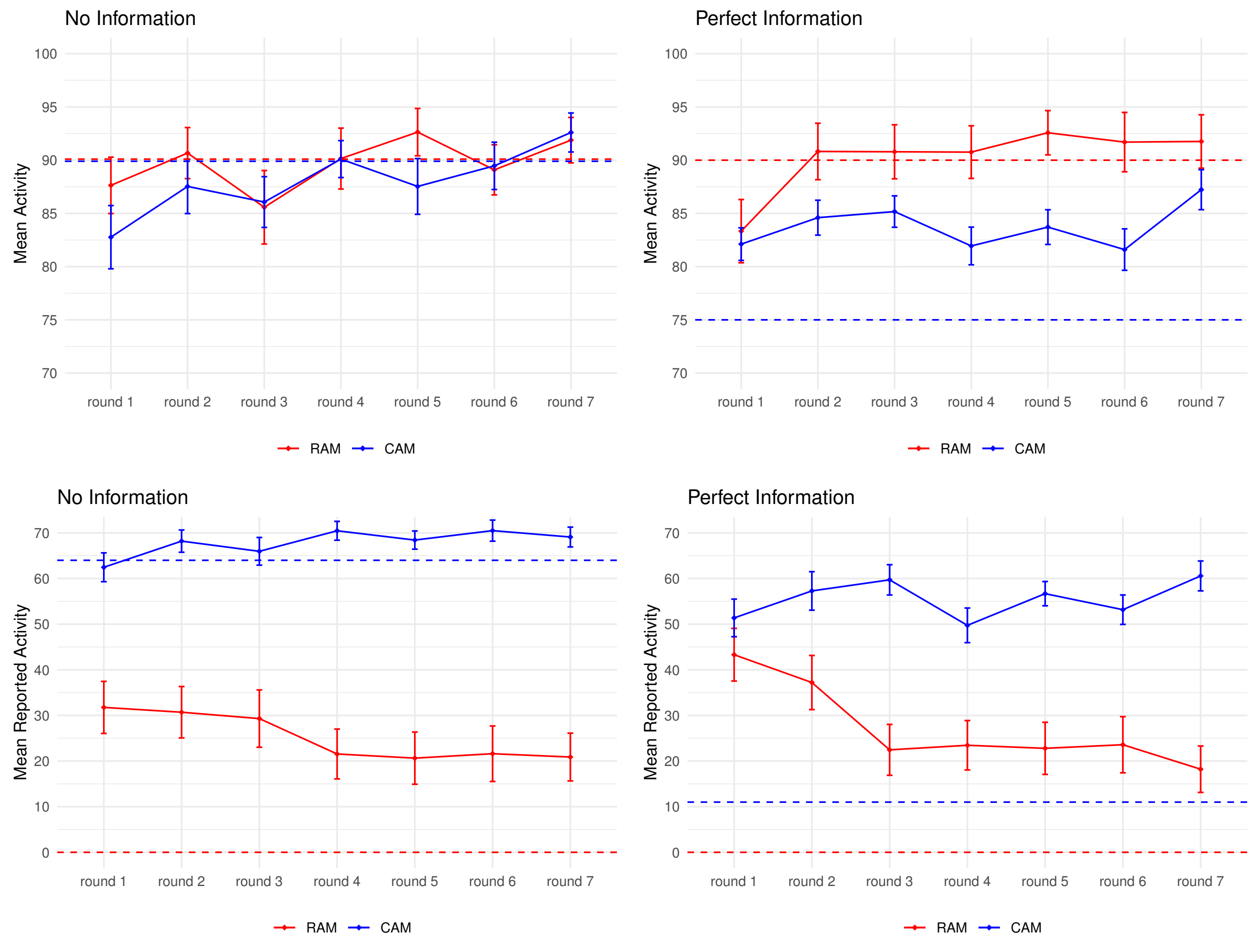


Figure 4: Evolution of the mean activity levels and mean reported activity levels under both the RAM (in red) and the CAM (in blue) across rounds, by information structures. Vertical bars indicate standard errors. Red dotted lines indicate the Nash Equilibrium for the RAM. Blue dotted lines indicate the Nash Equilibrium for the CAM.

To test our main hypotheses, we compare means across the RAM and the CAM treatments under each information condition, PI and NI. The empirical means (EM) of activity and report levels are given, by treatment, in Table 3 together with the Nash predictions derived from the model (NE).

In the absence of peer-information (NI), we find no significant differences in activity levels between the CAM and the RAM (two-sided Mann-Whitney tests:[14] p=0.495). However, we do find significantly higher reported activity levels under the CAM than under the RAM (MW

[14] MW test, hereafter.

test: p<0.001). This supports Hypothesis 1.

Table 3: Empirical means (EM) and Nash equilibrium (NE) of actual and reported levels of activity, by treatments.

Activity level	RAM-NI		RAM-PI		CAM-NI		CAM-PI	
	EM	NE	EM	NE	EM	NE	EM	NE
Actual	89.66	90	90.25	90	88.01	90	83.77***	75
	(1.481)		(1.427)		(1.594)		(1.077)	
Reported	25.20***	0	27.29***	0	67.88*	64	55.51***	11
	(4.520)		(3.291)		(1.676)		(2.376)	
Obs.	33		33		30		35	

Note: Table 3 displays the empirical means (EM) and Nash equilibrium (NE) of both actual and reported activity levels, by treatment. Standard errors in parentheses. Our unit of observation is the average across all rounds of a participant's level of activity. Stars indicates differences between EM and NE using one-sample Wilcoxon sign-ranks test. *p<0.05; **p<0.01; ***p<0.001.

Result 1 *In a setting in which participants have no information about each other's activity level (NI), the CAM leads to a) the same level of activity and b) a higher level of reported activity than the RAM (supports H1).*

In the presence of peer information (PI), we find that the CAM leads to significantly lower levels of activity (MW test: p<0.001) and higher levels of reported activity than the RAM (MW test: p<0.001). This supports Hypothesis 2.

Result 2 *In a setting in which participants have perfect information about each other's activity level, the CAM lead to a) a lower level of activity and b) a higher level of reported activity than the RAM (supports H2).*

To probe the above results further, we perform random-effects GLS regressions clustered at the participant level. The independent variables include a dummy variable equal to 1 if the participant was allocated to the CAM treatment and 0 otherwise plus rounds fixed effects. The GLS coefficients are displayed in columns (1) to (8) of Table 4. We use activity levels as the dependent variable in columns (1) to (4) and self-report levels in columns (5) to (8). Columns (1), (2), (5) and (6) show the results in the NI condition, columns (3), (4), (7) and (8) those for the PI condition. In columns (2), (4), (6) and (8) we additionally control for participants' demographics (gender, age, French as native language) as well as risk attitudes.[15]

Consistent with Results 1 and 2, columns (1) and (2) of Table 4 show no significant differences between the RAM and the CAM in activity levels in the NI condition (p=0.444 and p=0.828, respectively). In contrast, columns (3) and (4) show that the CAM induces lower activity levels than the RAM in the PI condition and the results are significant at the 0.1% level ($p<0.001$ in both specifications). For self-reporting, columns (5) to (8) show that the CAM induces significantly higher levels of reported activity than the RAM, both in the NI ($p<0.001$ in both models) and the PI condition ($p<0.001$ in both specifications).[16]

In summary, we find two key theoretical predictions confirmed by the experimental evidence: When participants have perfect information about each other's activity level, the CAM outperforms the RAM both with respect to the primary objective of aligning activity levels with the regulatory target and with respect to the secondary objective of encouraging more truthful self-reporting. When participants have no information about each others' activity, the CAM performs as well as the RAM on activity levels, but induces more truthful self-reporting.

[15]The missing cluster in columns (2) and (6) is due to one participant leaving the experiment without completing the post-experimental questionnaire.

[16]Appendices D.2 and D.3 show that our main results are robust to various sample restrictions.

Table 4: Effect of the audit mechanisms on actual and reported levels of activity, by information structure.

Dep. var:	Actual level of activity				Reported level of activity			
	NI		PI		NI		PI	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CAM	-1.653	-0.442	-6.479***	-7.864***	42.67***	41.92***	28.22***	25.57***
	(2.175)	(2.037)	(1.787)	(1.676)	(4.823)	(4.932)	(4.058)	(4.212)
Round FE	X	X	X	X	X	X	X	X
Ind. controls		X		X		X		X
Const.	86.10	106.22	86.04	103.39	26.06	23.66	32.93	49.76
	(2.129)	(11.85)	(2.031)	(6.151)	(4.930)	(23.54)	(4.650)	(21.97)
Obs.	441	434	476	476	441	434	476	476
Clusters	63	62	68	68	63	62	68	68

Note: Table 4 displays the GLS coefficients of random effects regressions clustered at the participant level. Standard errors in parentheses. Stars indicates significant differences from the RAM. *p <0.05, **p <0.01, ***p <0.001.

Point predictions. Point predictions constitute the most challenging test for the experiment given the many cognitive and affective factors that could induce a deviation in participants' behavior from the assumptions of the model. For tests of point predictions, we refer back to Table 3 to compare the empirical data against our theoretical predictions.

Comparing the predicted (NE) and observed (EM) reporting behavior in all treatments, we replicate the well-established empirical finding (Mazar et al., 2008; Goldstone and Chin, 1993) that individuals' self-reporting behavior typically deviates from pure payoff maximization (MW tests: RAM-NI: p<0.001; RAM-PI: p<0.001; CAM-NI: p=0.035; CAM-PI: p<0.001). The literature offers both moral and social preferences, such as lying-aversion (Gneezy, 2005), as candidate explanations.

Comparing predicted (NE) and observed (EM) activity levels, we find no difference for the RAM under either of the information structures (MW test: RAM-NI: p=0.851; RAM-

PI: p=0.376) and for the CAM under NI (MW test: p=0.367). In contrast, the CAM induces higher levels of activity than predicted by theory under perfect information (MW test: p<0.001). As in the case of reporting behavior, this points to behavioral mechanisms that the theory does not account for. We summarize as follows.

Result 3 *In a setting in which participants have perfect information about each others' activity level (PI), average activity levels under the CAM are significantly above the socially optimal level of activity (does not support H3).*

The deviation of the experimental data from the point predictions highlights the potential value added of refining the theoretical model. The parsimonious version developed in section 2 performs more than adequately in terms of predicting the direction of the treatment effects. In order to succeed with point predictions, the theory requires a considerably richer model of agent behavior in a complex decision situation. We offer some pointers towards the underlying mechanisms that such a refinement may want to consider in section 4.3.2.

4.3 Exploratory Analyses

Results 1 through 3 technically exhaust the design of the experiment. However, the experiment has produced additional evidence that can help inform future research. We focus on three aspects: Differences in heterogeneity across mechanisms (Gilpatric et al., 2011), underlying mechanisms, and the evidence on collusive behavior.

4.3.1 Heterogeneity

Compared to the RAM, the structure in the CAM is more complex. Higher variance in experiments involving contests and tournaments is a common finding in the literature, adding possible subtlety to comparisons based on empirical means only. One exception are Gilpatric et al. (2011) who find that the CAM leads to less variance in reporting than the RAM.

Visually, the experimental data suggests patterns in line with the findings of Gilpatric et al. (2011): In Figure 3, the box plots of activity and reporting levels exhibit less variance among observations under the CAM compared to the RAM. To explore this issue further, we first investigate treatment differences in the variance of activity choices. To do so, we follow Gilpatric et al. (2011) by estimating random-effects GLS regressions of the squared deviation from the mean activity level, that is $(e_{ij} - \bar{e_j})^2$, where $\bar{e_j}$ is the treatment-specific mean level of activity in round j, on the treatment dummies. We control for rounds fixed effects and standard errors are clustered at the individual level. The GLS coefficients are reported in column (1) of Table 5. In addition, we also investigate treatment differences in the variance of reported activity. To do so, we estimate the same random-effects GLS regression as before, using the squared deviation from the mean reported activity level as the dependent variable, that is $(r_{ij} - \bar{r_j})^2$, where $\bar{r_j}$ is the treatment-specific mean reported level of activity in round j. The GLS coefficients of this estimation are reported in column (2) of Table 5.

Column (1) in Table 5 shows no significant differences in variance between the CAM and the RAM in terms of activity levels in the absence of peer-information (p=0.211). In contrast, column (2) shows that the CAM leads to significantly less heterogeneity in activity levels under perfect peer information (p=0.014). Second, we replicate Gilpatric et al.'s (2011) finding that the CAM leads to significantly less heterogeneity in reported activity levels under both information structures (p <0.001 in both cases), as shown in columns (3) and (4). These results suggest that, in contrast to experimental findings from previous studies investigating competitive incentives, our competitive audit mechanism actually leads to relatively less heterogeneity in individual behavior.

Result 4 *The CAM does not induce more heterogeneity in actual nor reported activity choices.*

Table 5: Effect of information structure and audit mechanism on actual and reported levels of activity.

Dep. var:	$(e_{ij} - \bar{e_j})^2$		$(r_{ij} - \bar{r_j})^2$	
	NI	PI	NI	PI
	(1)	(2)	(3)	(4)
CAM	-56.54	-115.08*	-874.12***	-600.29***
	(45.25)	(46.95)	(156.27)	(142.20)
round FE	X	X	X	X
Const.	266.15	237.36	1098.40	1124.11
	(53.85)	(51.07)	(128.68)	(117.36)
Obs.	441	476	441	476
Clusters	63	68	63	68

Note: Table 5 displays the GLS coefficients of random effects regressions clustered at the participant level. Standard errors in parentheses. Stars indicates significant differences from the baseline (RAM-NI). *p <0.05, **p <0.01, ***p <0.001.

4.3.2 Underlying mechanisms

While the tests of Hypotheses 1 and 2 are successful, they do not interrogate the experimental data on the question whether the mechanisms underlying these results align with the theory used to derive the hypotheses. The theory makes rather specific predictions about patterns in individual-level choice data that a closer look at the evidence can uncover. This closer look comes with a disclaimer: Our experiment was not designed to test competing hypotheses about the mechanisms underlying. The exploratory analysis on the mechanism is therefore purely in the spirit of enhancing the evidence base for future experimentation in this area.

We focus on three predictions of the theory that focus on the reporting behavior: One is that under the RAM, one's own activity level as well as the competitors' activity level does not have a significant influence on reports. The reason is that neither affect the likelihood of being audited, regardless of the information structure (see Proposition 1). The second is that under the CAM with no information, participants change their reports solely according to their own activity level and not according to their peers' activity levels. The reason is trivial: The participant cannot observe others' activity levels. The third prediction is that under the CAM with perfect information, both one's own activity level and the competitors' activity level influence self-reporting. The reason are the direct and indirect components of the marginal benefit of reporting (equation 4) that link reports in a group. If the predictions are valid, their mechanisms should give rise to recognizable patterns in the individual-level choice data.

To detect the predicted patterns, we regress participant i's reported activity level in round t on i's actual activity level in round t and the average activity level of i's fellow group members, including rounds fixed effects. The GLS coefficients are displayed in Table 6 for RAM-NI (columns (1) and (2)), RAM-PI (columns (3) and (4)), CAM-NI (columns (5) and (6) and CAM-PI (columns (7) and (8)). In models (2), (4), (6) and (8) we control for participant i's gender, age, risk attitudes and whether French is i's native language.

Table 6 shows that under the RAM, reports are independent from the activity levels, in the NI and the PI condition ($p > 0.230$ in columns (1) to (4)). Under the CAM, on the other hand, participants condition their reports on their own activity levels in both conditions ($p < 0.001$ in columns (5) to (8)). In the PI condition, there is also tentative evidence that participants condition their reports on the average activity level observed in their group ($p = 0.060$ and $p = 0.048$ in column (7) and (8), respectively). In terms of effect directions, these findings constitute patterns that would be consistent with the theoretical mechanisms that underpin Hypotheses 1 and 2.

Table 6: Effect of activity levels on reported activity levels.

Dep. var: Activity reported	RAM-NI		RAM-PI		CAM-NI		CAM-PI	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Activity level	0.103	0.132	-0.049	-0.042	0.539***	0.519***	0.831***	0.820***
	(0.106)	(0.110)	(0.142)	(0.166)	(0.070)	(0.085)	(0.203)	(0.184)
GA activity level	0.065	0.091	-0.314	-0.319	-0.012	-0.011	0.376	0.395*
	(0.132)	(0.135)	(0.168)	(0.166)	(0.553)	(0.060)	(0.200)	(0.199)
Round FE	X	X	X	X	X	X	X	X
Ind. controls		X		X		X		X
Const.	17.00	-10.81	73.50***	88.64*	18.86***	31.94	-47.79	4.532
	(16.30)	(28.90)	(18.73)	(43.77)	(5.326)	(21.39)	(28.16)	(35.79)
Obs.	231	231	231	231	209	202	245	245
Clusters	33	33	33	33	30	29	35	35

Note: Table 6 displays the GLS coefficients of random effects regressions clustered at the participant level of i's reported activity level on i's actual activity levels and the average activity levels of i's fellow group members, including rounds fixed. We control for participant i's gender, age, risk attitudes and whether French is i's native language in columns (2), (4), (6) and (8). Standard errors in parentheses. Stars indicates significant differences from 0. *p<0.05, **p<0.01, ***p<0.001.

While consistent in terms of direction, the effect sizes are smaller than predicted. For instance, in the CAM-PI treatment, one can compare the theoretically predicted effect of an agent increasing her activity by one unit at the equilibrium on her own reporting with the empirically estimated coefficient. Using the calibration in table (1), the predicted effect size is 1.57. In contrast, the empirically estimated effect size in specifications (7) and (8) is 0.831 and 0.820, respectively (chi2 tests H0: $\beta = 1.57$: $p < 0.001$ in both models). Likewise in the CAM-PI treatment, the theoretically predicted effect of both peers increasing their activity level by one unit (*i.e.*, GA activity level increases by 1) on participant's own reporting is 2.01 units. In contrast, the empirically estimated effect in specifications (7) and (8) is only 0.376 and 0.395, respectively (chi2 tests H0: $\beta = 2$: $p < 0.001$ in both models). These observations are in line with Result 3: The CAM-PI treatment produces lower activity levels, but not as low as predicted by Hypothesis 3.

The failure of Hypothesis 3 points to the presence of mechanisms that the theory does not account for. Several mechanisms could be at play: For instance, if participants systematically report above the profit-maximizing level, then this can dilute the competitive effects induced by the CAM. Also, the CAM adopted in this paper is "calibrated" to implement socially optimal activity levels among risk-neutral agents. Because laboratory participants are typically risk-averse (Harrison and Rutström, 2008), deviations that reduce variation in expected profit are to be expected. Finally, several participants reported that their goal was to avoid being audited. Such 'audit aversion' may lead to higher activity levels in stage 1 in order to be able to report more in stage 2, which in turn leads to a lower audit probability in stage 3. One piece of evidence from the experimental data is that risk attitudes, while affecting self-reporting behavior, does not appear to be linked to the choice of activity level (see Table 7 in Appendix).

Ultimately, the analyses remain speculative since we cannot rule out alternative causal mechanisms through experimental design. Further research, however, can build on this evidence to establish the role of other drivers and to determine what features would enable the mechanism to implement full compliance.

4.3.3 Collusion

Recall from subsection 2.3 that the CAM enables the possibility of collusion among regulated agents. By colluding, agents behave "as if" audit probabilities were evenly distributed by reducing their reports to zero and adjusting their activity level to the RAM level. Results from Table 3 show that collusion is not a concern in our data. First, reported activity levels are substantially above zero in both CAM-NI (Wilcoxon sign-rank test: p <0.001) and CAM-PI (Wilcoxon sign-rank test: p <0.001) treatments. Second, while CAM fails to achieve the socially optimal level of activity, activity levels in CAM-PI are still significantly lower than activity levels observed under RAM (MW test: p <0.001). Note that the absence of communication and the fact that participants were rematched at the beginning of each round may have prevented the emergence of such an equilibrium. Hence, we cannot exclude that collusion can be a problem when these two features are relaxed.

5 Conclusion

Recent theoretical and experimental research has highlighted the potential of competitive audit mechanisms for harnessing the benefits of self-reporting for achieving regulatory targets while limiting misreporting. Such mechanisms are therefore of considerable interest to both academics and regulators. Our paper scrutinizes this potential with respect to two dimensions. One is a comparison of competitive audit mechanisms with the more conventional random audit mechanism not just in terms of truthful reporting, but also in terms of aligning activity levels with regulatory targets. The other is a better understanding of the role of peer information in the comparative performance. To make progress, the present paper integrates the key results of the existing literature into a parsimonious theoretical framework and translates this framework into an experimental design in which information structures can be exogenously manipulated in isolation.

The theoretical part of our investigation generates predictions that the competitive audit mechanism will outperform the random audit mechanism in terms of truthful self-reporting irrespective of the presence or absence of peer information, but in terms of activity levels only in its presence. In the latter case, however, socially optimal activity levels are predicted to be implementable. All except one of these predictions turn out to be accurate. The quasi-laboratory experiment that implements the theoretical framework in the form of an experimental design finds that on average, participants self-report more truthfully under the competitive mechanism in both information conditions. Without peer information, their mean activity levels do not differ between the mechanisms. When participants can observe peers' activity levels, the competitive mechanism brings activity levels closer into line with regulatory targets. The only theoretical prediction that fails its test is that the competitive mechanism implements the socially optimal activity levels: Participants still under-comply with the regulatory objectives. This could be the result of several factors, including risk attitudes, lying aversion, or a type of 'audit aversion'. These insights are important whenever competitive auditing is considered as a viable alternative for constrained regulators charged with enforcing policies. Regulators need to understand the circumstances under which competitive auditing leads to advantages over random auditing with regards to their primary objective of aligning activity levels with the regulatory target and with respect to their secondary objective of encouraging more truthful self-reporting.

Two stylized facts that are not accounted for by the theoretical framework emerge from the experimental data. First, while the equilibrium under competitive auditing may be threatened by collusion, the collusive equilibrium of no reporting is not observed in the data. In fact, we replicate the well-established finding that self-reporting is significantly higher than predicted by theory regardless of the information structure. This is consistent with the behavioral literature that people decisions can also be influenced by payoff-irrelevant motivations such as social considerations or lying-aversion. Second, we found that the competitive audit mechanism can induce less heterogeneity in individuals' decisions, both in terms of self-reported activity levels and in terms of actual activity levels. These effects constitute unexpected co-benefits of using a competitive audit mechanism and reaffirm the benefits of test-bedding regulations.

One limitation in our current study is that we could only examine the two limiting cases of peer information. In regulatory reality, most situations will fall somewhere between the two extremes of no or perfect peer information. It will be interesting to explore in future research just how good peer information has to be for competitive audit mechanisms to outperform a random mechanism that recommends itself to regulators through its simplicity.

References

Allen Jr, H. M. and Bunn III, W. B. (2003). Validating self-reported measures of productivity at work: a case for their credibility in a heavy manufacturing setting. *Journal of occupational and environmental medicine*, 45(9):926–940.

Bayer, R. and Cowell, F. (2009). Tax compliance and firms' strategic interdependence. *Journal of Public Economics*, 93(11-12):1131–1143.

Bock, O., Baetge, I., and Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, 71:117–120.

Burns, J., Movsisyan, A., Stratil, J. M., Coenen, M., Emmert-Fees, K. M., Geffert, K., Hoffmann, S., Horstick, O., Laxy, M., Pfadenhauer, L. M., et al. (2020). Travel-related control measures to contain the covid-19 pandemic: a rapid review. *Cochrane Database of Systematic Reviews*, (9).

Cason, T. N., Friesen, L., and Gangadharan, L. (2016). Regulatory performance of audit tournaments and compliance observability. *European Economic Review*, 85:288–306.

Cason, T. N., Friesen, L., and Gangadharan, L. (2020). Complying with environmental regulations: Experimental evidence. Technical report, University of Queensland, School of Economics.

Cason, T. N. and Gangadharan, L. (2013). Empowering neighbors versus imposing regulations: An experimental analysis of pollution reduction schemes. *Journal of Environmental Economics and Management*, 65(3):469–484.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

De Marchi, S. and Hamilton, J. T. (2006). Assessing the accuracy of self-reported data: an evaluation of the toxics release inventory. *Journal of Risk and Uncertainty*, 32(1):57–76.

Dohmen, T. J., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2005). Individual risk attitudes: New evidence from a large, representative, experimentally-validated survey. *IZA discussion paper*.

Earnhart, D. and Friesen, L. (2021). Use of competitive endogenous audit mechanisms by federal and state inspectors within environmental protection agencies. *Journal of Environmental Economics and Management*, in press.

Friesen, L. and Gangadharan, L. (2013). Designing self-reporting regimes to encourage truth telling: An experimental study. *Journal of Economic Behavior & Organization*, 94:90–102.

Gilpatric, S. M., Vossler, C. A., and Liu, L. (2015). Using competition to stimulate regulatory compliance: A tournament-based dynamic targeting mechanism. *Journal of Economic Behavior & Organization*, 119:182–196.

Gilpatric, S. M., Vossler, C. A., and McKee, M. (2011). Regulatory enforcement with competitive endogenous audit mechanisms. *The RAND Journal of Economics*, 42(2):292–312.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1):384–394.

Goldstone, R. L. and Chin, C. (1993). Dishonesty in self-report of copes made: moral relativity and the copy machine. *Basic and Applied Social Psychology*, 14(1):19–32.

Harford, J. D. (1987). Self-reporting of pollution and the firm's behavior under imperfectly enforceable regulations. *Journal of Environmental Economics and Management*, 14(3):293–303.

Harrison, G. W. and Rutström, E. E. (2008). Risk aversion in the laboratory. In *Risk aversion in experiments*. Emerald Group Publishing Limited.

Healy, P. J. (2006). Learning dynamics for mechanism design: An experimental comparison of public goods mechanisms. *Journal of Economic Theory*, 129(1):114–149.

Helland, E. (1998). The enforcement of pollution control laws: Inspections, violations, and self-reporting. *Review of Economics and Statistics*, 80(1):141–153.

Innes, R. (2017). Lie aversion and self-reporting in optimal law enforcement. *Journal of Regulatory Economics*, 52(2):107–131.

Malik, A. S. (1993). Self-reporting and the design of policies for regulating stochastic pollution. *Journal of Environmental Economics and Management*, 24(3):241–257.

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research*, 45(6):633–644.

Oestreich, A. M. (2015). Firms' emissions and self-reporting under competitive audit mechanisms. *Environmental and Resource Economics*, 62:949–978.

Oestreich, A. M. (2017). On optimal audit mechanisms for environmental taxes. *Journal of Environmental Economics and Management*, 84:62–83.

Probst, T. M. and Estrada, A. X. (2010). Accident under-reporting among employees: Testing the moderating influence of psychological safety climate and supervisor enforcement of safety practices. *Accident analysis & prevention*, 42(5):1438–1444.

Requate, T. and Waichman, I. (2011). "a profit table or a profit calculator?" a note on the design of cournot oligopoly experiments. *Experimental Economics*, 14(1):36–46.

Telle, K. (2015). Monitoring and enforcement of environmental regulations: lessons from a natural field experiment in norway. *Journal of Public Economics*, 99:24—-34.

Tirole, J. (1988). *The theory of industrial organization.* MIT press.

Vossler, C. A. and Gilpatric, S. M. (2018). Endogenous audits, uncertainty, and taxpayer assistance services: Theory and experiments. *Journal of Public Economics*, 165:217—-229.

A Proofs of Propositions

A.1 Proof of Proposition 1

At stage 3, since $-tr_i - k\theta(e_i - r_i)$ is strictly decreasing in r_i, agents choose $r_i = 0$. At stage 2, agents adjust their activity level e_i according to the marginal expected fine, i.e.: $g'(e^{k\theta}) = k\theta$. Since $g(.)$ is strictly concave and $k\theta < t$, we have $e^t > e^{k\theta}$.

A.2 Proof of Proposition 2

Assuming profit function (1) is differentiable, the set of interior first-order conditions ($\partial \Pi_i / \partial r_i = 0$) for a unique interior reporting solution are:

$$-t + p_i\theta - \frac{\partial p_i}{\partial r_i}(e_i - r_i)\theta = 0,\ \forall r_i \in [0, e_i],\text{ for } i = 1, ..., n. \tag{9}$$

We note that the partial derivative $\partial p_i / \partial r_i$ of the competitive audit mechanism in (3) is:

$$\frac{\partial p_i}{\partial r_i} = \frac{-\lambda(n-1)}{e^0 - r_i}. \tag{10}$$

Using (10) in (9) we get:

$$p_i\theta + \lambda(n-1)\theta\frac{e_i - r_i}{e^0 - r_i} = t \tag{11}$$

which is equal to equation (4) in the main paper. At symmetry we have that $r_i = r^*$ and $p_i = k$. Thus, we can re-write (11) as:

$$\frac{e_i - r_i}{e^0 - r_i} = \frac{\frac{t}{\theta} - k}{\lambda(n-1)} \tag{12}$$

Using $\lambda = (t/\theta - k)/((n-1)(2-N))$ (12) is:

$$\begin{aligned} \frac{e_i - r_i}{e^0 - r_i} &= (2 - N) \\ r_i &= \frac{e_i - e^0(2-N)}{(N-1)} \end{aligned}$$

which is positive as long as $e_i - e^0(2-N) > 0$ and thus larger than the reporting equilibrium under the RAM $r_i^{RAM} = 0$.

A.3 Proof of Proposition 3

Following from (8) and the according discussion, we need to show that $\partial r_j / \partial e_i = 1$ for $i = 1, ..., n$ and $i \neq j$ in a symmetric pure strategy equilibrium where $e_1 = ... = e_n$ and $r_1 = ... = r_n$. In that case, a necessary condition holds for all agents to choose socially efficient activity levels in equilibrium, namely $g'(e^t) = t$. The proof of existence of this equilibrium is beyond the scope of this paper. We do however show through the example in section xx that the set of equilibrium values is not empty.

First, define $s_i = e_i - r_i$ and $R_i = e_0 - r_i$ to simplify notation. Applying the audit mechanism proposed in xx, the set of interior first-order conditions with respect to reporting ($\partial E\Pi_i/\partial r_i = 0$) can be written as:

$$-\frac{t}{\theta} + p_i(\mathbf{r}) + \frac{\lambda(n-1)}{R_i}s_i = 0, \text{ for } i = 1, ..., n. \tag{13}$$

Total differentiation of system (13) yields the following matrix system:

$$\begin{bmatrix} -2\frac{\lambda(n-1)}{R_1} + \frac{\lambda(n-1)}{(R_1)^2}s_1 & \frac{\lambda}{R_2} & . & \frac{\lambda}{R_n} \\ \frac{\lambda}{R_1} & -2\frac{\lambda(n-1)}{R_2} + \frac{\lambda(n-1)}{(R_2)^2}s_2 & . & \frac{\lambda}{R_n} \\ . & . & . & . \\ \frac{\lambda}{R_1} & \frac{\lambda}{R_2} & . & -2\frac{\lambda(n-1)}{R_n} + \frac{\lambda(n-1)}{(R_n)^2}s_n \end{bmatrix} \begin{pmatrix} dr_1 \\ dr_2 \\ . \\ dr_n \end{pmatrix}$$
$$= \begin{bmatrix} -\frac{\lambda(n-1)}{R_1} & 0 & . & 0 \\ 0 & -\frac{\lambda(n-1)}{R_2} & . & 0 \\ . & . & . & . \\ 0 & 0 & . & -\frac{\lambda(n-1)}{R_n} \end{bmatrix} \begin{pmatrix} de_1 \\ de_2 \\ . \\ de_n \end{pmatrix} \tag{14}$$

At symmetry, where $e_1 = ... = e_n$ and $r_1 = ... = r_n$ matrix system (14) simplifies to:

$$\begin{bmatrix} -2\frac{\lambda(n-1)}{R_i} + \frac{\lambda(n-1)}{(R_i)^2}s_i & \frac{\lambda}{R_i} & ... & \frac{\lambda}{R_i} \\ \frac{\lambda}{R_i} & -2\frac{\lambda(n-1)}{R_i} + \frac{\lambda(n-1)}{(R_i)^2}s_i & ... & \frac{\lambda}{R_i} \\ ... & ... & ... & ... \\ \frac{\lambda}{R_i} & \frac{\lambda}{R_i} & ... & -2\frac{\lambda(n-1)}{R_i} + \frac{\lambda(n-1)}{(R_i)^2}s_i \end{bmatrix} \begin{pmatrix} dr_1 \\ dr_2 \\ ... \\ dr_n \end{pmatrix}$$
$$= \begin{bmatrix} -\frac{\lambda(n-1)}{R_i} & 0 & 0 & 0 \\ 0 & -\frac{\lambda(n-1)}{R_i} & 0 & 0 \\ 0 & 0 & ... & 0 \\ 0 & 0 & 0 & -\frac{\lambda(n-1)}{R_i} \end{bmatrix} \begin{pmatrix} de_1 \\ de_2 \\ ... \\ de_n \end{pmatrix} \tag{15}$$

Using the implicit function theorem in (15) allows to obtain the partial derivatives of reporting

of any agent j with respect to a change in activity of agent i at symmetry where $e_1 = ... = e_n$ and $r_1 = ... = r_n$:

$$\frac{\partial r_j}{\partial e_i} = \frac{-\frac{\lambda}{R_i}\frac{-\lambda(n-1)}{R_i}}{(-2\frac{\lambda(n-1)}{R_i} + \frac{\lambda(n-1)}{(R_i)^2}s_i)^2 + (n-2)(-2\frac{\lambda(n-1)}{R_i} + \frac{\lambda(n-1)}{(R_i)^2}s_i)\frac{\lambda}{R_i} - (n-1)(\frac{\lambda}{R_i})^2} \text{ at } e_1 = ... = e_n. \tag{16}$$

We neglect the qualifier "at $e_1 = ... = e_n$" in the following to simplify notation. Using (13) we can derive that $s_i = (k - \frac{t}{\theta})/\frac{-\lambda(n-1)}{R_i}$ at symmetry and thus:

$$\frac{\partial r_j}{\partial e_i} = \frac{\frac{\lambda}{R_i}\frac{\lambda(n-1)}{R_i}}{(-2\frac{\lambda(n-1)}{R_i} + \frac{\lambda(n-1)}{(R_i)^2}\frac{(k-\frac{t}{\theta})}{-\frac{\lambda(n-1)}{R_i}})^2 + (n-2)(-2\frac{\lambda(n-1)}{R_i} + \frac{\lambda(n-1)}{(R_i)^2}\frac{(k-\frac{t}{\theta})}{-\frac{\lambda(n-1)}{R_i}})\frac{\lambda}{R_i} - (n-1)(\frac{\lambda}{R_i})^2}. \tag{17}$$

We can simplify (17) also by using $\lambda = (t/\theta - k)/((n-1)(2-N))$:

$$\frac{\partial r_j}{\partial e_i} = \frac{1}{(n-1)N^2 - (n-2)(N)) - 1}. \tag{18}$$

From (18) we can now see that $r_j/\partial e_i = 1$ at $e_1 = ... = e_n$ and $r_1 = ... = r_n$ iff $(n-1)N^2 - (n-2)(N)) = 2$. This is indeed the case as shown in Figure (5) below. This concludes the Proof.

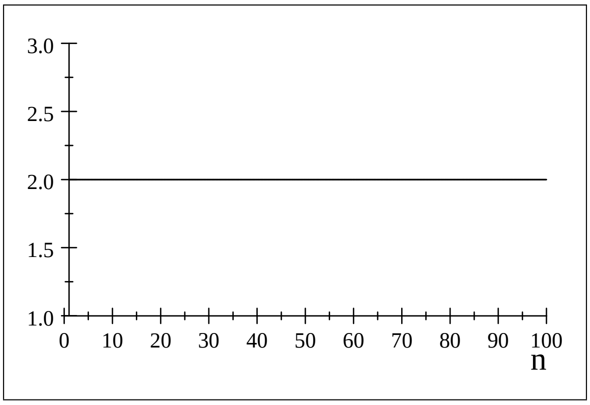


Figure 5: Plot of $(n-1)N^2 - (n-2)(N))$,
where $N = \left(n - 2 + \sqrt{n^2 + 4n - 4}\right)/(2\,(n-1))$.

B Simulations

Perfect Information. In the case of perfect information between regulated agents and CAM, there is no incentive to unilaterally deviate from the subgame-perfect Nash Equilibrium (SPNE) neither in terms of activity nor in terms of reporting applying the parameters presented in Table 1. Particularly, Panel (a) and Panel (c) of Figure 6 illustrate that the equilibrium candidate (identified by the first-order necessary condition) $e_1 = e_2 = e_3 = e^t = 75$ and $r_1 = r_2 = r_3 = r^* = 10.96$ lead to a global profit maximum for agent 1 ensuring that those choices are indeed the symmetric SPNE.

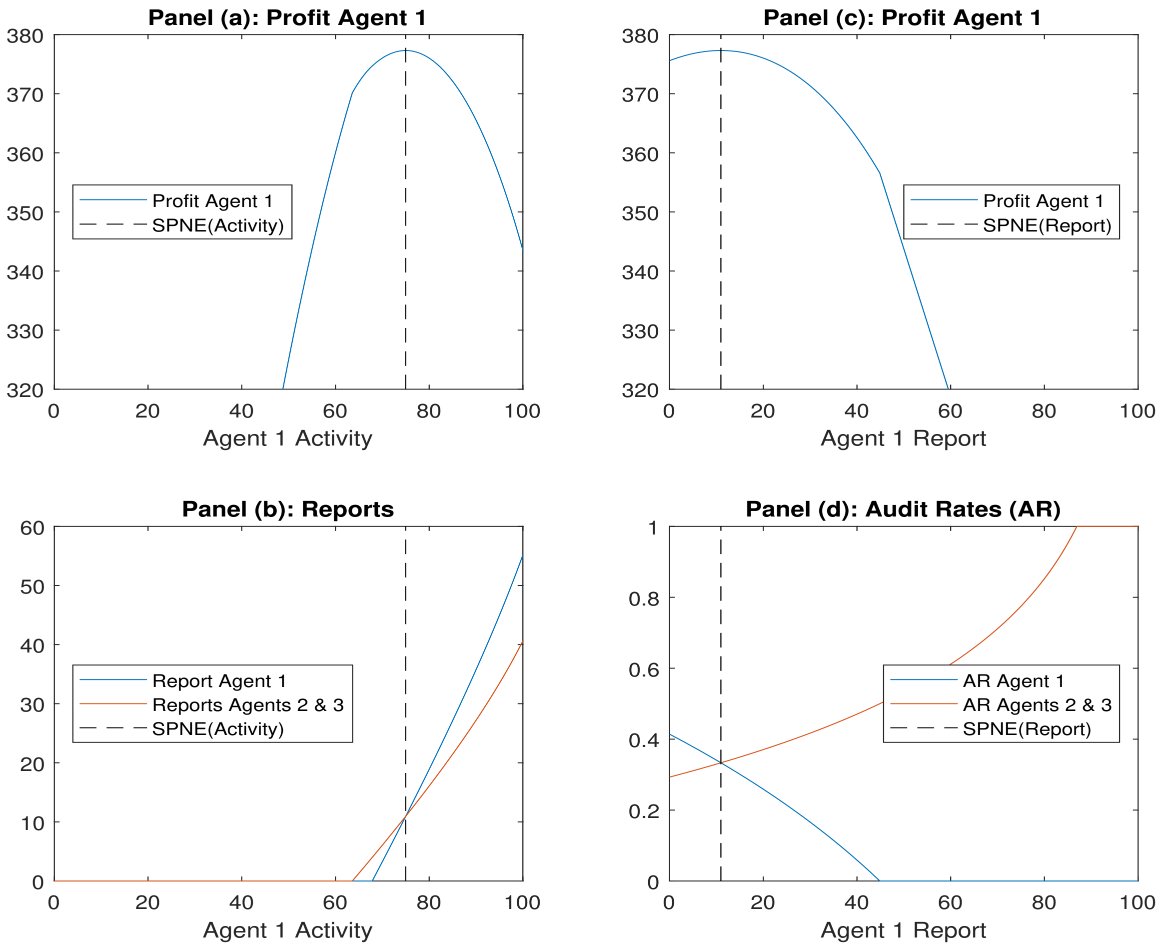


Figure 6: Impacts on profit when agent 1 unilaterally deviates from the SPNE. Panel (a) shows the impacts on profit and Panel (b) shows the impacts on reporting, when agent 1 varies the activity level e_1 from 0 to e^0 holding the activity levels of the other agents fixed at the optimal level $e_2 = e_3 = e^t$. Panel (c) shows the impacts on profit and Panel (d) shows the impacts on audit rates, when agent 1 varies the report r_1 from 0 to e^0 holding the reporting levels of the other agents fixed at the reporting equilibrium $r_2 = r_3 = r^*$ and also fixing all activity levels.

No Information. In the case of no information between agents and CAM, there is no incentive to unilaterally deviate from the Nash Equilibrium (NE) neither in terms of activity nor in terms of reporting applying the parameters presented in Table 1. Particularly, Figure 7 illustrates that the equilibrium candidate (identified by the first-order necessary condition) $e_1 = e_2 = e_3 = e^{CAM} = 90$ and $r_1 = r_2 = r_3 = r^{CAM} = 64$ lead to a global profit maximum for agent 1 ensuring that those choices are indeed the symmetric NE.

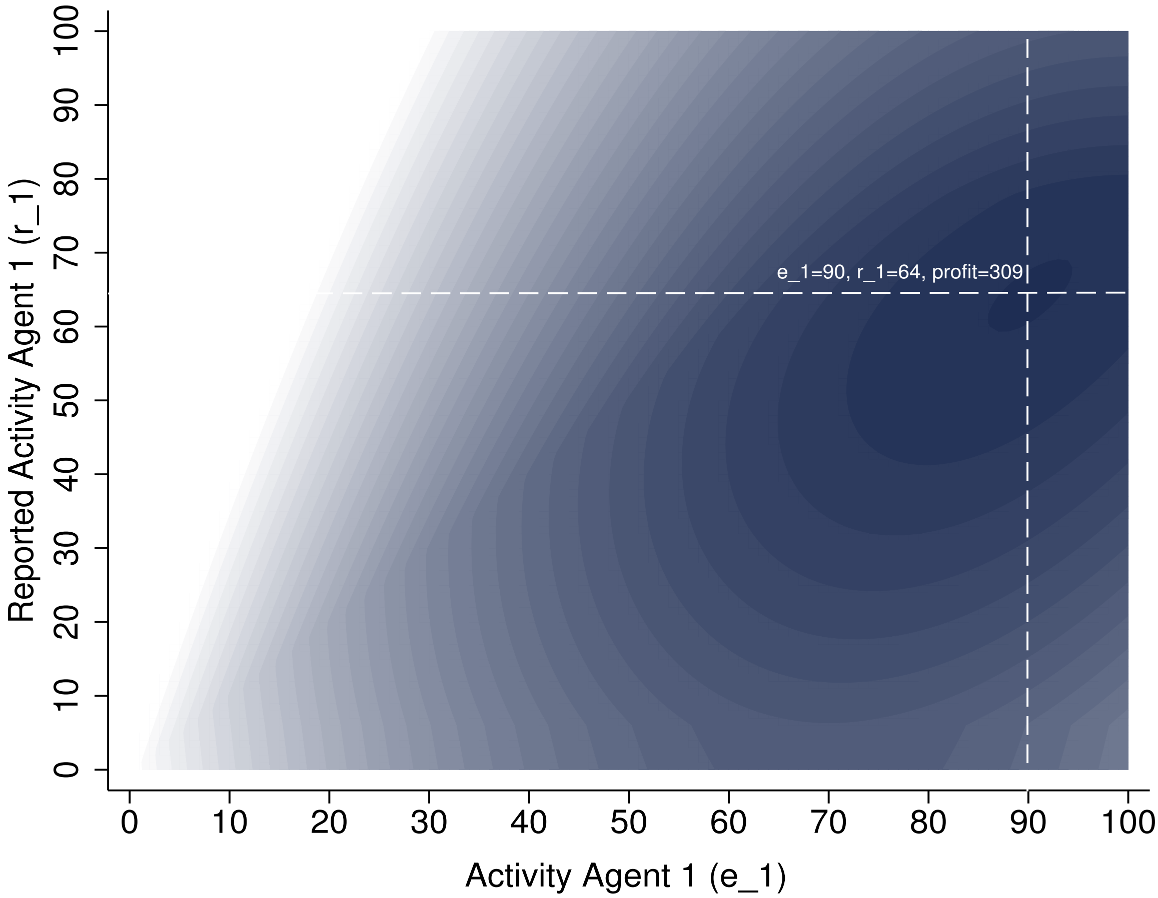


Figure 7: The figure illustrates the profit of agent 1, if agent 1 unilaterally deviates from the NE in terms of activity (e_1) or in terms of reporting (r_1). Darker blue areas show higher profits, lighter blue and white areas show lower and negative profits respectively. The activity levels of the other two agents are fixed at the equilibrium level $e_2 = e_3 = e^{CAM} = 90$ and also the reported activity levels of the other two agents are fixed at the equilibrium $r_2 = r_3 = r^{CAM} = 64$. The equilibrium choices are illustrated through the two white dashed lines.

C Detailed experimental procedure

Unfolding of a session. After enrolling for a session, participants were provided with a unique participant ID and the link to join the virtual lab.[17] 15 minutes prior to the beginning of the session, participants were asked to login to the virtual lab waiting room using their unique participant ID. An experimenter would check participants ID before letting them in the virtual lab. At the start of the session, participants were asked to turn on their webcam[18] while the experimenter explains the unfolding of the session. Participants were invited to send questions via private messages to the experimenter at any time during the experiment.[19] Participants were paid via Lyf Pay, a French peer-to-peer payment platform. After completing the experiment, participants were asked to provide a phone number or email address where they wished to receive the Lyf Pay link to retrieve their payment. Participants were invited to confirm the reception of the link before leaving the virtual lab. Participants could then choose to receive their earnings on their Lyf Pay account if they have one or directly on their bank account by providing their bank details on the Lyf Pay interface otherwise. Participants were aware of the payment procedure before signing-up for the experiment.

Attrition management. Due to the online nature of our experiment, we pre-registered the following attrition management procedure to prevent drop-outs from disrupting the session. If a participant fails to make a decision in the activity stage before the end of the allocated time, the missing value would be replaced by the equilibrium value. The time allocated to each page was set to the mean + 2 standard deviations of the time spent by participants in a pilot session to allow plenty of time for participants to make their decisions.

[17]The virtual lab was hosted on heiCONF, a web conferencing system for audio and video conferences based on BigBlueButton. The service is hosted and operated by Heidelberg University Computing Centre, which means that all data is stored in the university's server rooms.

[18]Participants were informed before signing-up that they would need to be able to take part in the experiment.

[19]Note that participants were prevented to send private messages to each other and to see each other's video feed. While the experimenter could see the video feed of every participant, participants could only see the video feed of the experimenter.

D Additional Analyses

D.1 Main Results (Demographics)

Table 7: Effect of the audit mechanisms on actual and reported levels of activity, by information structure.

Dep. var:	Actual level of activity		Reported level of activity	
	NI	PI	NI	PI
	(1)	(2)	(3)	(4)
CAM	-0.442	-7.864***	41.92***	25.57***
	(2.037)	(1.676)	(4.932)	(4.212)
Gender	-4.591	2.625	1.540	-4.598
	(2.755)	(1.605)	(6.277)	(4.044)
Age	-0.693*	-0.744***	1.004	-0.381
	(0.281)	(0.183)	(0.590)	(0.796)
Native French	1.089	-0.361	-16.32	4.454
	(4.576)	(2.229)	(9.669)	(6.044)
Risk Pref.	-0.250	-0.112	-1.113	-2.245*
	(0.513)	(0.412)	(1.122)	(0.895)
Round FE	X	X	X	X
Const.	106.22	103.39	23.66	49.76
	(11.85)	(6.151)	(23.54)	(21.97)
Obs.	434	476	434	476
Clusters	62	68	62	68

Note: Table 7 displays the GLS coefficients of random effects regressions clustered at the participant level. Standard errors in parentheses. Stars indicates significant differences from the RAM. *p<0.05, **p<0.01, ***p<0.001.

D.2 Robustness Checks: Comprehension questionnaire

We replicate our main results in the main text by restricting the sample to participants who did not failed the comprehension questionnaire. One participant failed the comprehension questionnaire in the NI x RAM treatment, 1 participant in the PI x RAM treatment and 2 participants in the CAM x NI. No participant failed the comprehension questionnaire in the CAM x PI treatment. The GLS coefficients are displayed in Table 8. In columns (2), (4), (6) and (8) we control for participants demographics (gender, age and whether French is their native language) as well as risk attitudes. Table 8 show that our main results are not driven by a lack of comprehension of the instructions as the GLS coefficients remains quantitatively the same.

Table 8: Effect of the audit mechanisms on actual and reported levels of activity, by information structure.

Dep. var:	Actual level of activity				Reported level of activity			
	NI		PI		NI		PI	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CAM	1.159	0.309	-6.492***	-7.931***	43.61***	43.06***	28.05***	25.56***
	(2.926)	(1.955)	(2.006)	(1.685)	(4.944)	(4.823)	(4.139)	(4.224)
Round FE	X	X	X	X	X	X	X	X
Ind. controls		X		X		X		X
Const.	86.19	97.45***	86.14	103.73***	25.00	8.406	33.14	49.81*
	(2.149)	(10.31)	(2.076)	(6.142)	(4.942)	(24.57)	(4.753)	(22.12)
Obs.	420	420	469	469	420	420	469	469
Clusters	60	60	67	67	60	60	67	67

Note: Table 8 displays the GLS coefficients of random effects regressions clustered at the participant level. Standard errors in parentheses. Stars indicates significant differences from the RAM. *p<0.05, **p<0.01, ***p<0.001.

D.3 Robustness Checks: Attrition Procedure

As explained in Appendix C, we implemented a back up strategy in case a participant drops-out from a session. A total of 4 participants dropped-out from a session before the beginning of the first round (3 participants in the CAM x NI treatment and 1 participant in the CAM x PI treatment). No participant dropped-out from the experiment after the beginning of the first round. As pre-registered, these participants actual and reported activity levels were replaced by the equilibrium values in all rounds.

In this section, we investigate whether our attrition procedure had a significant effect on participants decisions after encountering an automated play. To do so, we created two dummy variables. The variable "met AP $t-1$" equals 1 if the participant were matched with an automated player in the previous round, and 0 otherwise. The variable "met AP before t" equals 1 if the participant encountered automated play in any round prior to the current round, and 0 otherwise. We then replicate the models in Table 4 in the main text after imposing restrictions on the sample. Results are displayed in Table 9. In columns (1), (3), (5) and (7), we exclude observations in round t for participants who encountered automated play in round $t-1$. In columns (2), (4), (6) and (8), we exclude observations for all subsequent rounds for participants who encountered automated play in a particular round. Table 9 shows that our main results are not sensitive to our attrition procedure as the GLS coefficients remain quantitatively the same.

Table 9: Effect of the audit mechanisms on actual and reported levels of activity, by information structure.

Dep. var:	Actual level of activity				Reported level of activity			
	NI		PI		NI		PI	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CAM	-1.133	-2.352	-6.633***	-6.605***	42.65***	39.75***	27.91***	26.51***
	(2.135)	(2.431)	(1.788)	(1.846)	(4.827)	(5.196)	(4.072)	(4.089)
Round FE	X	X	X	X	X	X	X	X
Const.	85.86	86.44	86.12	86.11	26.07	27.45	33.09	33.81
	(2.141)	(2.245)	(2.028)	(2.056)	(4.927)	(5.094)	(4.645)	(4.699)
Obs.	408	339	464	434	408	339	464	434
Clusters	63	63	68	68	63	63	68	68

Note: Table 9 displays the GLS coefficients of random effects regressions clustered at the participant level. Standard errors in parentheses. Stars indicates significant differences from the RAM. *p<0.05, **p<0.01, ***p<0.001.

In addition, we conducted GLS regressions with random effects and clustered at the participant level of the effect of these two dummy variables on participants' activity level (Table 10) and reported activity levels (Table 11) in both treatments in which drop-outs occurred. The results displayed in Tables 10 and 11 are consistent with the above statement that our main results are not driven by the attrition procedure as none of the two attrition dummies has a significant effect on participants' activity levels. Attrition dummy "met AP before $t-1$" takes on value one if the participant encountered automated play in the previous round and zero otherwise. Attrition dummy "met AP before t" takes on value one if the participant ever encountered automated play in any of the previous rounds and zero otherwise.

Table 10: Effect of automated plays on activity.

Dep. var: Activity	CAM x NI		CAM x PI	
met AP $t-1$	-3.883	–	3.278	–
	(2.413)		(2.671)	
met AP before t	–	-1.765	–	1.318
		(2.601)		(2.033)
round	0.886	1.486*	0.12	0.206
	(0.503)	(0.657)	(0.348)	(0.298)
Constant	85.61***	82.92***	83.32***	82.72***
	(2.681)	(2.650)	(1.929)	(1.528)
Obs.	180	210	210	245
Clusters	30	30	35	35
R-squared	0.02	0.03	0.01	0.02

Note: Table 10 displays the GLS coefficients of random effects regressions clustered at the participant level. Standard errors in parentheses. Stars indicates significant differences from 0. *p<0.05, **p<0.01, ***p<0.001.

Table 11: Effect of automated plays on reported activity.

Dep. var: Reported activity	CAM x NI		CAM x PI	
met AP $t-1$	-0.886		3.550	
	(1.691)		(5.167)	
met AP before t		-0.910		3.347
		(2.602)		(4.180)
round	0.446	1.088	0.107	0.392
	(0.577)	(0.657)	(0.670)	(0.600)
Constant	66.93***	63.97***	55.51***	53.36***
	(3.193)	(3.068)	(3.833)	(3.215)
Obs.	180	210	210	245
Clusters	30	30	35	35
R-squared	0.01	0.02	0.01	0.02

Note: Table 11 displays the GLS coefficients of random effects regressions clustered at the participant level. Standard errors in parentheses. Stars indicates significant differences from 0. *p<0.05, **p<0.01, ***p<0.001.

E Instructions and Screenshots

All original and translated instructions are available on OSF.[20]

E.1 Instructions CAM Perfect Info (English translation)

Thank you for taking part in this experiment on decision making.

Please ensure that your mobile phone remains switched off. It is important for the integrity of the experiment that you refrain from communicating with other participants and make your decisions in a timely manner.

All your decisions are strictly anonymous.

Your payment for this experiment is composed of 3 parts. First, you will receive 2 euros for joining the session on time; Second, you will receive 3 euros for completing the experiment; Third, you will receive a variable payment of up to 20 euros which depends on your decisions, the decisions of other participants in this session and an element of chance. At the end of the experiment, the experimenter will send you a Lyf Pay link that you can use to retrieve your payment.

This experiment is divided into 7 paid rounds. You will be paid based on your decisions in 4 randomly chosen rounds among these 7 rounds. Each decision you make is therefore important because it has a chance of determining the amount of money you earn.
In this experiment, you will be in groups. Your group will always be composed of yourself and 2 other participants in this session. Groups are randomly formed at the beginning of each round, but you will never be matched with the same participant more than once.

You will make decisions privately, that is, without consulting other group members.

A timer will be displayed on each page. If you fail to validate your decision before the end of the allocated time, the program will record the last values that were on the screen when the time expired and you will be automatically forwarded to the next page. The timer can differ from pages to pages and from round to round.
If you have a question as we read through the instructions or any time during the experiment, you can send a private message to the experimenter via the videoconference platform (left click on the experimenter's name 'send a private message').

Your earnings in the experiment are denominated in ECU (Experimental Currency), which will be exchanged at a rate of (CAM) 20 ECU = 1 euro at the end of the experiment.

Overview

[20]Link: `https://osf.io/92qwt/?view_only=4920328a2ec74a04934be305fb2142c7`

In many contexts, economic agents are asked to self-report their activities. This experiment captures the two distinct decision stages involved in this context: in each round, you make two decisions. One decision is the level of activity to produce. The second decision is the level of activity to report. Both decisions affect your earnings for particular round.

Changing your activity level can influence your earnings. On the one hand, you pay a fee on each unit of activity that you report. On the other hand, you pay a penalty on each unit of activity that you did not report if an audit reveals that you under-reported your activity level. The greater the under-reporting, the higher the penalty.

Audits take place at the end of each round. Per round, exactly one participant in your group will be audited. An audit has no consequence for a participant that did not under-report their chosen activity level. On the other hand, a participant whom the audit reveals to have under-reported will pay a penalty. The likelihood that you will be audited is not fixed. The more activity you report compared to the others in the group, the less likely it is that you will be audited.

Your earnings in one period depend on your decisions –and the decisions of others – in that particular period only.

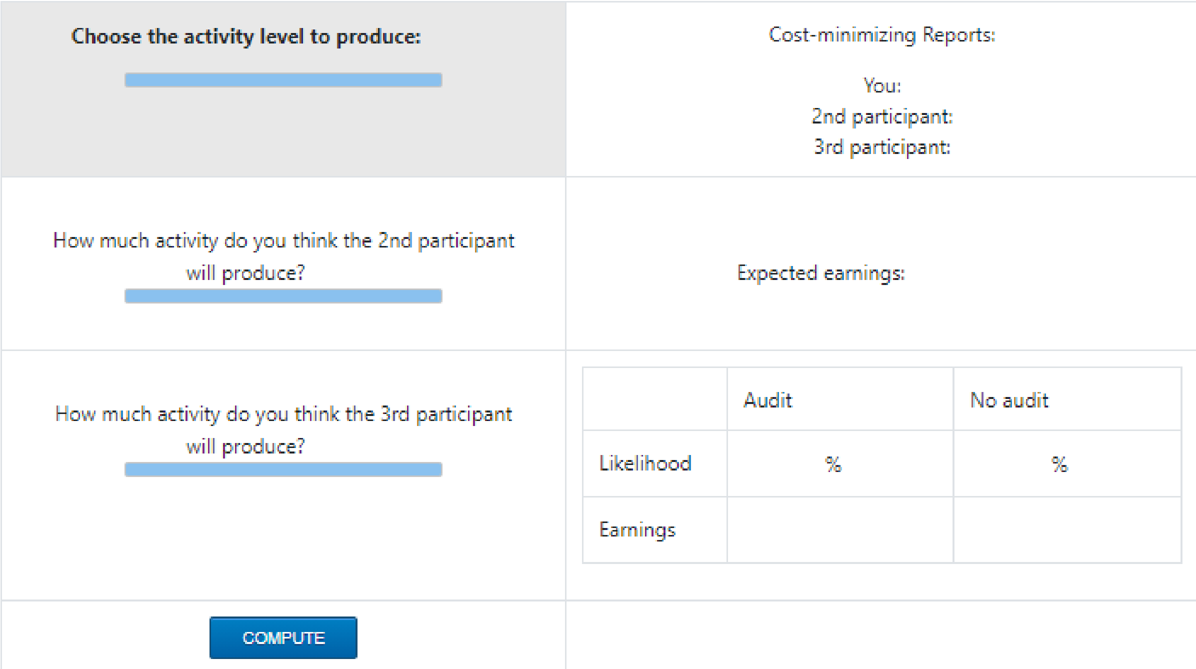
Each round is composed of 4 stages: a Production stage, an Information stage, a Report stage and an Audit stage.

Your two decisions take place in the Production stage and the Report stage. The details of each stage is described below.

Production stage
Your task (and those of others in the group) in this stage is to choose the level of activity to produce. Each unit of activity generates a private revenue for you. This revenue is determined by the following formula:
Revenue = 10 x activity level – 1/20 x (activity level) ^{2} - 200
This formula is the same for everyone in the group. You can see an example of the Production stage below.

Decision
You must indicate your choice of activity level between 0 and 100 on the slider in the top-left corner of the screen.
You must also indicate your beliefs about the activity level of the two other participants in your group. Note that the sliders do not have default values. You must click on the slider to reveal the handle.

Information
In order to help you make your decision, you will receive information about the consequences of your choices.



After you indicate your choice of activity and your beliefs about the activity level of the two other participants in your group, use the 'COMPUTE' button to compute your cost-minimizing report, your likelihood of being audited and your payoffs for this round, given your production, your beliefs about the activity level of the two other participants in your group and their own cost-minimizing reports.

The cost-minimizing report for each participant in the group is displayed at the top-right corner of your screen. Your cost-minimizing report is the report that leads to the highest expected payoff given your chosen activity level, your guess about the activity level of the two other participants in your group and their own cost-minimizing report.

Your cost-minimizing report only maximizes your expected payoff conditional on the other participants in your group also reporting their own cost-minimizing reports and is, therefore, suggestive.

The likelihood that you will be audited in this round, your payoff if you are audited and your payoff if you are not audited will be displayed in the table at the bottom-right corner of the screen.

Finally, the computer program will compute your expected earnings which summarizes the information displayed in the table. Expected earnings are the sum of your earnings under both scenarios multiplied by their respective likelihoods and is computed as follows:

Expected earnings = Likelihood of audit x earnings if audit occurs + likelihood of no-audit x earnings if no audit occurs

To help you keep track of your expected payoff, the symbol "↑" will be displayed on the screen if your current choices lead to a higher expected payoff than the previous ones. In contrast, the symbol "↓" will be displayed on the screen if your current choices lead to a lower expected payoff than the previous ones. The symbol "=" will be displayed on the screen if your current choices lead to the same expected payoff than the previous ones.

You can change your activity level and your beliefs about the other participants in your group as many times as you want. However, you will need to press the 'COMPUTE' button each time to update the information provided on the right-hand side of the screen.

When you are satisfied with your choice, press the 'NEXT' button to validate your decision and move on to the next stage.

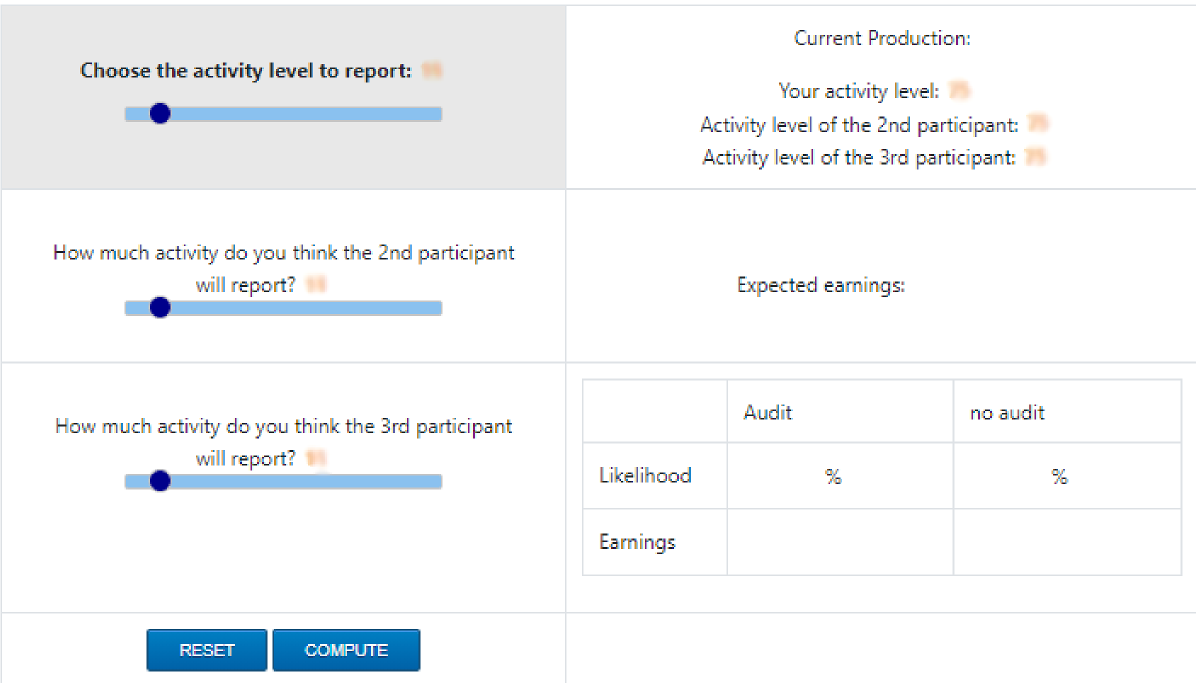
Information stage

After every participant in your group has made a decision in the Production stage, your own activity level in this round will be displayed on the screen.

You will also receive information on the activity level of your fellow group members in this round. The chosen level of activity of each of the two other participants in your group will be displayed on your screen.

Report stage

Your task (and those of others in the group) in this stage is to choose how much activity level to report, using a screen similar to the one below.



You pay a fee of 2.5 ECU for each unit of activity level reported. This fee is deducted from your earnings.

Decision

You must indicate how many units of activity level to report between 0 and 100 on the slider in the top-left corner of the screen.

In addition, you must also indicate your beliefs about the report of the two other participants in your group.

By default, the cost-minimizing reports are displayed on the sliders.

Information

As in the Production Stage, pressing the 'COMPUTE' button will reveal the likelihood that you will be audited, your payoff if you are audited, your payoff if you are not audited, as well as your expected payoff for this round.

You can change your report and your guesses about the reports of the two other participants in your group as many times as you want. However, you will need to press the 'COMPUTE' button each time to update the information provided on the right-hand side of the screen. You can reset the sliders to their default values at any time by pressing the 'RESET' button.

At the top-right corner of the screen, you can see the level of activity that you chose in the previous stage, as well as the chosen activity level of the two other participants in your group. When you are satisfied with your choice, press the 'NEXT' button to validate your decision and move on to the next stage.

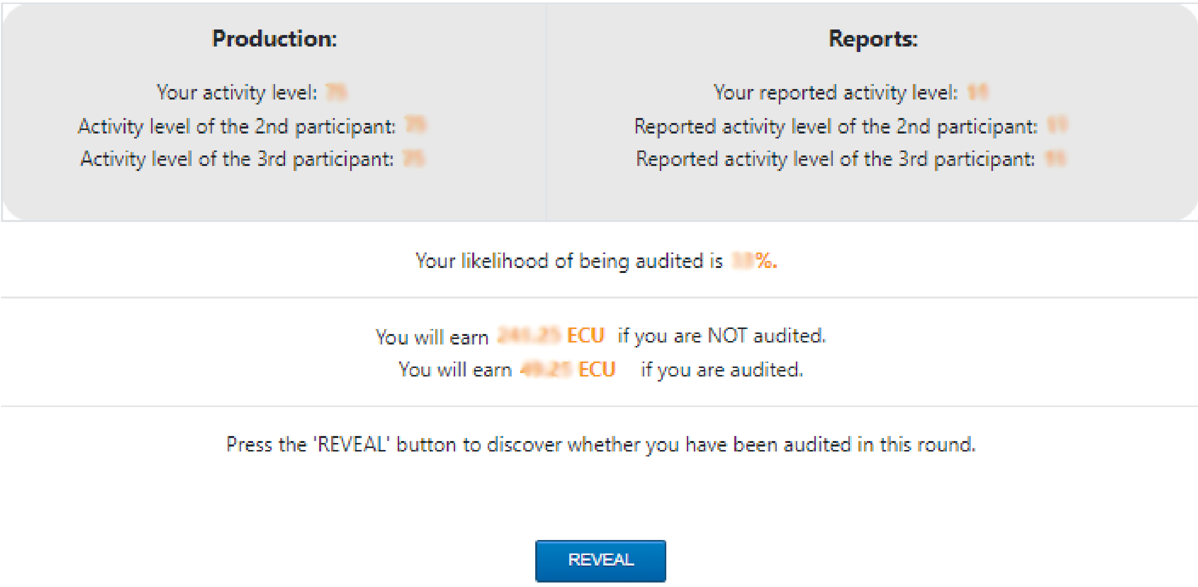
Audit stage

After all the members of your group have submitted their reports, exactly one group member will be audited.

The picture below shows an example of the audit screen. At the top of the screen, there is a summary of the activity level and reports chosen by you and the two other participants in your group. In the middle of the screen, you can see your likelihood of being audited in this round, as well as your payoff if you are audited and your payoff if you are not audited in this round.

Pressing the "REVEAL" button reveals whether you have been audited and your earnings for this round.

To help you keep track of your decisions in each round, you can note down the information presented on this screen on your Personal Record Sheet. A message will be displayed on your screen to remind you to do so.



Determining who is audited

The computer program will select one participant from your group to audit based on your report and the reports of each of the two other participants in your group.

The more activity level you report compared to your fellow group members, the less likely you are to be audited. Conversely, the less activity level you report compared to your fellow group members, the more likely you are to be audited.

The specific formula used by the computer is displayed below.

Likelihood of audit = min max [1/3 + 0.35*(ln (1-your report / 1 – the report of the second participant in your group) + ln (1 – your report / 1 – the report of the third participant in your group))]

What happens if you are audited

An audit reveals your true chosen level of activity.

If you did NOT under-report your activity level (i.e. you reported at least as much as your chosen level of activity), then the audit has no consequence for you.

If you under-reported your activity level (i.e. you reported less than your chosen level of activity), then you pay a penalty of 3 ECU per unit of under-reported activity level.

Your Earnings

As described above, each period your overall earnings for the current round depend upon your activity level, your reported activity level, and whether you are audited or not in this round.

Thus, after you have submitted your report, three things can happen:

1. You are not audited.
2. You are audited and you did not under-report.
3. You are audited and you under-reported.

We summarize below how your earnings will be calculated under each scenario.

Your earnings if you are not audited:
Revenue: 10 x activity level – 1/20 x (activity level)^2 - 200
- Fee: 2.5 x reported activity level
- Penalty: 0
= Period Earnings

Your earnings if you are audited and you did not under-report:
Revenue: 10 x activity level – 1/20 x (activity level)^2 - 200
- Fee: 2.5 x reported activity level
- Penalty: 0
= Period Earnings

Your earnings if you are audited and you under-reported:
Revenue: 10 x activity level – 1/20 x (activity level)^2 - 200
- Fee: 2.5 x reported activity level
- Penalty: 3 (activity level - reported activity level)
= Period Earnings

To help you familiarize yourself with the interface, you will answer a comprehension questionnaire about these instructions before the beginning of the experiment. Please take the time to read these instructions again.

E.2 Comprehension Questionnaire

To help participants to familiarize themselves with the interface, they were asked to complete a comprehension questionnaire before entering the first round of the experiment. The comprehension questionnaire was composed of two different scenarios that mimics the Production stage that participants will face in the experiment. We asked participants the 3 following questions for each scenarios: ”What is the activity level you have to report to minimize your cost?”, ”What is your likelihood of being audited?” and ”What are your expected earnings?”. Both scenarios are the same across information structure.

The scenarios for the Perfect Information treatments were:
Scenario 1: *Let's suppose that you choose to produce an activity level of 80. Now let's assume that the 2nd participant in your group will report an activity level of 40 and the 3rd participant in your group will also report an activity level of 40.*
Scenario 2: *Now, let's suppose that you decide to increase your activity level from 80 to 85 and your beliefs about the other participants in your group remain unchanged.*

The scenarios for the No Information treatments were:
Scenario 1: *Let's suppose that you choose to produce an activity level of 80. Now let's assume that the 2nd participant in your group will choose an activity level of 80 and the 3rd participant in your group will also choose an activity level of 80.*
Scenario 2: *Now, let's suppose that you decide to increase your activity level from 80 to 85 and your beliefs about the other participants in your group remain unchanged.*

Below we present a screenshot of the comprehension questionnaire for the CAM Perfect Info treatment.

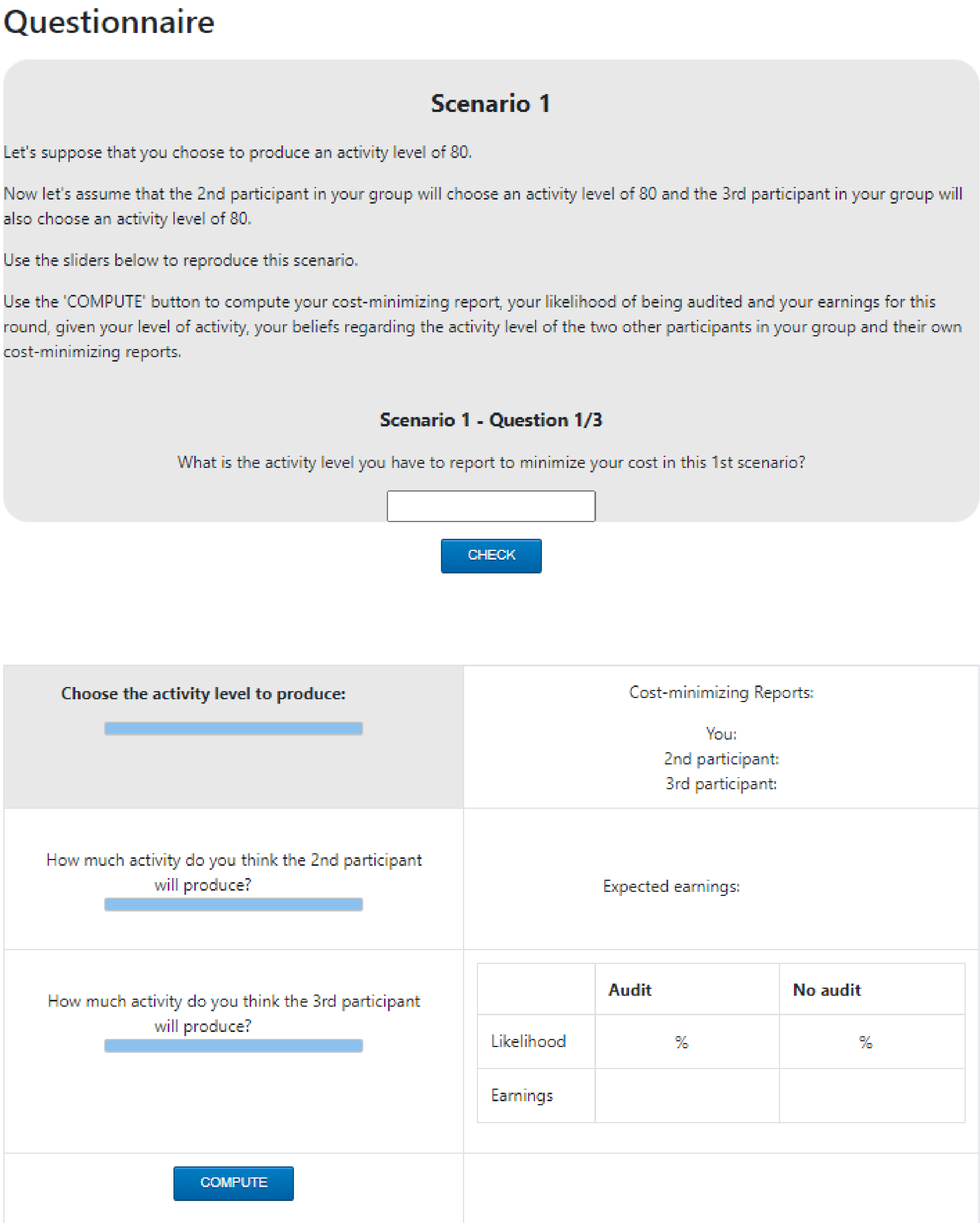


Figure 8: Comprehension Questionnaire (CAM Perfect Information treatment)

E.3 Profit calculator

In order to help participants to make informed decisions, we provided them with a profit calculator that allowed them to simulate how their choices would affect their expected payoffs, both in the production and the report stage.

In each stage, participants were required to input at least one “estimate” regarding the decisions made by other participants. In the production stage, participants were required to input their estimate regarding the activity levels made by other participants in the *Perfect Information* (PI) treatment and the reported activity levels made by other participants in the *No Information* (NI) treatment. The profit calculator would then inform participants about their own cost-minimizing report, their probability of being audited, their probability of not being audited, their payoff in each case, as well as their expected payoffs for the current round. In the PI treatment, participants were additionally informed about the cost-minimizing reports of their fellow group members.

In the report stage, participants participants were required to input their estimate regarding the activity levels made by other participants both in the PI treatment and the NI treatment. In the former, the slider default was set to the cost-minimizing reports of each of the two remaining group members. The profit calculator would then inform participants about their probability of being audited, their probability of not being audited, their payoff in each case, as well as their expected payoffs for the current round. Participants could submit multiple scenarios, allowing them to assess how incremental changes in activity and reported activity levels affect the aforementioned variables before logging their decision.