Discussion Paper Series – CRC TR 224

Discussion Paper No. 114
Project B 04

Costless Information and Costly Verification: A Case for Transparency

Deniz Kattwinkel*
Jan Knoepfle**

July 2019

*University of Bonn, denizkattwinkel@gmail.com
**University of Bonn, J.K@uni-bonn.de

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Costless Information and Costly Verification: A Case for Transparency

Deniz Kattwinkel and Jan Knoepfle*

31st July 2019

Abstract

A principal has to take a binary decision. She relies on information privately held by a completely biased agent. The principal cannot incentivize with transfers but can learn the agent's information at a cost. Additionally, the principal privately observes a signal correlated with the agent's type. Transparent mechanisms are optimal: unlike in standard results with correlation, the principal's payoff is the same as if her signal was public. They take a simple cut-off form: favorable signals ensure the agent's preferred action. Signals below this cut-off lead to the nonpreferred action unless the agent appeals. An appeal always triggers type verification. **Keywords:** Mechanism Design without Transfers, Costly Verification, Robust Mechanism Design, Transparency **JEL Codes:** D61, D82, K40

---

*Kattwinkel: University of Bonn, denizkattwinkel@gmail.com. Knoepfle: University of Bonn, J.K@uni-bonn.de. We thank Marina Halac, Stephan Lauermann, Benny Moldovanu, and Sven Rady for continual guidance and support. We likewise thank Ian Ball, Andreas Kleiner, Daniel Krähmer, Martin Pollrich, Andre Speit, Dezsö Szalay, Tymon Tatur, and participants at the 11th BiGSEM Workshop on Economic Theory (Bielefeld), the BGSE Micro Workshop (Bonn), the Jerusalem Summer School in Economic Theory 2017, the Yale Theory Breakfast, the Stony Brook Conference on Game Theory 2018, and the EEA-ESEM 2018 (Cologne) for insightful discussions. Both authors thank the Yale Economics Department for their hospitality during the respective research visits, when parts of this paper were written. This work was supported by the German Research Foundation (DFG) through CRC TR 224 (Project B04).

A principal has to take a binary decision for which she relies on an agent's private information. The agent prefers one of the two actions independent of his information. Prior to the decision, the principal privately observes a signal about the agent's information. She cannot incentivize the agent through monetary transfers but has the opportunity to reveal his information at a cost.

Examples for this setting include: a human resource department decides whether to hire a candidate, a judge decides whether to acquit or convict a defendant, or a competition authority decided whether to grant or deny a company permission to merge with or acquire another firm.

While one party —the agent— has a clear preference toward one action (the candidate wants to be hired, the defendant wants to be acquitted, and the company wants to merge), the preferences of the other party —the principal— depend on information that is privately held by the agent. Here, one may think of the candidate's ability, the defendant's guilt, or the company's competitive position in the market.

Often, monetary transfers to elicit the agent's private information are not feasible (for practical or moral reasons[1]), but the principal can learn the information at a cost, for example, by conducting an assessment center, a trial, or a market analysis. However, verification is costly, so the principal has an incentive to economize on it.

Typically, costly information acquisition is not the only way to learn the agent's private information. The potential employer receives references or recommendation letters from previous supervisors, the judge can inspect the outcome of pretrial investigations, and the competition authority has sector-specific knowledge derived from its supervisory function. That is, the principal privately observes factors that are correlated with the agent's type.

This paper investigates how the principal can use this private, costless information in conjunction with the costly verification to maximize her expected payoff from the decision if monetary transfers are not feasible.

[1]The assumption is that payments cannot depend on the agent's report. Even though a public sector job entails payments, if the payment is fixed, it cannot be used to incentivize truthful reports of the candidate's ability.

Preview. We show that the optimal Bayesian incentive compatible (BIC) mechanism takes a simple cut-off structure: if the principal observes a signal that makes her sufficiently certain that the agent's preferred action is also better for her, she takes it, independent of the type report. If the signal falls below the cut-off, she takes the nonpreferred action by default but gives the agent the possibility to appeal. An appeal is always verified and induces the agent-preferred action whenever his type exceeds a threshold. This appeal threshold is set such that the agents who appeal are only those whose type makes it worthwhile for the principal to implement the agent-preferred action and pay the verification cost.

This mechanism is ex-post incentive compatible (EPIC). It would also be incentive compatible if the principal's signal was known to the agent. This result implies that the principal does not benefit from the privacy of her signal. It advocates for transparent procedures.

We extend our result to settings where the information that the principal privately observes has a positive direct effect on her utility from the agent's preferred choice. The structure of the optimal mechanism remains, and, again, transparency comes without loss for the principal. If, in contrast, the direct effect is negative, the principal benefits from hiding his information. The equivalence between EPIC and BIC optimal mechanisms breaks down; we show that the simple EPIC mechanism is no longer optimal in the larger class of BIC mechanisms.

Literature. In settings where monetary transfers are feasible, the principal can design a lottery rewarding the agent for guessing the value of her privately observed signal correctly. Different agent types hold different beliefs over the signal distribution and, therefore, reveal their type by guessing the signal they deem most likely. If the agent's liability is not limited, the principal can increase reward and loss in the lottery to such an extent that the incentives to win the lottery exceed any incentives regarding the allocation decision. In doing so, she can learn the agent's type at arbitrarily small costs. Mechanisms with monetary transfers and correlated information have been discussed by

Crémer and McLean (1988), Riordan and Sappington (1988), Johnson et al. (1990), and McAfee and Reny (1992), who all establish conditions on the information structure that ensure full surplus extraction by the principal. As all surplus can be extracted, revenue maximization leads to ex-post efficient allocations. Neeman (2004) discusses the genericity of the above-mentioned conditions and shows that full surplus extraction is possible only if every preference type is "determined" by his belief over the correlated characteristics. Even though this condition is fulfilled, in our setting with costly verification instead of monetary transfers, full surplus extraction is not feasible and implementing the ex-post efficient allocation is not optimal for the principal.

The full surplus-extracting lotteries require potentially unbounded transfers. For the case of bounded transfers or limited liability, Demougin and Garvie (1991) show that the qualitative results, the application of rewards as bets on the signal, still apply. Different from our setting, the principal gains by maintaining her signal private.

In the absence of monetary transfers, Bhargava et al. (2015) show how positively correlated beliefs among voters allow overcoming the impossibility of nondictatorial voting rules established by Gibbard (1973) and Satterthwaite (1975).

Our result is in line with the findings in other settings where monetary transfers are not feasible but correlated information is absent. The literature (Glazer and Rubinstein, 2004; Ben-Porath, Dekel, and Lipman, 2017; Erlanson and Kleiner, 2017; Hart, Kremer, and Perry, 2017; Halac and Yared, 2019) has found optimal mechanisms to take a simple cut-off structure and to be EPIC in the sense that the agents would also report truthfully if they were informed about the other agents' type realizations before their report.

The possibility for the mechanism designer to verify an agent's private information at a cost was first introduced by Townsend (1979) considering a principal-agent model for debt contracts, which was extended to a two-period model by Gale and Hellwig (1985). These early models of state verification feature both, monetary transfers and verification. Glazer and Rubinstein (2004) introduce a setting where the principal has to take a binary decision depending on the multidimensional private information of the agent. Here, the principal cannot use monetary transfers, but she can learn about one dimension before making her decision.

Our model is most closely related to that of Ben-Porath, Dekel, and Lipman (2014), who model costly verification and consider the case of allocating a good among finitely many agents whose types are independently distributed; see also the discussion section. Erlanson and Kleiner (2017) study a collective decision problem with costly verification and show that the optimal mechanism is EPIC and can be implemented by a simple weighted majority voting rule. Mylovanov and Zapechelnyuk (2017) consider an allocation problem without monetary transfers in which the principal learns the agents' types without cost but only posterior to the allocation decision and has the ability to punish untruthful reports up to a limit. Halac and Yared (2019) consider a delegation problem and specify conditions on the verification cost that ensure optimality of a threshold mechanism with an escape clause.

Erlanson and Kleiner (2017) show further that the equivalence between BIC and EPIC mechanisms holds more generally rather than only for optimal mechanisms. This relates to Gershkov, Goeree, Kushnir, Moldovanu, and Shi (2013) and Manelli and Vincent (2010), who show equivalence between BIC and DIC mechanisms in settings with monetary transfers. All these results have been derived under the assumption that the private information of players is independently distributed; we deviate from this assumption by introducing correlation between the agent's type and the principal's signal.

As the principal has private information, our model is also related to the informed principal problem cf. Myerson (1983) and Maskin and Tirole (1990). With monetary transfers, Severinov (2008) and Cella (2008) show that correlated information allows for an efficient solution to the informed principal problem. We assume that the principal designs mechanisms with full commitment over allocation procedures before observing the signal. A priori, there is no informed principal problem in our model, but in the discussion section, we show that the optimal mechanism we derive also solves the informed principal problem.

Roadmap After an example highlighting the difficulties that correlation adds to the canonical verification setup, and showing how our findings advocate for transparency in pretrial investigations, section 2 sets up the model. The characterization of optimal mechanisms starts in section 3 for the class of transparent mechanisms, and section 4 shows that this mechanism is optimal in the broader class of BIC mechanisms. We then extend the analysis to a specification in which the principal's valuation may be affected by the signal realization, and discuss the relation to favored-agent mechanisms and to the informed principal problem.

I Example

With the following numerical example, we illustrate (1) how the principal can exploit correlation to lower the minimal verification cost required for implementing an allocation and (2) why the optimal allocation does not leave scope for such an improvement. Consider a defendant in court (the agent) who privately knows whether he is of the guilty or innocent type $t \in \{G, I\}$. The judge (principal) privately observes signal realization $s \in \{g, i\}$ as a result of pretrial investigations and asks the agent to plead either guilty or innocent (to report his type). Following the signal and plea, the judge decides whether to conduct a costly trial (to reveal the agent's type) and whether the defendant should be acquitted (as a function the trial's outcome, in case it was conducted). A trial requires verification cost $c > 0$. The defendant's utility is 1 from acquittal and 0 if he is convicted, irrespective of his type. Type and signal are jointly distributed according to

$$\begin{pmatrix} f_{G,g} & f_{G,i} \\ f_{I,g} & f_{I,i} \end{pmatrix} = \begin{pmatrix} \frac{2}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{6} \end{pmatrix}.$$

For illustration, we fix acquittal probabilities and consider the optimal verification schedule for the case when the defendant observes the outcome of the pretrial investigation and for the case when he does not observe this signal.

Let the guilty type be acquitted with probability $1/2$, at both signals g and i, and the innocent type with probability 1, independent of the signal. Denote by $z_{t,s} \in [0,1]$ the probability with which the type-signal combination (t,s) is verified. After verification, the judge acquits the innocent and convicts the guilty type.[2] In a transparent mechanism, the agent observes the principal's signal before making a report. The cost-minimal verification probabilities that ensure truthful reporting are[3]

$$\begin{pmatrix} z_{G,g} & z_{G,i} \\ z_{I,g} & z_{I,i} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1/2 & 1/2 \end{pmatrix}.$$

This induces a verification cost of $\frac{1}{6} \cdot 1/2\ c + \frac{2}{6} \cdot 1/2\ c = \frac{1}{4}\ c$.

If, instead, the signal realization is not known to the agent when he makes his type report, the principal can save verification costs. The above mechanism fulfills type G's incentive constraint by verifying report I with equal probability after both signal realizations. The principal can exploit the fact that type G's subjective belief puts more weight on signal g, and shift verification probability from the type-signal combination (I,i) to (I,g). The verification probabilities

$$\begin{pmatrix} z_{G,g} & z_{G,i} \\ z_{I,g} & z_{I,i} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 3/4 & 0 \end{pmatrix}$$

produce verification costs of only $\frac{1}{6} \cdot 3/4\ c = \frac{1}{8}\ c$ but ensure truthful reporting. To see this, consider the following (Bayesian) IC constraints:

$$\begin{array}{lllllll} \frac{2}{3} \cdot 1/2 & + & \frac{1}{3} \cdot 1/2 & \geq & \frac{2}{3} \cdot (1 - z_{I,g}) & + & \frac{1}{3} \cdot (1 - z_{I,i}) \quad \text{and} \\ \frac{1}{3} \cdot 1 & + & \frac{2}{3} \cdot 1 & \geq & \frac{1}{3} \cdot (1/2 - z_{G,g}) & + & \frac{2}{3} \cdot (1/2 - z_{G,i}) \quad . \end{array}$$

[2] This assumption is inessential and made for simplicity here. As will be shown in the main body of the paper, verification optimally affects only misreporting types, with the harshest possible punishment. That is, whenever verification reveals a misreport, the defendant is not acquitted.

[3] Type $t = I$ is acquitted for sure and, therefore, never has an incentive to misreport type $t = G$. Type G, who knows signal s, is willing to report his type truthfully if $1/2 \geq 1 - z_{I,s}$.

The first constraint hinders type G from reporting I. Note that $\frac{2}{3}$ is the agent's posterior belief that the signal matches his type. As before, the second constraint, hindering type I from reporting G, is satisfied independent of the verification. However, the two constraints illustrate the complication that correlation adds to the analysis: different types assign distinct probabilities to the signal realizations. Therefore, the expected utility of a given type report depends on the agent's real type. This complicates the characterization of incentive compatibility in comparison to other costly verification models.[4]

With this second verification schedule, the above-mentioned allocation,

$$\begin{pmatrix} 1/2 & 1/2 \\ 1 & 1 \end{pmatrix},$$

is not transparently implementable. If type G knows that the signal is i, he can be acquitted with probability 1 by misreporting I. This illustrates how a nontransparent procedure potentially allows the lowering of verification costs by exploiting correlation. The idea parallels the design of transfer lotteries used in Crémer and McLean (1988) and others to extract the agent's surplus.

However, conditional on the defendant's guilt, the outcome of the pretrial investigation should not affect the judge's preferences over acquittal or conviction. Therefore, we can achieve the same ex-ante expected allocation value but adjust how the probability to acquit the guilty type is distributed over the signal realizations: consider ex-post acquittal probabilities

$$\begin{pmatrix} 1/4 & 1 \\ 1 & 1 \end{pmatrix},$$

and note that the ex-ante probability for both types and, therefore, the principal's expected allocation value remain unchanged. With the above-mentioned

[4] In Ben-Porath, Dekel, and Lipman (2014) and other models of costly state verification, the absence of correlated information allows to fully characterize incentive compatibility by focusing on the type with the lowest expected utility. This technique cannot be applied to our setting. To make this explicit, consider an allocation rule that acquits the agent whenever type and signal do not match (without verification). Both types would strictly prefer to misreport. Without correlated information, such a situation cannot arise.

verification probability of $z_{I,g} = \frac{3}{4}$, this allocation can be transparently implemented. The guilty type is indifferent between truth-telling or misreporting after observing either signal.[5] This exemplifies how the allocation can be rearranged over signal realizations so that the suggested improvement in verification does not impede transparency. In the main body of the paper, we show how this can be done for the optimal allocation without violating incentives with arbitrary numbers of types and signals.

The acquittal probability in this example was fixed somewhat arbitrarily to allow for informative illustration. To consider the original problem of choosing allocation and verification probabilities jointly, the principal needs to trade off verification costs and allocation value. The paper proves that this trade-off is optimally solved by a simple cut-off mechanism. In this court example, for a range of parameter values, the optimal cut-off mechanism would feature allocation

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix},$$

implemented through verification schedule

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

This mechanism resembles the proceedings of a pretrial: The case is dismissed if the signal for the defendant's innocence is strong enough, i.e., the charge is weak. If the signal for innocence is below this cut-off, the agent can plead guilty and is convicted, or he can request a trial by pleading not guilty, after which he is acquitted if he is indeed found to be not guilty and convicted

[5]The observant reader may notice how the fact that the principal's value remains unchanged, hinges on the assumption that the signal does not have a direct effect on the allocation value for given types. This is justified in most applications, in which the signal is simply informative about the underlying fundamental type, as the outcome of pretrial investigations gives information about the defendant's guilt but should not itself affect the value of a guilty verdict. In the extensions, we show how our main result carries over with a direct effect that goes in the same direction as the informational effect, for example, when confirming the outcome of the pretrial investigations carries some value on its own.

otherwise.

An important implication of our result is that as in this simple example, the justice system cannot gain from keeping the charge secret during a pretrial. This is established practice in modern codes of procedures but was not always the case. Compare, for example, today's Austrian criminal code of procedure[6] with the code of 1803[7]. While the modern code grants the defendant the right to learn about all potential charges, the version from 1803 gives the court of inquiry much more discretion in the extent of information release to the defendant, stating that he has to be informed only as far as necessary to notify him that he is accused.

Kittler (2003) argues that this observation is in line with the broader development of continental European criminal procedure from the medieval inquisitorial proceedings, which exhibited secret charges, to modern forms of criminal law proceedings.[8]

II Model

For concreteness, in the remainder of the paper, we take the binary decision to be the allocation of a single indivisible good. The principal (she) decides whether to allocate the good to the agent (he).

Types The agent is characterized by his type T, which takes values $t \in \mathbb{R}$. We assume that the type is the agent's private information and that the set of possible types $\mathcal{T}$ is finite.

The principal privately observes a signal S with realizations $s \in \mathcal{S} \subset \mathbb{R}$, finite and ordered. This signal contains information about the agent's type: T and S are jointly distributed according to distribution $(f_{t,s})_{t\in\mathcal{T},s\in\mathcal{S}} >> 0$. The

[6]§6 (2) StPO: www.jusline.at/gesetz/stpo/paragraf/6

[7] II.3 Von Untersuchung des Beschuldigten und dem Verhöre §331

[8]Maybe the most famous defendant who is not informed about the charges he faces is Josef K., the protagonist in Franz Kafka's novel *The Trial*. In fact, Kittler (2003) suggests that Kafka, who took multiple courses in legal history before completing his law degree at the University of Prague, might have based this *process* not on the proceeding standards during his time but on medieval ones.

signal satisfies the *Monotone Likelihood Ratio Property*:

$$(\textbf{MLRP}) \;\; \forall t, t' \in T \text{ with } t < t' : \frac{f_{t',s}}{f_{t,s}} \text{ is nondecreasing in } s.$$

This is equivalent to requiring that T and S be *affiliated*. It implies that a higher signal is more indicative of a higher type.

Preferences The principal derives valuation $v(t)$ when allocating the good to an agent of type t. We normalize the value she derives from not allocating to 0. Therefore, v represents the *net* value for the principal. We assume that $v(t)$ is nondecreasing and that there are $t', t'' \in \mathcal{T}$ with $v(t') < 0 < v(t'')$ (otherwise, the principal could implement the efficient allocation decision without the agent's private information).

The signal S provides costless information about the agent's type to the principal, but does not affect her payoff from allocating the good directly. In an extension, we investigate the case in which the signal also directly affects the value of the allocation.

An agent of type t receives utility $u(t) > 0$ from the good. The agent's payoff from not receiving the good is zero.

Verification The principal has the option to learn the realization of T after paying a cost $c > 0$. Verification is perfect; the principal always learns the exact type.[9]

Solution Concept The principal can announce and commit to a verification and allocation mechanism before she learns her private signal. The realization of the signal is contractible. The agent learns his type (but not the signal) and then plays a Bayesian best response in the game that is induced by the

[9]It turns out that with commitment over subsequent allocation decisions, whether the principal learns the agent's real type or just learns whether the information he provided is wrong does not alter the results, as long as she is certain of what she learned. Erlanson and Kleiner (2017) consider the extension to the case with imperfect verification.

mechanism. We are interested in characterizing mechanisms that maximize the principal's ex-ante expected payoff in equilibrium.

II.A Mechanisms

As a first step, we can reduce the class of mechanisms that we have to consider. The arguments here are similar to those of Ben-Porath, Dekel, and Lipman (2014) but take into account the correlation between the signal and the type. The proofs and a formal version of this section are relegated to the Appendix.

Revelation Principle No loss is incurred when focusing on direct three-stage mechanisms of the following form:

1. The agent reports his type.
2. Based on this report and the signal realization, the mechanism specifies whether to verify the agent's type.
3. The mechanism specifies whether to allocate the good based on the report, the signal, and the outcome of the verification, if it was conducted.

Optimal Mechanism We can restrict the set of potential optimal mechanisms even further. As the principal wants to minimize the expected verification costs, optimal direct mechanisms need to satisfy two intuitive properties:

- Maximal Punishment: If an agent is revealed to have reported $\hat{t}$ different from his actual type t, he is awarded the good with probability 0.
- Minimal Verification: Following (t, s), the agent is verified only if, after his report is verified to be true, he receives the good for sure.

The above properties imply that we can focus on direct mechanisms $(\boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R}_+^{2|\mathcal{T}\times\mathcal{S}|}$ of the following form: $(\boldsymbol{x}, \boldsymbol{z})$ specifies for every combination of agent report and signal realization (t, s) the probabilities of three distinct events:

- With probability $x_{t,s}$, the agent gets the good <u>without</u> being verified.

- With probability $z_{t,s}$, the agent is verified and receives the good if and only if his report was truthful.
- With probability $1 - x_{t,s} - z_{t,s}$, the agent does not receive the good and is not verified.

Feasibility requires that the total allocation probability $x_{t,s} + z_{t,s} \leq 1$ for all $(t, s) \in \mathcal{T} \times \mathcal{S}$.

II.B The Agent's Problem

The agent's preferences are such that he only cares about the probability of receiving the good. Consider the incentive problem of an agent of type t. He does not know the signal realization. If he reports truthfully, he faces the random allocation probability $x_{t,S} + z_{t,S}$. Whether his report is verified is irrelevant for him. If, however, t reports $\hat{t} \neq t$, he receives the good with random probability $x_{\hat{t},S}$ only if he is not verified. Therefore, the Bayesian incentive constraint (BIC) reads as follows:

$$\forall t, \hat{t} \in \mathcal{T} : \quad u(t) \cdot \mathbb{E}_S \left[x_{t,S} + z_{t,S} \mid T = t \,\right] \;\; \geq \;\; u(t) \cdot \mathbb{E}_S \left[x_{\hat{t},S} \mid T = t \,\right].$$

As we assume that every type derives strictly positive utility from the good ($u(t) > 0$), it follows that the intensity of type t's preferences can be eliminated from the IC constraint: the agent simply maximizes his expected allocation probability. The utility he derives from the good is the same irrespective of whether he reported truthfully or not:

$$(BIC_{t,\hat{t}}) : \qquad \mathbb{E}_S \left[(x_{t,S} + z_{t,S} - x_{\hat{t},S}) \, \mathbb{1}_{\{T=t\}} \right] \; = \; \sum_{s \in \mathcal{S}} f_{t,s} \left[x_{t,s} + z_{t,s} - x_{\hat{t},s} \right] \; \geq \; 0.$$

The expected allocation probability at a certain misreport is not independent of the true type, as different types have different conditional beliefs over the distribution of S. The interim expectations are therefore insufficient to describe the mechanism.[10] In most settings without transfers and independence (Ben-Porath, Dekel, and Lipman, 2014; Erlanson and Kleiner, 2017) all types have the same interim expectations about utilities from deviations. In checking the incentive compatibility of a mechanism, it is then sufficient to restrict attention to deviations of the type with the lowest expected utility. This approach does not work in our setting.

II.C The Principal's Problem

The principal designs a mechanism that maximizes her expected utility from the allocation net of the costs of verification. If the good is assigned without verification, she gains $v(T)$. In the case of allocation with prior verification, she additionally pays cost c. Hence, the principal's problem can be stated as the following linear program:

$$
\begin{aligned}
(LP) \quad & \max_{(\boldsymbol{x},\boldsymbol{z})\geq 0} \mathbb{E}\,[\; x_{T,S}\; v(T) + z_{T,S}\; (v(T)-c)\;] \\
& \text{s.t. } \forall t,\hat{t} \in \mathcal{T}: \qquad (BIC_{t,\hat{t}}) \qquad \text{and} \\
& \forall (t,s) \in \mathcal{T}\times\mathcal{S}: \; x_{t,s} + z_{t,s} \leq 1.
\end{aligned}
$$

III Optimal Transparent Mechanisms

We call a direct mechanism transparent if the mechanism is such that the agent would report his type truthfully even if he had learned the realization of S before reporting. In our setting, transparency coincides with ex-post incentive compatibility (EPIC). The ex-post incentive constraints read as follows:

$$
\forall s \in \mathcal{S},\; \forall t,\hat{t} \in \mathcal{T}: \qquad (EPIC(s)_{t,\hat{t}}): \qquad x_{t,s} + z_{t,s} - x_{\hat{t},s} \;\geq\; 0.
$$

The Bayesian incentive constraint $(BIC_{t,\hat{t}})$ is a weighted sum of the corresponding EPIC constraints over all signal realizations. A transparent mech-

[10] This is a common feature of mechanism design with correlation: the different expected utility stemming from different beliefs is precisely how mechanisms with money exploit correlation to extract surplus.

anism is therefore necessarily BIC. In this section, we solve for the optimal transparent mechanisms. In the remainder of the paper, we show that this mechanism will also be optimal in the wider set of Bayesian incentive compatible mechanism.

Lemma 1. *The optimal transparent mechanism is as follows: For all $s \in \mathcal{S}$,*

$$\begin{cases} x_{t,s} = 1,\ z_{t,s} = 0 & \text{if} \quad \mathbb{E}_T[\ (v(T) \mid S = s\ ] > \mathbb{E}_T[\ (v(T) - c\ )^+ \mid S = s\ ] \\ x_{t,s} = 0,\ z_{t,s} = \mathbb{1}_{\{v(t) > c\}} & \text{otherwise.} \end{cases}$$

The good is allocated without verification whenever the signal alone (without considering the type report) makes the principal sufficiently optimistic about the allocation value. If she is not convinced by the signal, she will only allocate after the successful verification of the agent's report. This happens if the reported allocation value exceeds the costs of verification. The induced allocation rule is not ex-post efficient. At low signals and high types, the allocation value may be positive but smaller than the verification cost so that the good is not allocated. At high signals and low types, the good may be allocated even though $v(t) < 0$.

Implementation The optimal transparent mechanism can be implemented as a provisional decision: the principal bases her initial decision only on the signal but gives the agent the option to appeal this decision, if the allocation net verification costs is profitable. If the agent appeals, the principal verifies the agent's claim and allocates if she finds that the agent told the truth.

Appealing is weakly dominant for the agent in this implementation game, even if the appeal has no chance of success. One can prevent this multiplicity of equilibria if the designer commits to allocating the good with small probability ϵ after a negative provisional decision that is not appealed by the agent. For any $\epsilon > 0$, the unique best response of the agent is to only appeal if the appeal will be successful. For ϵ converging to zero, the loss in efficiency for the principal (compared with the optimal transparent mechanism) goes to zero, as well.

Proof of Lemma 1. For a given $s \in \mathcal{S}$, let $\boldsymbol{x}_{\cdot,s}$ denote the vector $(x_{t,s})_{\{t \in \mathcal{T}\}}$ and similarly for $\boldsymbol{z}_{\cdot,s}$.

<u>Step 0:</u> For any $s \in \mathcal{S}$, the optimal $(\boldsymbol{x}_{\cdot,s}, \boldsymbol{z}_{\cdot,s})$ can be determined separately, as all constraints only involve allocation and verification probabilities for the same signal realization. The principal's optimal expected value is the weighted sum of the values of these subproblems:

$$(LP(s)) \quad \max_{(\boldsymbol{x}_{\cdot,s}, \boldsymbol{z}_{\cdot,s}) \geq 0} \mathbb{E}_T \left[x_{T,s}\, v(T) + z_{T,s}\, (v(T) - c) \mid S = s \right]$$
$$\text{s.t. } \forall t, \hat{t} \in \mathcal{T}: \ (EPIC(s)_{t,\hat{t}}) \quad \text{and}$$
$$\forall t \in \mathcal{T}: \quad x_{t,s} + z_{t,s} \leq 1.$$

<u>Step 1:</u> *For any $s \in \mathcal{S}$ and for all $t, \hat{t} \in \mathcal{T}$: $x_{t,s} = x_{\hat{t},s}$, i.e., the allocation probability $\boldsymbol{x}_{\cdot,s}$ has to be constant in the report.*

Suppose to the contrary that there were reports t and $\hat{t}$ with $x_{\hat{t},s} > x_{t,s}$. Ex-post incentive compatibility implies that for all $\tilde{t} \in \mathcal{T}$, we have $x_{\tilde{t},s} + z_{\tilde{t},s} \geq x_{\hat{t},s} > x_{t,s}$. Hence, there could not be a type with a binding incentive constraint regarding the report t. This, in turn, implies that optimally, $z_{t,s} = 0$. If it were positive, $z_{t,s}$ could be lowered and $x_{t,s}$ could be increased, at least until the strict inequality above binds. This leaves the allocation probabilities unchanged but lowers verification costs.

The incentive constraints of type t now take the form $x_{t,s} + 0 \geq x_{\tilde{t},s}$ for all reports $\tilde{t}$ and, in particular, for report $\hat{t}$, contradicting the above hypothesis. Hence, we must have that for all t, $\hat{t}$: $x_{t,s} = x_{\hat{t},s} \equiv \chi_s$.

<u>Step 2:</u> With constant $\boldsymbol{x}_{\cdot,s}$, all incentive constraints are automatically fulfilled, as the unverified allocation probability is the same for any possible report. The principal's problem reads as follows:

$$(\text{LP(s)}) \quad \max_{(\chi_s, \boldsymbol{z}_{\cdot,s}) \geq 0} \sum_{t \in \mathcal{T}} f_{t,s} \left[\chi_s\, v(t) + z_{t,s}\, (v(t) - c) \right]$$
$$\text{s.t.} \quad \forall t \in \mathcal{T} : \chi_s + z_{t,s} \leq 1.$$

In this simplified program, $z_{t,s}$ will be set as high as possible, i.e., to $1-\chi_s$ if $(v(t,s)-c)$ is positive and to 0 otherwise, yielding the following:

$$(LP(s)) \quad \max_{\chi_s \in [0,1]} \quad \chi_s \cdot \sum_{t \in \mathcal{T}} f_{t,s}\, v(t) \quad + \quad \sum_{t \in \mathcal{T}} f_{t,s}\, (1-\chi_s)\, (v(t)-c)^+.$$

<u>Step 3:</u> Expressed in terms of conditional expectations, the problem is linear in χ_s:

$$(LP(s)) \quad \max_{\chi_s \in [0,1]} \; \chi_s \cdot \mathbb{E}_T\Big[v(T) |\; S = s\;\Big] + (1-\chi_s) \cdot \mathbb{E}_T\Big[(v(T)-c)^+ |\; S = s\;\Big].$$

Generically, the optimal value of χ is either 0 or 1, depending on which of the expectations is larger.

□

Because of the MLRP, the principal is more optimistic about the agent's type when she observes higher signals. This results in an intuitive cut-off form of the optimal transparent mechanism.

Corollary 1.
The optimal transparent mechanism is given by the following cut-off rule:

- *If the signal is above the cut-off $\overline{s}$ the good is allocated without verification (x=1).*
- *If the signal is below the cut-off $\overline{s}$, the good is allocated if and only if the allocation value net verification $v(t,s)-c$ is positive. In this case, the agent is always verified (z=1).*

Formally,

$$x_{t,s} = \mathbb{1}_{\{s \,\geq\, \bar{s}\}} \text{ and } z_{t,s} = \mathbb{1}_{\{s \,<\, \bar{s}\}} \cdot \mathbb{1}_{\{v(t) \,>\, c\}}.$$

The cut-off $\bar{s}$ is uniquely characterized by

$$\bar{s} = \min\Big\{ s \;\Big|\; \mathbb{E}_T[\; (v(T) \mid S = s\;] > \mathbb{E}_T[\; (v(T) - c\;)^+ \mid S = s\;] \Big\}\,.$$

Proof. As S and T are affiliated, the function $s \mapsto \mathbb{E}_T[v(T) - (v(T)-c)^+ |\ S = s\ ]$ is nondecreasing. Hence, there is a unique cut-off $\bar{s}$ such that

$$\begin{cases} E_T[v(T) - (v(T)-c)^+ |\ S = s\ ] \geq 0, & \forall s \geq \bar{s} \\ E_T[v(T) - (v(T)-c)^+ |\ S = s\ ] < 0, & \forall s < \bar{s}. \end{cases}$$

□

Figure 1 sketches the cut-off mechanism. If the signal realization is above the cut-off, $\bar{s}$, the principal is optimistic about T and allocates the good to the agent without verification ($x = 1$), irrespective of his reported type. If the signal is below the cut-off, the agent can receive the good only after his type report is verified ($z = 1$) to be above a threshold so that $v(t,s) - c$ is positive. It is easy to see that this mechanism is transparently implementable. Either the signal is such that the agent gets the good independent of his type (the shaded blue area to the right of $\bar{s}$), or he can only get the good after being verified (the shaded yellow area above the horizontal line and to the left of $\bar{s}$) so that misreporting a higher or lower type cannot be beneficial even when the agent knows which signal s has realized.

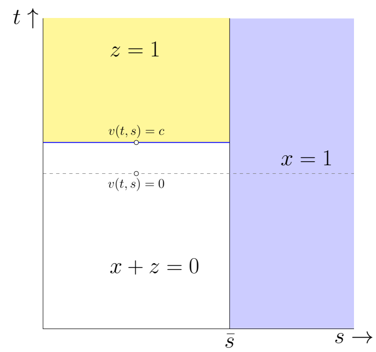


Figure 1: Transparent cut-off mechanism

IV Optimal BIC Mechanisms

This section shows that the optimal transparent mechanism is also optimal in the broader class of BIC mechanisms. This implies that the principal cannot exploit the privacy of her signal; transparency comes at no cost.

Proposition 1.

The optimal transparent mechanism is optimal in the class of Bayesian mechanisms.

To gain intuition on why the principal cannot exploit different types' beliefs, we refer back to the example in the introduction. Analogous to the method of Crémer and McLean (1988), the way in which the principal could potentially save verification costs with a nontransparent mechanism was to verify the agent who reports a high type only if the signal is low and not to reveal the signal realization to the agent.[11] Two features of the costly verification setting rule out that such an improvement exists. First, recall that in the example, the improvement was constructed by shifting verification probability to the lower signal, which is more likely to occur from the low type's perspective, and that this was the only type with an incentive to deviate. The lack of monetary transfers implies that the relevant incentive constraints in verification mechanisms are those that impede reports from less-favorable toward more favorable types. Hence, it is favorable for all *relevant* incentive constraints to shift verification probability toward lower signals.[12] Second, the signal realization does not affect the principal's valuation conditional on the agent's type. This allows a shift in the allocation probability of lower types toward exactly those signal realizations after which a misreport may be fruitful (high signals that feature less verification).[13]

The optimal transparent mechanism in the last section already distributes allocation and verification probabilities in the way that this intuition suggests. The proof of Proposition 1 makes use of the first feature by relaxing the problem and considering only a specific class of upward incentive constraints. Within the relaxed problem, we make use of the second fact to apply feasible deviations to an arbitrary mechanism that make the principal better off, and finally yield the optimal transparent mechanism.

[11]As otherwise, a low type would misreport after observing a high signal.

[12]If higher types had an incentive to deviate downward, it may save verification cost to verify after observing high signals for some type reports. This does not occur in our setting.

[13]We want to highlight that in settings with monetary transfers, even payoff-irrelevant correlated signals allow for full surplus extraction (Riordan and Sappington, 1988). In the next section, we consider the case in which the signal has a direct effect on the principal's value.

Proof. We show that the cut-off mechanism from Corollary 1 with

$$\bar{s} = \min \left\{ s \;\middle|\; \mathbb{E}_T[\ (v(T) \mid S = s\ ] > \mathbb{E}_T[\ (v(T) - c\ )^+ \mid S = s\ ] \right\}$$

solves the following relaxation of the problem, which proves that it is a solution to the original LP. Define the set of profitable types as those t with a positive allocation value,

$$\mathcal{T}^+ \equiv \{t \in \mathcal{T} | v(t) > 0\},$$

and the unprofitable types accordingly as $\mathcal{T}^- \equiv \mathcal{T} \backslash \mathcal{T}^+$. Both sets are non-empty by the assumption that v crosses 0. Otherwise, the optimal mechanism is trivial.

The relaxed problem includes only those incentive constraints that prevent types in $\mathcal{T}^-$ from misreporting types in $\mathcal{T}^+$. Hence, it reads as follows:

$$\begin{aligned}
\text{(LP.r)} \quad \max_{(\boldsymbol{x},\boldsymbol{z}) \geq 0} \quad & \sum_{t \in \mathcal{T}} \sum_{s \in \mathcal{S}} f_{t,s} \left[x_{t,s}\ v(t) + z_{t,s}\ (v(t) - c) \right] \\
\text{s.t.} \quad & \forall t \in \mathcal{T}^-,\ \forall \hat{t} \in \mathcal{T}^+ : \quad (BIC_{t,\hat{t}}) \quad \text{and} \\
& \forall (t,s) \in \mathcal{T} \times \mathcal{S} : \qquad x_{t,s} + z_{t,s} \leq 1.
\end{aligned}$$

In the remainder of the proof, we derive feasible changes to a solution to the relaxed problem, which do not lower the principal's value and which finally lead to the cut-off mechanism. We make repeated use of the following notation: we denote changes in the allocation probability by $dx_{t,s}$ so that the new probability after the change is given by $x_{t,s} + dx_{t,s}$. $dx_{t,s}$ may be positive or negative. Analogously for $dz_{t,s}$. Further, $d(BIC_{t,\hat{t}})$ denotes the change in surplus utility that type t receives from reporting the truth rather than misreporting $\hat{t}$, which is induced by a change of the above form. Recall that the constraint $(BIC_{t,\hat{t}})$ reads as $\sum_s f_{t,s} \left[x_{t,s} + z_{t,s} - x_{\hat{t},s} \right] \geq 0$ so that $d(BIC_{t,\hat{t}})$ denotes the change to the left-hand side of this inequality.

The value for the principal is given by

$$V=\sum_{t\in\mathcal{T}}\sum_{s\in\mathcal{S}} f_{t,s}\left[x_{t,s}\, v(t)+z_{t,s}\,(v(t)-c)\right],$$

and dV will denote the induced change to this value.

Step 1: The optimal mechanism in the relaxed problem features $\forall t\in\mathcal{T}^-\ \forall s\in\mathcal{S}:\ z_{t,s}=0$:

Suppose $z_{t,s}>0$ for some type $t\in\mathcal{T}^-$. Shifting probability mass from $z_{t,s}$ to $x_{t,s}$ such that the overall allocation probability stays constant,

$$0<dx_{t,s}=-dz_{t,s},$$

saves the principal verification costs and does not distort the incentives, as type t's incentive to misreport remains the same, and all incentive constraints to misreport a type $t\in\mathcal{T}^-$ are ignored in the relaxed problem.

Step 2: There is an optimal mechanism in the relaxed problem featuring a cut-off form for $\boldsymbol{x}_{\hat{t},\cdot}$:

$$\forall \hat{t}\in\mathcal{T}^+\ \exists \tilde{s}(\hat{t})\in\mathcal{S}:\ x_{\hat{t},s}\begin{cases}=0 & \text{if } s<\tilde{s}(\hat{t})\\ \in[0,1) & \text{if } s=\tilde{s}(\hat{t})\\ =1 & \text{if } s>\tilde{s}(\hat{t})\end{cases}.$$

Take a feasible IC mechanism of the relaxed problem featuring that for some $\hat{t}\in\mathcal{T}^+$, $\exists s<s'\in\mathcal{S}$ such that $x_{\hat{t},s}>0$, $x_{\hat{t},s'}<1$.

Modify the mechanism only at two points, shifting allocation probability mass from $x_{\hat{t},s}$ to $x_{\hat{t},s'}$, i.e. $dx_{\hat{t},s}<0$ and $dx_{\hat{t},s'}>0$. Choose these shifts in a proportion, such that for the highest unprofitable type, $\tilde{t}\equiv\max\mathcal{T}^-$, the incentive to misreport $\hat{t}$ remains unchanged:

$$0\overset{!}{=}d(IC_{\tilde{t},\hat{t}})=-f_{\tilde{t},s}\,dx_{\hat{t},s}-f_{\tilde{t},s'}\,dx_{\hat{t},s'}=0\Leftrightarrow dx_{\hat{t},s}=-\frac{f_{\tilde{t},s'}}{f_{\tilde{t},s}}\,dx_{\hat{t},s'}.$$

For all types $t \in \mathcal{T}^-$, we have $t \leq \tilde{t}$, and, therefore,

$$d(BIC_{t,\hat{t}}) = -f_{t,s}\, dx_{\hat{t},s} - f_{t,s'}\, dx_{\hat{t},s'} = f_{t,s} \left[\frac{f_{\tilde{t},s'}}{f_{\tilde{t},s}} - \frac{f_{t,s'}}{f_{t,s}} \right] dx_{\hat{t},s'} \geq 0$$

by the monotone likelihood ratio property.

The principal's value changes in the following way:

$$\begin{aligned} dV =& f_{\hat{t},s}\, dx_{\hat{t},s}\, v(\hat{t}) + f_{\hat{t},s'}\, dx_{\hat{t},s'}\, v(\hat{t}) = f_{\hat{t},s} \left[-\frac{f_{\tilde{t},s'}}{f_{\tilde{t},s}}\, dx_{\hat{t},s'} \right] v(\hat{t}) + f_{\hat{t},s'}\, dx_{\hat{t},s'}\, v(\hat{t}) \\ =& f_{\hat{t},s} \left[\frac{f_{\hat{t},s'}}{f_{\hat{t},s}} - \frac{f_{\tilde{t},s'}}{f_{\tilde{t},s}} \right] dx_{\hat{t},s'}\, v(\hat{t}) \geq 0, \end{aligned}$$

since $dx_{\hat{t},s'} > 0$ and $\hat{t} \in \mathcal{T}^+$ which implies both $v(\hat{t}) \geq 0$ and $\hat{t} > \tilde{t}$.

The proposed shift is clearly feasible if in the original mechanism, $x_{\hat{t},s'} + z_{\hat{t},s'} < 1$.

In the case that $x_{\hat{t},s'} + z_{\hat{t},s'} = 1$, it can still be implemented by shifting in addition mass from $z_{\hat{t},s'}$ to $z_{\hat{t},s}$ to ensure that $x_{\hat{t},s'} + z_{\hat{t},s'}$ and $x_{\hat{t},s} + z_{\hat{t},s}$ remain constant:

$$dx_{\hat{t},s'} + dz_{\hat{t},s'} = 0 \quad \text{and} \quad dx_{\hat{t},s} + dz_{\hat{t},s} = 0.$$

This implies $dz_{\hat{t},s'} < 0$ and $dz_{\hat{t},s} > 0$. This is feasible, as $x_{\hat{t},s'} < 1$ and $x_{\hat{t},s'} + z_{\hat{t},s'} = 1$ imply that $z_{\hat{t},s'} > 0$. As $x_{\hat{t},s} > 0$, we must further have $z_{\hat{t},s} < 1$ by feasibility.

The above changes in x imply for z the following:

$$dx_{\hat{t},s} = -\frac{f_{\tilde{t},s'}}{f_{\tilde{t},s}} dx_{\hat{t},s'} \Leftrightarrow dz_{\hat{t},s} = \frac{f_{\tilde{t},s'}}{f_{\tilde{t},s}}(-dz_{\hat{t},s'}).$$

The incentives for any lower type to misreport his type as $\hat{t}$ are weakened in the same way as above because $z_{\hat{t},s}$ and $z_{\hat{t},s'}$ do not play a role in the constraints that prevent misreport $\hat{t}$.

Finally, the principal's value now changes by

$$
\begin{aligned}
dV =& f_{\hat{t},s}\left[dx_{\hat{t},s}\, v(\hat{t}) + dz_{\hat{t},s}\, (v(\hat{t})-c)\right] + f_{\hat{t},s'}\left[dx_{\hat{t},s'}\, v(\hat{t}) + dz_{\hat{t},s'}\, (v(\hat{t})-c)\right] \\
=& -c\left[f_{\hat{t},s}\, dz_{\hat{t},s} + f_{\hat{t},s'}\, dz_{\hat{t},s'}\right] \\
=& -c\, f_{\hat{t},s}\left[\frac{f_{\tilde{t},s'}}{f_{\tilde{t},s}} - \frac{f_{\hat{t},s'}}{f_{\hat{t},s}}\right](-dz_{\hat{t},s'}) \geq 0,
\end{aligned}
$$

as, by MLRP, the term in squared brackets is negative and, by assumption, $-dz_{\hat{t},s'} \geq 0$.

Step 3: There is an optimal mechanism in the relaxed problem featuring $\boldsymbol{x}_{\hat{t},\cdot} = \boldsymbol{x}_{\hat{\hat{t}},\cdot}$ for all $\hat{t}, \hat{\hat{t}} \in \mathcal{T}^+$:

By the cut-off structure established in Step 2, $\boldsymbol{x}_{\hat{t},\cdot} = (0,\ldots,0,x_{\hat{t},\tilde{s}(\hat{t})},1,\ldots,1)$ for all $t \in \mathcal{T}^+$. Suppose to the contrary that $x_{\hat{t},\tilde{s}(\hat{t})} + \sum_{s>\tilde{s}(\hat{t})} 1 > x_{\hat{\hat{t}},\tilde{s}(\hat{\hat{t}})} + \sum_{s>\tilde{s}(\hat{\hat{t}})} 1$ for some $\hat{t}, \hat{\hat{t}} \in \mathcal{T}^+$.

Replacing $\boldsymbol{x}_{\hat{\hat{t}},\cdot}$ by $\boldsymbol{x}_{\hat{t},\cdot}$ does not generate new incentives to misreport, but it increases the principal's expected value, as it increases the allocation probability for profitable types. If feasibility is hurt, i.e., $x_{\hat{t},s} + z_{\hat{\hat{t}},s} > 1$ for some $s \in \mathcal{S}$, decrease $z_{\hat{t},s}$ until $x_{\hat{t},s} + z_{\hat{\hat{t}},s} = 1$. This is also a strict improvement for the principal, as she saves verification costs.

Step 4: There is an optimal mechanism in the relaxed problem featuring $\boldsymbol{x}_{\hat{t},\cdot} = \boldsymbol{x}_{\hat{\hat{t}},\cdot}$ for all $\hat{t}, \hat{\hat{t}} \in \mathcal{T} = \mathcal{T}^+ \cup \mathcal{T}^-$:

Fix some unprofitable type $t \in \mathcal{T}^-$. By Step 1, we have $\boldsymbol{z}_{t,\cdot} = 0$. Optimally, the principal wants to choose the lowest possible allocation probability for the unprofitable types. However, she needs to grant him at least the same interim allocation probability that he could achieve by misreporting to be a profitable type $\hat{t} \in \mathcal{T}^+$ (By steps 2–3, we know that $\boldsymbol{x}_{\hat{t},\cdot}$ is the same for all $\hat{t} \in \mathcal{T}^+$). As the signal realization has no effect on the allocation value, the principal is indifferent between any allocation vector $\boldsymbol{x}_{t,\cdot}$ which induces the same interim allocation probability:

$\mathbb{E}[\boldsymbol{x}_{t,S}|T = t]$. Formally,

$$\mathbb{E}[v(t,S)\boldsymbol{x}_{t,S}|T = t] = v(t)\mathbb{E}[\boldsymbol{x}_{t,S}|T = t].$$

Therefore, she can grant the unprofitable types just the same allocation lottery, they would face if they would misreport to a profitable type: $\boldsymbol{x}_{t,\cdot} = \boldsymbol{x}_{\hat{t},\cdot}$.

In the section about optimal EPIC mechanisms, we have shown how, for given s, a constant allocation $\boldsymbol{x}_{\cdot,s}$ in t implies that the principal's problem is reduced to the choice between allocating without verification at all reports and allocating after verification only if the report is such that $v(t) - c \geq 0$. This concludes the proof for the first case of the theorem, showing that the cut-off mechanism solves the relaxed problem and is therefore optimal in the original problem.

□

V Direct Signal Effects

In the previous sections, we characterized optimal mechanisms under the assumption that the signal that the principal privately observes had no direct effect on her allocation value.

In this section, we deviate from this assumption: suppose that the principal derives valuation $v : \mathcal{T} \times \mathcal{S} \to \mathbb{R}$ when allocating the good to an agent of type t at signal s. We normalize the value she derives from not allocating again to 0 and assume that $t \mapsto v(t,s)$ is nondecreasing for any $s \in \mathcal{S}$. Note that the agent's problem does not change, so neither do the incentive constraints for the transparent mechanism $(EPIC(s)_{t,\hat{t}})$ nor for the Bayesian $(BIC_{t,\hat{t}})$. The only change is in the objective of the principal, which reads as follows:

$$\max_{(\boldsymbol{x},\boldsymbol{z})\geq 0} \mathbb{E}\left[\, x_{T,S}\; v(T,S) + z_{T,S}\; (v(T,S) - c)\,\right].$$

We first characterize the optimal transparent mechanism. This characterization and the proof (in the Appendix) are analogous to Lemma 1. The reason is that by transparency the optimal allocation and verification vector can again be determined for each signal realization $s \in \mathcal{S}$ separately so that the dependence of the value of the allocation on the signal does not change the optimal solution.

Lemma 2. *The optimal transparent mechanism is as follows: For all $s \in \mathcal{S}$,*

$$\begin{cases} x_{t,s} = 1,\ z_{t,s} = 0 & \textit{if} \quad \mathbb{E}_T[\ (v(T,s) \mid S = s\ ] > \mathbb{E}_T[\ (v(T,s) - c\ )^+ \mid S = s\ ] \\ x_{t,s} = 0,\ z_{t,s} = \mathbb{1}_{\{v(t,s)>c\}} & \textit{otherwise.} \end{cases}$$

The characterization of the optimal transparent mechanism in Lemma 2 holds true for any functional form of the principal's value $v(\cdot,\cdot)$, which is increasing in the agent's type. For the remainder of the paper, we distinguish between two cases in which the direct effect is either positive or negative. We refer to the setting where there is no direct effect as case 1.

Case 2: Positive Direct Effect

First, we assume that for all t, $s \mapsto v(t,s)$ is increasing. For example the valuation could take the functional form $v(t,s) = t+s-r$, where the principal's value from allocating increases linearly in type and signal and r represents her reservation value from not allocating.

In the court example, this corresponds to a situation in which it was beneficial for the justice system that the final verdict confirm the initial charge.

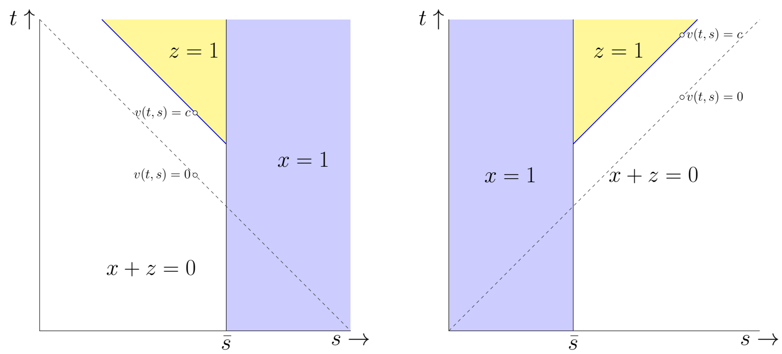
Case 3: Negative Direct Effect

By contrast, in the case in which for all t, $s \mapsto v(t,s)$ is decreasing, the direct effect is negative. Here, the direct and indirect effects go in opposite directions. This direction seems natural, for example if the principal observes the allocation value of an outside option s, which is affiliated with the allocation value of the agent t. In this example her net utility from allocating could read as follows: $v(t,s) = t - s$. We also assume that in this case, $v(t,s)$ is sufficiently decreasing such that,

$$\textbf{(Assumption 3)} \quad s \mapsto \mathbb{E}[v(T,s)|S=s] \text{ is decreasing.}$$

V.A Cut-off Mechanisms

First, we establish that in this case, the optimal transparent mechanisms again have a cut-off structure.



(a) Increasing $v(t,\cdot)$ (b) (Sufficiently) Decreasing $v(t,\cdot)$

Figure 2: Transparent cut-off mechanisms if there is a direct effect

Corollary 2.

(i) If the direct effect of the signal is positive (Case 2: $s \mapsto v(t,s)$ is increasing), then the optimal transparent mechanism is given by the following cut-off rule:

$$x_{t,s} = \mathbb{1}_{\{s \geq \bar{s}\}} \text{ and } z_{t,s} = \mathbb{1}_{\{s < \bar{s}\}} \cdot \mathbb{1}_{\{v(t,s) > c\}}.$$

(ii) If the direct effect of the signal is sufficiently negative (Case 3: $s \mapsto$ is$v(t,s)$ is decreasing) and $s \mapsto \mathbb{E}[v(T,s)|S=s]$ is decreasing (Assumption 3), then the optimal transparent mechanism is given by the following cut-off rule:

$$x_{t,s} = \mathbb{1}_{\{s \leq \bar{s}\}} \text{ and } z_{t,s} = \mathbb{1}_{\{s > \bar{s}\}} \cdot \mathbb{1}_{\{v(t,s) > c\}}.$$

The cut-off $\bar{s}$ is, in both cases, uniquely characterized by

$$\bar{s} = \min \left\{ s \,\middle|\, \mathbb{E}_T[\ (v(T,s) \mid S = s\ ] > \mathbb{E}_T[\ (v(T,s) - c\ )^+ \mid S = s\ ] \right\}.$$

V.B Optimal Bayesian Mechanisms

If the direct effect is positive, the optimal Bayesian mechanism is again given by the optimal transparent mechanism.

Proposition 2.
If the direct effect of the signal is positive (Case 2: $s \mapsto v(t,s)$), the optimal transparent mechanism is also optimal in the class of BIC mechanisms.

The proof is relegated to the Appendix. We again introduce a relaxation by defining a set of profitable types $\mathcal{T}^+$, but now, the relaxed problem has to include incentive constraints ensuring that no type t from the entire $\mathcal{T}$ has an incentive to misreport a $\hat{t} \in \mathcal{T}^+$, which is higher than t. That is, this relaxed problem ignores all downward incentive constraints and all constraints preventing misreports toward types within $\mathcal{T}^- \equiv \mathcal{T} \backslash \mathcal{T}^+$. This requires altering the construction of profitable perturbations, but the intuition is similar to the previous case. The transparent mechanism requires shifting allocation probability toward higher signals. If the principal's value is increasing in the signal, this shift is profitable and, therefore, already a feature of optimal BIC mechanisms.

The equivalence between transparent and Bayesian optimal mechanisms does not extend to the case where the direct effect is negative and $s \mapsto v(t,s)$ is decreasing (Case 3). In this case, the principal may incur loss when she chooses the EPIC optimal mechanism or, equivalently, when she publicizes the realization of the signal before the agentâs report.

Proposition 3.
If the direct effect of the signal is negative (Case 3: $s \mapsto v(t,s)$ is decreasing), and the optimal transparent mechanism $(\boldsymbol{x}, \boldsymbol{z})$ features $x_{\hat{t},s} > 0$ and $z_{\hat{t},s'} > 0$ for some $\hat{t} \in \mathcal{T}$ and some $s < s' \in \mathcal{S}$, then there exists a mechanism $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}})$ with a strictly higher value, which is BIC. Hence, the principal profits strictly from S being private.

Note that under assumption 3 ($\mathbb{E}_T[\ (v(T,s) \mid S = s\ ]$ is decreasing in s), the optimal transparent mechanism has the properties stated in the proposition whenever it is nontrivial, i.e. whenever $\boldsymbol{x}$ and $\boldsymbol{z}$ are positive at some combinations (t,s). The proof is again relegated to the Appendix.

As noted previously, the principal saves verification costs by shifting verification probability at high reports toward low signals. Low types, who have an incentive to misreport, find such signals more likely. Transparency then requires a shift in the allocation probability for low types toward the other signal realizations to ensure that at those signals indicating no verification of high reports, the low types have no incentive to deviate. If the principal's value is decreasing in the signal, shifting allocation toward higher signals comes at a cost and, therefore, transparency comes at a cost.

VI Discussion

VI.A Favored-agent mechanism

Ben-Porath, Dekel, and Lipman (2014) study the problem of allocating a single indivisible good among several agents. The principal's utility from allocating to an agent is this agent's private information which is assumed to be independently distributed across agents. The principal can perfectly verify any agent's type at an agent-specific cost. Ben-Porath, Dekel, and Lipman (2014) show that optimal mechanisms are so-called *favored-agent mechanisms*, which allocate the object to a predetermined (*favored*) agent if no other agent reports a type above his individual threshold. Whenever there is a type report above the threshold from an agent other than the favored agent, the highest type, net of verification costs, is verified and receives the good conditional on having reported the truth. In our setting, we can interpret the value s as coming from a second agent's type whose verification cost is zero.

We can interpret $v(t, s) = t+s$ as the (net) value of allocating to player one whose type is t. If we let $-s$ be the type of a second player, MLRP between type and signal means that the two players have negatively correlated values. Case 2 of our setting is fulfilled, and the EPIC mechanism from Figure 2(a) is optimal. This EPIC mechanism is essentially a favored-agent mechanism in which player one is the favored agent. If player two's type, $-s$, is low, i.e. s is above the cut-off, agent one always gets the good without being verified. If the signal is below a cut-off, agent one gets the good only if he is verified to have the highest type, net of verification costs, i.e., if $t - c \geq -s - 0$.

Similarly, we can interpret $v(t, s) = t-s$ as the value of allocating to player one if player two's type is s, so that MLRP implies positive correlation. As the utility function $v(t, s) = t - s$ is decreasing in s, case 3 is fulfilled and therefore (proposition 3) the EPIC favored agent mechanism is not generally, the optimal mechanism.

The difference between the present paper and Ben-Porath, Dekel, and Lipman (2014) is the correlation between players' types through MLRP. In contrast to their setting, here, different types hold different beliefs and therefore expect different interim allocations for a fixed type report. Hence, one cannot reduce the mechanism to interim allocation probabilities for each report.

Despite the technical differences in solving for the optimal mechanism, we find a close connection to favored-agent mechanisms (which are optimal under independence) and we infer that they remain optimal under negative correlation but not under positive correlation.

VI.B Informed principal problem

By our assumptions on the principal's commitment, the mechanism proposed by the designer does not convey information to the agent. However, the fact that the principal's signal can be made public without loss implies that the informed principal game has a separating equilibrium in which the agent perfectly learns the principal's type from the proposal. This implies that our mechanism constitutes a solution to the informed principal problem for Cases 1 and 2 when the EPIC mechanism is optimal.

VII Concluding Remarks

This paper studies the role of information in a mechanism design model, in which the principal may use costly verification instead of monetary transfers to incentivize the revelation of private information. We show that a transparent mechanism is optimal. It is without loss for the principal to make her information public before contracting with the agent. Our result gives a rationale for the use of transparent procedures in a variety of applications from hiring to procedural law. This is in contrast with results on correlation in mechanism design problems with money.

In an extension in which the principal's private information also affects her preferences, we characterize the mechanism and show that the above qualities remain if the information and direct effect work in the same direction. In the opposite case, we show how the principal can benefit by ensuring that her signal remains private. Interesting directions for future analysis include the question of whether, with correlation, the equivalence between BIC and transparent (EPIC) mechanisms holds more generally than for optimal mechanisms.

References

Elchanan Ben-Porath, Eddie Dekel, and Barton L. Lipman. Optimal allocation with costly verification. *American Economic Review*, 104(12):3779–3813, 2014.

Elchanan Ben-Porath, Eddie Dekel, and Barton L Lipman. Mechanisms with evidence: Commitment and robustness. *Working Paper*, 2017.

Mohit Bhargava, Dipjyoti Majumdar, and Arunava Sen. Incentive-compatible voting rules with positively correlated beliefs. *Theoretical Economics*, 10(3):867–885, 2015.

Michela Cella. Informed principal with correlation. *Games and Economic Behavior*, 64(2): 433–456, 2008.

Jaques Crémer and Richard P. McLean. Full extraction of the surplus in bayesian and dominant strategy auctions. *Econometrica*, 56(6):1247–1257, 1988.

Dominique M Demougin and Devon A Garvie. Contractual design with correlated information under limited liability. *The RAND Journal of Economics*, pages 477–489, 1991.

Albin Erlanson and Andreas Kleiner. Costly verification in collective decisions. 2017.

Douglas Gale and Martin Hellwig. Incentive-compatible debt contracts: The one-period problem. *Review of Economic Studies*, 52(4):647–663, 1985.

Alex Gershkov, Jacob K. Goeree, Alexey Kushnir, Benny Moldovanu, and Xianwen Shi. On the equivalence of bayesian and dominant strategy implementation. *Econometrica*, 81(1): 197–220, 2013.

Allan Gibbard. Manipulation of voting schemes: a general result. *Econometrica*, 41(4): 587–601, 1973.

Jacob Glazer and Ariel Rubinstein. On optimal rules of persuasion. *Econometrica*, 72(6): 1715–1736, 2004.

Marina Halac and Pierre Yared. Commitment vs. flexibility with costly verification. *Working Paper*, 2019.

Sergiu Hart, Ilan Kremer, and Motty Perry. Evidence games: truth and commitment. *The American Economic Review*, 107(3):690–713, 2017.

Scott Johnson, John W Pratt, and Richard J Zeckhauser. Efficiency despite mutually payoff-relevant private information: The finite case. *Econometrica*, pages 873–900, 1990.

Wolf Kittler. Heimlichkeit und schriftlichkeit: Das österreichische strafprozessrecht in franz kafkas roman der proceş. *The Germanic Review: Literature, Culture, Theory*, 78(3): 194–222, 2003.

Alejandro M. Manelli and Daniel R. Vincent. Bayesian and dominant-strategy implementation in the independent private-values model. *Econometrica*, 78(6):1905–1938, 2010.

Eric Maskin and Jean Tirole. The principal-agent relationship with an informed principal: The case of private values. *Econometrica: Journal of the Econometric Society*, pages 379–409, 1990.

R Preston McAfee and Philip J Reny. Correlated information and mecanism design. *Econometrica*, pages 395–421, 1992.

Roger B Myerson. Mechanism design by an informed principal. *Econometrica: Journal of the Econometric Society*, pages 1767–1797, 1983.

Tymofiy Mylovanov and Andriy Zapechelnyuk. Optimal allocation with ex post verification and limited penalties. *American Economic Review*, 107(9):2666–94, September 2017.

Zvika Neeman. The relevance of private information in mechanism design. *Journal of Economic Theory*, 117(1):55–77, 2004.

Michael H Riordan and David EM Sappington. Optimal contracts with public ex post information. *Journal of Economic Theory*, 45(1):189–199, 1988.

Mark Allen Satterthwaite. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of economic theory*, 10(2):187–217, 1975.

Sergei Severinov. An efficient solution to the informed principal problem. *Journal of Economic Theory*, 141(1):114–133, 2008.

Robert M. Townsend. Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, 21(2):265–293, October 1979.

Appendix A

Revelation Principle

The revelation principle presented here is close to the revelation principle in Ben-Porath, Dekel, and Lipman (2014), but it takes into account possible issues arising from the correlation between the signal and the type realization.
Pick any (possibly dynamic) mechanism G and an agent strategy s_A that is a best response to this mechanism. Then, there is an equivalent incentive compatible, direct, two-stage mechanism characterized by the pair of functions (e, a),

$$e : \mathcal{T} \times \mathcal{S} \to [0, 1],$$
$$a : \mathcal{T} \times \mathcal{T} \cup \{\emptyset\} \times \mathcal{S} \to [0, 1],$$

of the following form:

1. The agent reports his type $\hat{t} \in \mathcal{T}$.
2. Given her signal realization s, the principal verifies the agent's type with probability $e(\hat{t}, s)$.
3. Depending on the result of this revision $t \in \mathcal{T} \cup \{\emptyset\}$, where $\emptyset$ encodes the event that there was no revision, the principal allocates the good to the agent with probability $a(\hat{t}, t, s)$.

Instead of G, the principal could commit to the following mechanism:

- The agent reports a type $\hat{t} \in \mathcal{T}$.
- Given this report and her signal's realization, s, the principal calculates the marginal probability of verification in the equilibrium in the original game under the condition that the agent's type was $\hat{t}$ and the principal's signal was s:[14]

$$e(\hat{t}, s) := \mathbb{P}(\text{there is verification}|s_A(\hat{t}), S = s).$$

- The principal verifies the agent's true type with this probability: $e(\hat{t}, s)$.
  - If she finds that the agent reported the truth, $\hat{t} = t$, or if she did not verify $t = \emptyset$, she allocates the good with probability that equals the marginal probability of allocation in the original mechanism, conditional on the type being equal to $\hat{t}$ and the signal being equal to s.

$$a(\hat{t}, \hat{t}, s) = a(\hat{t}, \emptyset, s) = \mathbb{P}(\text{allocation}|s_A(\hat{t}), S = s).$$

  - If she verifies and finds out that the agent misreported, i.e., $t \notin \{\hat{t}, \emptyset\}$, the allocation probability is determined in the following way:
    The principal constructs a lottery over all stages in the original mechanism, which have the principal verify the agent with positive probability in equilibrium, conditional on the event that the agent played according to $s_A(\hat{t})$ and

[14] This means the probability that there was verification at any point in the game, specified by G and played by the agent according to $s_A(\hat{t})$, under the condition that signal s realized.

the signal was s.
The probabilities of the lottery are chosen such that they equal the probability of verifying at this stage for the first time, conditional on the event that there is verification at some point in the game.
Now, she chooses one of these stages according to the above probabilities. She simulates the game from this point onward, assuming that the game had reached this stage and it was found at this point that the agent's true type was t, by letting the simulated agent behave according to what is described in $s_A(t)$ for behavior after this knot and the verification. The strategy s_A contains a plan for the behavior of the agent from this stage onward. The principal simulates his own behavior, as prescribed in the original mechanism.
Note that given any signal realization, this reproduces exactly the same allocation profile as the following deviation strategy for type $t \neq \hat{t}$ (which he could play without knowing the true signal realization s):
The agent of type t imitates type $\hat{t}$'s behavior $s(\hat{t})$ until the first verification, and then sticks to the behavior that the equilibrium strategy prescribes for his type.

If the agent reports the truth, the marginal probabilities of verification and allocation and, therefore, the expected utilities of the agent and the principal are the same in both mechanisms.
However, truth-telling is optimal for the agent in the constructed mechanism, as misreporting yields the exact same outcome as the above-described deviation strategy in the original game and can therefore be not profitable.

There are two further observations that help simplify the class of possible optimal mechanisms. In short, in any optimal mechanism, the principal will chose he highest possible punishment for detected misreports and the highest possible reward for detected truth-telling.

1. Maximal punishment: $t \notin \{\hat{t}, \emptyset\} \Rightarrow a(\hat{t}, t, s) = 0$
As the mechanism is direct, in equilibrium, the agent will not lie; therefore, decreasing $a(\hat{t}, t, s)$ for $t \notin \{\hat{t}, \emptyset\}$ will not affect the expected utility of the mechanism designer. This deviation only increases the incentives to report truthfully. Therefore in can be assumed WLOG that the optimal mechanism features maximal punishment.

2. Maximal reward: $e(\hat{t}, s) > 0 \Rightarrow a(\hat{t}, \hat{t}, s) = 1$.
Suppose $a(\hat{t}, \hat{t}, s) < 1$. One could now lower the probability of verification $de(\hat{t}, s) < 0$ while increasing the probability of allocation after confirming the report as true $da(s, \hat{t}, \hat{t}) > 0$ such that $d(e(\hat{t}, s)a(\hat{t}, \hat{t}, s)) = 0$.
Lowering the verification probability would only increase the incentives to misreport and the overall allocation probability after report $\hat{t}$ and signal s, if there was allocation with positive probability conditional on no verification, i.e. $a(s, \hat{t}, \emptyset) > 0$. However, in this case, this allocation could be lowered $da(s, \hat{t}, \emptyset) < 0$ such that $d((1 - e(\hat{t}, s))a(s, \hat{t}, \emptyset)) = 0$, and the incentives to misreport and the overall allocation probability would remain constant. As these procedure would save verification costs while keeping all unconditional allocation probabilities constant, the fact that an optimal mechanism features non-maximal reward can be ruled out.

These observations fix the allocation after verification. Effectively the mechanism designer therefore has to choose only the verification probability $e(t, s)$ and the allocation probability, conditional on no verification $a(s, t, \emptyset)$.
For convenience, define $z_{t,s} = e(t, s)$, the joint probability of verification and allocation, and $x(t, s) = (1 - e(t, s))a(t, \emptyset, s)$, the joint probability of no verification and allocation.
Note that the set of mechanisms described by

$$\{(x_{t,s}, z_{t,s})_{t\in\mathcal{T},s\in\mathcal{S}} | \forall t \in \mathcal{T}\ \forall s \in \mathcal{S} :\ 0 \leq x_{t,s} + z_{t,s} \leq 1\}$$

is equivalent to all maximal reward and punishment, two-stage, direct mechanisms.[15]

Appendix B

Proofs not included in the paper

Proof of proposition 2

Step 0: (Relaxation) Define the following cut-off in the signal space:

$$\overline{s} = \min\{s|\mathbb{E}[v(T,s)|S=s] > \mathbb{E}[(v(T,s)-c)^+]\},$$

where $(a)^+ = \max\{0,a\}$ and we use the convention that $\min\emptyset = \max\mathcal{S}$. Note that $v(t,s) - (v(t,s)-c)^+ = \min\{c, v(t,s)\}$ is increasing in both components. Due to the MLRP it follows that $E[v(T,s)|S=s] - \mathbb{E}[(v(T,s)-c)^+$ is increasing in s.

Next, define the set of profitable types $\mathcal{T}^+ = \{t \in \mathcal{T}|\, v(t,\overline{s}) \geq 0\}$. We denote all types that are not profitable by $\mathcal{T}^- = \mathcal{T} - \mathcal{T}^+$.

The relaxed problem ignores certain incentive constraints. It optimizes the same objective function but only subject to:

$$\forall \hat{t} \in \mathcal{T}^+\ \forall t \in \mathcal{T} \text{ with } t < \hat{t}\text{: } (IC_{t\hat{t}})$$

Step 1: There is an optimal solution to the relaxed problem that takes a cut-off form in x for all $t \in \mathcal{T}$:

$$\forall t \in \mathcal{T}\ \exists\ \tilde{s}(t) \in \mathcal{S}:\ x_{t,s} = \begin{cases} 0 & \text{if } s < \tilde{s}(t) \\ x_{t,s} \in [0,1) & \text{if } s = \tilde{s}(t) \\ 1 & \text{if } s > \tilde{s}(t) \end{cases}.$$

Proof. Suppose there is a (relaxed) incentive compatible mechanism which has for some $t \in \mathcal{T}$ and $s' < s'' \in \mathcal{S}$, that $x_{t,s} > 0$ and $x_{t,s'} < 1$.

In the following, we consider a shift in allocation probability from $x_{t,s'}$ to $x_{t,s''}$ that keeps the overall allocation probability for type t constant:

$$f_{t,s'}\ dx_{t,s'} + f_{t,s''}\ dx_{t,s''} = 0 \Leftrightarrow \underbrace{dx_{t,s'}}_{<0} = -\frac{f_{t,s''}}{f_{t,s'}} \underbrace{dx_{t,s''}}_{>0}.$$

First, for any type $t^- < t$ the probability of receiving the good without verification after a

[15] The inverse mapping is given by

$$e(t,s), a(t,\emptyset,s)) = \left(z_{t,s}, \frac{x_{t,s}}{1-z_{t,s}}\right).$$

Note that the value of $a(t,\emptyset,s)$ does not play any role in the mechanism if $e(t,s) = z_{t,s} = 1$ and can therefore be chosen arbitrarily.

misreport t decreases in the new mechanism:

$$f_{\tilde{t},s'}dx_{t,s'} + f_{\tilde{t},s''}dx_{t,s''} = -f_{\tilde{t},s'}\left[\frac{f_{t,s''}}{f_{t,s'}} - \frac{f_{\tilde{t},s''}}{f_{\tilde{t},s'}}\right]dx_{t,s''} \leq 0.$$

The last inequality holds since the likelihood ratio is increasing. The shift yields type t the same allocation probability, so he cannot have a new incentive to misreport. Therefore, all relaxed incentive constraints survive.

Second, the modified mechanism yields the principal a higher expected value:

$$f_{t,s'}\ dx_{t,s'}\ v(t,s') + f_{t,s''}\ dx_{t,s''}\ v(t,s'') = f_{t,s''}dx_{t,s''}\left[-v(t,s') + v(t,s'')\right] > 0$$

The proposed shift is clearly feasible if in the original mechanism $x_{t,s''} + z_{t,s''} < 1$. In the case that $x_{t,s''} + z_{t,s''} = 1$, it can still be implemented by shifting in addition mass from $z_{t,s''}$ to $z_{t,s'}$ such that $x_{t,s''} + z_{t,s''}$ and $x_{t,s'} + z_{t,s'}$ stay constant:

$$dx_{t,s'} = -dz_{t,s'} \ \text{ and } \ dx_{t,s''} = -dz_{t,s''}.$$

As we assume $x_{t,s''} < 1$ and $x_{t,s''} + z_{t,s''} = 1$, we have that $z_{t,s'} > 0$. Since $x_{t,s'} > 0$ we also have $z_{t,s'} < 1$.

The incentives for any lower type to misreport his type as t are weakened in the same way as above since $z_{t,s}$ and $z_{t,s'}$ do not play a role in these constraints. The incentive for t to misreport is not affected since the total allocation probability $x + z$ is kept constant.

Further, the principal's expected value is not changed by this shifts:

$$f_{t,s'}\left[dx_{t,s'}\ v(t,s') + dz_{t,s'}\ (v(t,s') - c)\right] + f_{t,s''}\left[dx_{t,s''}\ v(t,s'') + dz_{t,s''}\ (v(t,s'') - c)\right]$$
$$= -c\ f_{t,s''}\left[\frac{f_{t,s''}}{f_{t,s'}} - \frac{f_{t,s''}}{f_{t,s'}}\right](-dz_{t,s'}) = 0.$$

The reason is, that the allocation $(x + z)$ remains the same with these shifts, so that only the verification cost changes. Yet, the change in verification is such that it does change the expected verification probability for the true type t and therefore neither the expected verification cost for the principal. □

Step 2: The optimal mechanism in the relaxed problem features,

$$\forall t \in \mathcal{T}^- \ \forall s \in \mathcal{S}: \ z_{t,s} = 0.$$

Proof. In the relaxed problem we disregard all incentive constraints that prevent the agent to misreport his type as $t \in \mathcal{T}^-$. If there were some $t \in \mathcal{T}^-$ and $s \in \mathcal{S}$ with $z_{t,s} > 0$, shifting probability mass from $z_{t,s}$ to $x_{t,s}$ by

$$\underbrace{dz_{t,s}}_{<0} = -\underbrace{dx_{t,s}}_{>0},$$

would save the principal verification costs while keeping the overall allocation probability constant. It would therefore not affect the incentive constraints in the relaxed problem. □

Step 3: We can assume that the optimal mechanism also takes a cut-off form in $x + z$:

$$\forall t \in \mathcal{T}\ \exists \underline{s}(t) \in \mathcal{S}:\ x_{t,s} + z_{t,s} = \begin{cases} 0 & \text{if } s < \underline{s}(t) \\ x_{t,s} + z_{t,s} \in [0,1) & \text{if } s = \underline{s}(t) \\ 1 & \text{if } s > \underline{s}(t) \end{cases}.$$

Proof. For $t \in \mathcal{T}^-$, this property follows immediately from the previous two steps with $\underline{s}(t) = \tilde{s}(t)$.

Suppose for $t \in \mathcal{T}^+$ that there exist $s' < s'' \in$ with $z_{t,s'} > 0$ and $x_{t,s''} + z_{t,s''} < 1$. To rule out this possibility, consider a shift in mass from $z_{t,s'}$ to $z_{t,s''}$ in a way that the allocation probability for a truth-telling agent of type t remains constant, i.e.,

$$\underbrace{dz_{t,s''}}_{>0} = \frac{f_{t,s'}}{f_{t,s''}}(-\underbrace{dz_{t,s'}}_{<0}).$$

Note that this shift is feasible by assumption and that it will keep all relaxed incentive constraints unchanged, since the true type t receives the same expected allocation probability, and $\boldsymbol{z}_{t,\cdot}$ does not play a role in the IC constraints preventing misreport t.

From the principal's point of view, it is favorable because it keeps the verification probability and thus the costs constant, while shifting allocation mass from (t, s') to the more favorable type–signal pair (t, s''), i.e.

$$dV = f_{t,s'} dz_{t,s'}[v(t,s') - c] + f_{t,s''} dz_{t,s''}[v(t,s'') - c] = 0 \cdot c + f_{t,s'}[v(t,s'') - v(t,s')](-dz_{t,s'}) > 0.$$

□

Step 4: The following restriction of the relaxed problem is without loss for the principal. Require, additionally, that the IC constraints hold point-wise. That is $\forall t \in \mathcal{T}$, $\forall \hat{t} \in \mathcal{T}^+$ with $t < \hat{t}$ and $\forall s \in \mathcal{S}$:

$$(EPIC(s)_{t,\hat{t}}) = x_{t,s} + z_{t,s} - x_{\hat{t},s} \geq 0.$$

An optimal solution to the restricted relaxed problem therefore also solves the relaxed problem.

Proof. By the above steps, the (Bayesian) IC constraints in the relaxed problem can be written as follows:[16]

$\forall t \in \mathcal{T}\ \forall \hat{t} \in \mathcal{T}^+$ with $\hat{t} > t$:

$$\begin{aligned}(IC_{t,\hat{t}}) &= \sum_{s \in \mathcal{S}} f_{t,s}(x_{t,s} + z_{t,s}) - \sum_{s \in \mathcal{S}} f_{t,s} x_{\hat{t},s} \\ &= f_{t,\underline{s}(t)}(x_{t,\underline{s}(t)} + z_{t,\underline{s}(t)}) + \sum_{s > \underline{s}(t)} f_{t,s}\, 1 - (f_{t,\tilde{s}(\hat{t})}\, x_{\hat{t},\tilde{s}(\hat{t})} + \sum_{s > \tilde{s}(\hat{t})} f_{t,s}\, 1) \overset{!}{\geq} 0.\end{aligned}$$

[16] Making use of the fact that $z_{t,s} = 0$ for all $t \in \mathcal{T}^-$.

This condition clearly requires that $\underline{s}(t) \leq \tilde{s}(\hat{t})$ and, in the case of equality, $x_{t,\underline{s}(t)} + z_{t,\underline{s}(t)} \geq x_{\hat{t},\tilde{s}(t)}$. Because by the definition of $\underline{s}(t)$, $x + z$ is equal to 0 below and equal to 1 above this threshold, we can conclude that for all s, $x_{t,s} + z_{t,s} \geq x_{\hat{t},s}$, which implies $(EPIC(s)_{t,\hat{t}})$. □

Step 5: Consider an optimal mechanism of the above cut-off structure that satisfies $(EPIC(s)_{t,\hat{t}}) \geq 0$ for all $t \in \mathcal{T}$, for all $\hat{t} \in \mathcal{T}^+$ with $\hat{t} > t$, and for all $s \in \mathcal{S}$. Then, it also must hold for $t, \hat{t} \in \mathcal{T}^-$ with $t < \hat{t}$, that $(EPIC(s)_{t,\hat{t}}) \geq 0$. This means that no unprofitable type has an incentive to report a higher unprofitable type.

Proof. Assume that for some $s \in \mathcal{S}$ there are types $t < \hat{t} \in \mathcal{T}^-$ such that the constraint $(EPIC(s)_{t,\hat{t}})$ is violated.

Define $s' \equiv \min\{s \in \mathcal{S} | \exists t < \hat{t} \in \mathcal{T}^- \colon x_{t,s} + z_{t,s} < x_{\hat{t},s}\}$ to be the lowest signal for which some type t profits from a higher report $\hat{t} \in \mathcal{T}^-$. Let $t' \equiv \min\{t \in \mathcal{T}^- | \exists \hat{t} \in \mathcal{T} \text{ with } \hat{t} > t \colon x_{t',s'} < x_{\hat{t},s'}\}$ be the smallest type with $EPIC(s')$ incentives to misreport his type to some type $\hat{t} \in \mathcal{T}^-$.

Since $z_{t',s'} = 0$ for the unprofitable type t' (Step 3), this implies $x_{t,s'} < x_{\hat{t},s'}$. As $t', \hat{t} \in \mathcal{T}^-$, it follows that $\hat{t}$'s EPIC(s') constraints are slack for all reports in $\mathcal{T}^+$. Having $x_{\hat{t},s'} > x_{t',s'} \geq 0$ can therefore only be optimal in the relaxed problem if $v(\hat{t}, s') \geq 0$. This implies that $s' > \bar{s}$ since $\mathcal{T}^-$ is precisely defined as the set of types t with $v(t, \bar{s}) < 0$.

Since $x_{t',s'} < x_{\hat{t},s'} \leq 1$, taking the cut-off structure from step 1 into account we can infer that for all $s < s'$ it holds that $x_{t',s'} = 0$. In particular we have $x_{t',\bar{s}} = 0$.

By the minimality of s' we get that $0 = x_{t',\bar{s}} \geq x_{\tilde{t},\bar{s}}$ for all $\tilde{t} \in \mathcal{T}^-$ with $\tilde{t} > t'$. By the minimality of t' we get that $x_{\tilde{t},\bar{s}} \leq x_{t',\bar{s}} = 0$ for all $\tilde{t} \in \mathcal{T}^-$ with $\tilde{t} < t'$. Furthermore, by $EPIC(\bar{s})$ IC compatability, it follows that $0 = x_{t,\bar{s}} \geq x_{t',\bar{s}}$ for all $t' \in \mathcal{T}^+$. So we have that $x_{t,\bar{s}} = 0$ for all t which by definition of $\bar{s}$ cannot be optimal in the restricted problem (for which this mechanism is also optimal) as $\mathbb{E}_T[\ (v(T, \bar{s}) \mid S = \bar{s}\ ] > \mathbb{E}_T[\ (v(T, \bar{s}) - c\ )^+ \mid S = \bar{s}\ ] \geq 0$. □

Step 6: Consider an optimal mechanism of the above cut-off structure that satisfies $(EPIC(s)_{t,\hat{t}}) \geq 0$ for all $t \in \mathcal{T}$, for all $\hat{t} \in \mathcal{T}^+$ with $\hat{t} > t$, and for all $s \in \mathcal{S}$. Then, for $t, \hat{t} \in \mathcal{T}$ with $t > \hat{t}$ it mus also satisfy $(EPIC(s)_{t,\hat{t}}) \geq 0$. This means that no type benefits from reporting any <u>lower</u> type.

Proof. Assume that there are types $t'' > t' \in \mathcal{T}$such that $x_{t'',s} + z_{t'',s} < x_{t',s}$ for some s. WLOG let t'' be the lowest type for which such a downward deviation is profitable.

Optimality of the relaxed mechanism requires then that $\mathbb{E}_T[\ v(T, s)\mathbb{1}_{\{T \leq t'\}} \mid S = s\ ] > 0$. Otherwise the principal would be better off by lowering x for all types below t'' (Note that $x_{t'',s} + z_{t'',s} < x_{t',s}$ implies that types $t < t''$ cannot have binding upwards constraints towards reports higher than t'' as this would violate the upward constraints for t''). Monotonicity of the value in the type in turn implies that $v(t, s) > 0$ for all $t > t'$. This contradicts optimality as the designer could increase $x_{t,s}$ for all higher types without violating any incentives. Either by just increasing $x_{t,s}$ if $x_{t,s} + z_{t,s} < 1$ or by lowering $z_{t,s}$ at the same time to save verification costs. □

Step 7: This concludes the proof. We have shown that optimal solution to the relaxed problem is EPIC (Step 4). Therefore the principal's expected value in the original problem cannot exceed the expected value from this optimal EPIC solution. In step 5 and 6 we ruled out the two possible violations of the original (ex-post) incentive constraints that can arise in a solution to the relaxed problem. Hence the candidate solution is also EPIC in the original problem. In particular, it is Bayesian incentive compatible and therefore a solution to the original problem.

Proof of proposition 3

Proof. To prove the claim, we construct an improvement that will not violate the Bayesian incentive constraints. This suffices to show that the principal strictly profits from ensuring that the realization of S remains private because the improved mechanism will implement the same allocation at lower verification costs. Consider the shift of mass from $z_{\hat{t},s'}$ to $z_{\hat{t},s}$ and —in order to maintain the overall allocation $\boldsymbol{x}+\boldsymbol{z}$ unchanged— vice versa for $x_{\hat{t},s'}$ and $x_{\hat{t},s}$:

$$dx_{\hat{t},s'} + dz_{\hat{t},s'} = 0 \text{ and } dx_{\hat{t},s} + dz_{\hat{t},s} = 0.$$

To ensure that the Bayesian incentive constraints of all types $t < \hat{t}$ are not violated by the shift, we require that

$$\forall t < \hat{t}: \; d(IC_{t,\hat{t}}) = -f_{t,s}\, dx_{\hat{t},s} - f_{t,s'}\, dx_{\hat{t},s'} \geq 0 \Leftrightarrow dx_{\hat{t},s'} \leq \frac{f_{t,s}}{f_{t,s'}} \left(-dx_{\hat{t},s}\right).$$

The proposed change has $(-dx_{\hat{t},s}) > 0$, and $\frac{f_{t,s}}{f_{t,s'}}$ is decreasing in t. Hence, the right-hand side of the above expression is minimized at $t' = \max\{t \in \mathcal{T} | t < \hat{t}\}$. Note that $\hat{t} \neq \min\{t \in \mathcal{T}\}$, as this would imply that at $v(t,s') > c > 0$ at all levels of t so that the optimal mechanism would allocate without verification after this signal.

Setting $dx_{\hat{t},s'} = \frac{f_{t',s}}{f_{t',s'}} \left(-dx_{\hat{t},s}\right)$ ensures that the incentives to misreport toward $\hat{t}$ are weakened for all lower types.

The above changes in x imply for z:

$$dx_{\hat{t},s} = \frac{f_{t',s'}}{f_{t',s}}(-dx_{\hat{t},s'}) \Leftrightarrow (-dz_{\hat{t},s}) = \frac{f_{t',s'}}{f_{t',s}} dz_{\hat{t},s'}.$$

The principal's value changes by:

$$\begin{aligned} dV =& f_{\hat{t},s} \left[dx_{\hat{t},s}\, v(\hat{t},s) + dz_{\hat{t},s}\, (v(\hat{t},s) - c)\right] + f_{\hat{t},s'} \left[dx_{\hat{t},s'}\, v(\hat{t},s') + dz_{\hat{t},s'}\, (v(\hat{t},s') - c)\right] \\ =& -c\left[f_{\hat{t},s}\, dz_{\hat{t},s} + f_{\hat{t},s'}\, dz_{\hat{t},s'}\right] \\ =& -c\, f_{\hat{t},s} \left[\frac{f_{t',s'}}{f_{t',s}} - \frac{f_{\hat{t},s'}}{f_{\hat{t},s}}\right] (-dz_{\hat{t},s'}) > 0. \end{aligned}$$

The second equality follows because the allocation remains the same with these shifts, so that only the verification cost changes.

Lastly, note that in the optimal EPIC mechanism, $z_{\hat{t},s} = 1$ implies $z_{t,s} = 1$ for all $t > \hat{t}$ and that $\boldsymbol{x}_{\cdot,s}$ is constant in the report at all s. Therefore, the fact that $z_{\hat{t},s} = 1$ in the original mechanism implies that the Bayesian IC constraints for higher types to lie downward to $\hat{t}$ are slack so that we can always find a shift in magnitude small enough to not violate these constraints.

The only case in which these constraints are not slack in the optimal EPIC mechanism is when several types receive exactly the same allocation. In this case, the above improvement can be applied to the highest report in this class. □

Proof of Lemma 2

Proof. Similar to Step 0 in Lemma 1, for any $s \in \mathcal{S}$, the optimal $(\boldsymbol{x}_{\cdot,s}, \boldsymbol{z}_{\cdot,s})$ can be determined separately, as all constraints only involve allocation and verification probabilities for the same signal realization.

This results in $|\mathcal{S}|$ separate problems, one for each possible signal realization $s \in \mathcal{S}$. In these separate problems $v(t, s)$ is a function of t only. Therefore, for all $s \in \mathcal{S}$, all steps in the proof of Lemma 1 can be replicated with $v(T)$ replaced by $v(T, s)$.

□