Costly Verification and Correlated Information

Deniz Kattwinkel and Jan Knoepfle*

University of Bonn

16th April 2018

Abstract

A principal has to take a binary decision. She relies on information privately held by an agent who always prefers one of the two actions. The principal cannot use monetary transfers to incentivise truthful reports but has the possibility to reveal the agent's information at a cost. Additionally, the principal privately observes a signal which is correlated with the agent's type. We show that optimal mechanisms take a simple cut-off structure: If the principal observes a signal above the cut-off, she takes the agent's preferred action, independent of the type report. If the signal falls below the cut-off, she takes the non-preferred action unless the agent's type is verified to be above a certain threshold. The cut-off mechanism is robustly implementable. In contrast to standard results on mechanism design with correlation and monetary transfers, the principal does not exploit the fact that different types have different beliefs. Without loss for the principal, the signal realisation can be made public before the agent reports his type.

Keywords: Mechanism design without transfers, Costly verification, Correlated information **JEL Codes:** D82

*Knoepfle: University of Bonn, Lennéstr. 43, 53113 Bonn, Germany, (e-mail: J.K@uni-bonn.de). Kattwinkel: University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany. We are indebted to Stephan Lauermann, Benny Moldovanu, Sven Rady, and Andre Speit for insightful discussions and suggestions. We also thank Marina Halac, Andreas Kleiner, Daniel Krähmer, Martin Pollrich, 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, and the Yale Theory Breakfast for fruitful comments.

Introduction

Consider a decision maker who faces two possible actions. For example, the Human Resource Department decides whether to hire a candidate, the judge acquits or convicts a defendant, or the competition authority grants or denies a company permission to merge with or acquire another firm.

While one party – the agent – has a clear preference towards one action (the candidate wants to be hired, the defendant acquitted, and the company to merge), the preferences of the other party – the principal – depend on information which is privately held by the agent. 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 (be it for practical or moral reasons[1]), but at a given cost, the principal can learn it. For example, by conducting an assessment centre, a trial, or a market analysis. However, since this verification is costly, the principal has an incentive to economise on it.

Typically, prior to the decision, the potential employer receives references or a recommendation letter from previous supervisors, the judge can inspect the outcome of pre-trial 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 information.

We study such situations in the following mechanism design model and establish procedures that maximize the principal's expected value from allocation net of verification costs:

A principal has to take a binary decision for which she relies on the 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 noisy signal about the agent's information. She cannot influence the agent's actions by monetary transfers, but has the opportunity to

[1] Even though a public sector job entails payments, if the payment is fixed, it cannot be used to incentivise truthful reports of the candidate's ability.

reveal his information at a cost.

Given that hiring procedures, the judicial system, as well as consumer protection legislation are established universally and before reviewing information about any individual case, we take the perspective that the principal has full commitment and fixes the mechanism *before* she observes her private information. Relations to the informed principal problem, which arises otherwise, are discussed below.[2]

For the main part of the paper we consider the case where the principal's information comes in form of a signal that is "payoff-irrelevant" in the sense that it is informative about the agent's type but does not alter the principal's preferences at a *given* agent type.[3] At the end of the paper we show when and how our main result extends to the case of "payoff-relevant" information. That is, when the principal's preferences over the two actions at any *given* agent type are affected by her privately observed characteristics (which are correlated with the agent's type).

Our analysis relies on the mechanism design model introduced by Ben-Porath, Dekel, Lipman (2014) (henceforth BDL) which studies optimal verification mechanisms for the case of I agents with independently distributed types. We consider the case of one agent and introduce correlation between the agent's and principal's private information.

An optimal mechanism in the class of Bayesian incentive compatible (BIC) mechanisms is characterised and shown to take a simple cut-off form: If the signal is above a certain cut-off, the agent receives the good without verification, independent of his type report. At lower signals, the agent can receive the good only at reports above a certain threshold and after this report is verified. This mechanism is similar in structure to *favoured agent mechanisms* which BDL show to be optimal in the case of independent types.

[2]It turns out that the proposed cut-off mechanism solves the informed principal problem in the sense that it constitutes an equilibrium of the corresponding principal-agent game where the principal proposes a mechanism after observing her private information and the agent updates his beliefs conditional on the proposed mechanism.

[3]Once his ability is known, the content of a recommendation letter should not alter the willingness to hire a candidate. Similarly, the pre-trial investigation gives information about the level of guilt but should not affect the judge's preferences at a given level of guilt.

The cut-off mechanism is also ex post incentive compatible (EPIC); the agent's beliefs over the signal distribution are irrelevant for implementability. This implies that the principal does not lose by making the realisation of her signal public. In the context of the above examples, this result advocates transparent procedures. It does not pay to keep the applicant in the dark about the references or other information which the potential employer has already evaluated in order to assess his ability level. The fact that pre-trial investigations and the exact charge have to be presented to the defendant before conducting a trial does not constrain the judicial system's efficiency.[4] Note that this does not imply, that the principal does not benefit from the signal. The more accurate belief about the agent type after observing the signal allows more efficient mechanisms but she does not benefit from the fact that the agent does not observe the signal before his report, as would be the case for mechanisms with monetary transfers (**?**).

Even though the presence of correlated information seems natural in most situations, most work within the mechanism design literature assumes that all players' private information is distributed independently. In settings with monetary transfers, this focus is partly due to results[5] which show that even slight correlation allows the principal to extract all surplus and therefore yields ex post efficient allocation rules. With money, the structure of optimal

[4] A pre-trial in which the accused is not informed about the charge he is facing seems to be kafkaesque. Probably because it is precisely the subject matter of Kafka's novel The Trial. As argued in **?** the description of the justice system in the novel is not pure fiction, but based on historic institutions.
Comparing today's Austrian criminal code of procedure (§6 (2) StPO)

> Der Beschuldigte hat das Recht, alle gegen ihn vorliegende Verdachtsgründe zu erfahren[...]

with the code of 1803 (II.3 Von Untersuchung des Beschuldigten und dem Verhöre §331),

> Wenn der in die Untersuchung gezogene Verhörte angibt, kein Ursache zu wissen, warum er vor der Behörde stehe, ist ihm die zur Schuld gelegte Übertretung so weit, und von dem woraus ein rechtlicher Verdacht gegen ihn entspringt, so viel vorzuhalten, als nötig ist, ihn in die Kenntniss der Beschuldigung zu setzen.

While the first quote states that the defendant has the right to learn about all potential charges he faces, the latter gives the court of inquiry much more discretion in the extent of information release to the defendant, stating that he has to be informed as far as necessary to notify him that he is accused.
? argues that this observation is in line with the broader development of continental European criminal procedure code from the medieval inquisitorial proceedings, which exhibited secret charges, to modern forms of criminal law proceedings.

[5] See discussion of the related literature below.

mechanisms in the presence of correlation differs significantly from the case of independent distributions.

We show that this difference does not arise in the context of mechanisms with costly verification. Correlated information complicates the analysis of such mechanisms significantly: With independently distributed types, BDL and **?** show that incentive constraints take a tractable structure. Incentive compatibility for all other types follows by ensuring incentive compatibility for the "worst-off type". With correlation however, this is not the case. The expected value of any given (mis-)report depends on the real type of the agent since his type affects the belief over the signal distribution. This difference in beliefs is precisely what optimal mechanisms with monetary transfers exploit to design payment schemes, in form of lotteries, that allow for full surplus extraction by the principal. Our main result shows that without monetary transfers, the principal cannot exploit these differences. Optimal mechanisms take a similar structure with and without correlation even though the characterisation of incentive compatibility differs.

At the end of the paper we study how our results extend, when the principal's private information is "payoff-relevant" so that it directly affects her preferences. The above-mentioned result carries over to this setting if and only if the direct effect works in the same direction as the information effect. That is, when the principal's private information indicates a high agent type, it also increases the principal's utility of the agent-preferred action at any *given* type. Conversely, if the information effect and direct effect point in opposite directions, the principal strictly profits from keeping the signal private.[6] In this case, the optimal mechanism exploits the fact that different types hold different beliefs and therefore fulfils the incentive constraints in expectation over the signal but not ex post.

The relation to the literature is discussed in the next section. Section 3 sets up the model and presents in detail how the correlation complicates the characterisation of incentive compatibility constraints. Section 4 first characterises optimal EPIC mechanisms for both

[6] If the optimal mechanism is not "trivial" in the sense that verification occurs never or always when the good is allocated. See section on BIC mechanism with preference effect for details

cases, "payoff-irrelevant" and "payoff-relevant" signals. These will serve as a useful benchmark for the following analysis of optimal Bayesian incentive compatible (BIC) mechanisms with payoff-irrelevant signals in section 5. The last section considers BIC mechanisms for the case where the principal's private information has a direct effect on her preferences.

Related Literature:

This work falls in the intersection of two strands within the mechanism design literature which have previously led to very different results.

On one hand, existing results on mechanism design without money have found that optimal mechanisms take a simple structure which makes only limited use of the specific type distribution and therefore, implementability does not hinge on beliefs of the agents. See discussion below, (**?????**) among others. To the best of our knowledge, all these results have been derived under the assumption that the private information of players is independently distributed.

On the other hand, results in mechanism design problems which do allow for monetary transfers, show that correlation can be beneficial to the principal. Different agent types hold different beliefs over the signal distribution which may be exploited by the principal. Reward schemes in form of lotteries can be designed, which specific types value more as a result of their beliefs. Under mild conditions on the information structure the principal can always allocate efficiently and extract all information rents. (For example **?**, **?** and **?**). These strong results give a rationale for the literature's focus on independent distributions.

Given the above contrast, our paper studies optimal mechanisms for situations without transfers and with correlation. In particular, we answer the question whether these mechanisms take a "simple" structure and are robustly implementable as has been shown for other verification problems, or if they present an involved structure that depends crucially on the exact distribution of private information, which is typical for mechanisms with correlation and monetary transfers.

The possibility for the mechanism designer to verify an agent's private information at a cost was first introduced by **?** considering a principal-agent model for debt contracts, which was extended to a two-period model by **?**. These early models of state verification feature both, monetary transfers and verification. **?** introduce a game between a principal 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 learn about on dimension before taking her decision.

The model presented in the present paper is most closely related to **?** 2014 (henceforth BDL) who model costly verification and consider the case of allocating a good among I agents whose types are independently distributed. BDL show that all optimal mechanisms are so-called *favoured agent mechanisms*, which allocate the object to a predetermined agent if no other agent reports a type above his individual threshold. In case there is a type report above the threshold, the highest type is verified and receives the good conditional on having reported the truth. *Favoured agent mechanisms* are robust in the sense that they are ex post incentive compatible (EPIC),[7] i.e. even after information that is otherwise private is revealed to the agent, telling the truth remains optimal. Our model is based on BDL; considering the case of a single agent and introducing correlation between the agent's and the principal's information.

? 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. **?** consider an allocation problem without monetary transfers where 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. **?** consider a delegation problem and specify conditions on the verification cost that ensure optimality of a threshold with escape clause.

Mechanisms with monetary transfers and correlated information have been discussed by

[7]Favoured Agent Mechanisms in fact also ensure that truth-telling is a weakly dominant strategy. However, EPIC is the relevant robustness notion in single player setting we consider.

?, ?, ?, and ? which all establish conditions on the information structure that ensure full surplus extraction by the principal. Since all surplus can be extracted through payment schemes in forms of lotteries over the realisations of characteristics which are correlated with the agent's preferences, revenue maximising mechanisms ensure that the ex post efficient allocation is implemented. ? 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. In our model, full surplus extraction is not possible even when these conditions are fulfilled. The lotteries mentioned above potentially require unbounded transfers in order to extract all surplus. For the case of bounded transfers or limited liability, ? show that the qualitative results, the application of rewards as "bets" on the signal, still apply. However, when no transfers are allowed, the lotteries which ensure implementation in their model are not feasible and it is ex ante unclear if and to what extent the principal can profit from correlation among the types.

The situation described in this paper is also connected to the informed principal problem. cf. ? and ?. Since 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. 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. Correlated information is shown to allow for an efficient solution to the informed principal problem with monetary transfers by ? and ?.

? also show that the equivalence between BIC and EPIC mechanisms established by BDL holds more generally rather than only for optimal mechanisms. This relates to and uses techniques first applied by ? (2013) to show equivalence between BIC and DIC mechanisms in settings with monetary transfers which has also been established by ?. With correlation it is not straightforward in which cases and how these techniques can be adapted. The second part of our paper shows that when the principal's preferences are affected by the correlated information the equivalence between BIC and EPIC for optimal mechanisms does not necessarily carry over. In the cases where the optimal BIC mechanism turns out to be EPIC, the question whether, in the presence of correlation, the equivalence holds for all rather than solely optimal mechanisms remains an interesting question for further analysis.

Model

Even though our model captures binary decision problems more generally, to make the illustration clearer, in the rest of the paper we focus on the jargon related to the hiring example from the introduction.

The principal (she) decides whether to allocate a task to the agent (he). The principal's value of outsourcing the task depends on private information the agent holds. Let random variable T be the agent's type. Realisations of T are denoted by $t \in \mathcal{T}$ which is finite and ordered. The principal is informed about the agent's type through noisy signal S with realisations $s \in \mathcal{S}$, finite and ordered. T and S are jointly distributed according to distribution $(f_{t,s})_{t \in \mathcal{T}, s \in \mathcal{S}} >> 0$.[8] The signal satisfies the *Monotone Likelihood Ratio Property*[9]:

(**MLRP**) For any pair t, t' with $t < t'$: $\frac{f_{t',s}}{f_{t,s}}$ is weakly increasing in s.

This implies that a higher signal is more indicative of a higher type.

The principal derives valuation $v : \mathcal{T} \times \mathcal{S} \to \mathbb{R}$ when hiring an agent of type t at signal s. v represents the *net* value for the principal. Hence, we assume that $v(t, s) < 0$ for some combinations (t, s). Otherwise, the analysis is trivial as the principal always allocates the good without any verification. We assume throughout the whole analysis that higher types are more suited to fulfil the task so that $v(\cdot, s)$ is increasing for all realisation s. For most

[8]The choice of discrete types is due to ease of exposition. All arguments in our proofs can be carried out with continuous type and signal spaces with positive density but require more careful probabilistic argumentation.

[9]For the case of two random variables this is equivalent to requiring that T and S are *affiliated*.

of the paper, S represents a noisy, "payoff-irrelevant" signal without affecting the principal's preferences directly. Hence, we assume that $v(t, \cdot)$ is constant for all t and consequently use notation $v(t)$ for this case. At the end of the paper, we do allow the signal to have a direct impact on the principal's preferences. We also characterise the optimal BIC mechanism and show that results for constant $v(t, \cdot)$ carry over in the case of increasing $v(t, \cdot)$; that is, when a signal that indicates a higher T also increases the principal's willingness to hire the agent for every type. If $v(t, \cdot)$ is decreasing on the other hand, we show that ex post implementation comes at a cost for the principal.

There are no monetary transfers to incentivise the agent. However, the principal can make us of a verification technology. More precisely, she has the possibility to find out the realisation of T after paying cost $c > 0$. Verification is perfect; the principal always learns the exact type.[10]

An agent of type t gets utility $u : \mathcal{T} \to \mathbb{R}_{++}$ from the good. Characterising incentive constraints in the next subsection shows that the intensities of the types' utility do not play an important role. As long as $u(t) > 0$ for all t, the agent simply maximises the probability of receiving the object. The agent's outside option is normalised to zero so that it is optimal to participate in the mechanism for all types.

We consider solutions to this problem in a mechanism design context in which the principal has full commitment over verification and allocation procedures chosen before the signal realisation. The agent does therefore not infer any information about S from the principal's actions. This timing of mechanism and signal is natural for situations that occur repeatedly; like hiring decisions. In our point of view, organisations typically fix a hiring procedure before receiving information about any specific candidate. If the signal were observed before the mechanism is specified, the information conveyed to the agent through the proposal would imply an informed principal problem which we do not consider in this paper. Note

[10]It turns out that with commitment over subsequent allocation decisions, it does not alter the results whether she learns the agent's real type or just learns whether the information he provided is wrong, as long as she is certain of what she learned. **?** consider the extension to the case with imperfect verification.

however, that the EPIC mechanisms we derive solve the informed principal problem since it is without loss for the principal to signal her type with the proposal of the mechanism.

We make use of the revelation principle and an optimality argument for the principal similar to the one presented in BDL (see Appendix A) which allow us to focus on direct revelation mechanisms that require only one cheap-talk message from the agent, his type report, and are characterised solely by verification and allocation probabilities depending on (t, s). As the Principal wants to minimise expected verification costs, optimal direct mechanisms need to satisfy two intuitive properties:[11]

1. 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.

2. Minimal Verification:

   Following (t, s) the agent is only verified if - after his report is verified to be true - he receives the good for sure.

The first property implies that in a direct revelation mechanism a candidate who is revealed to be misrepresenting his information never receives a job offer.[12] Minimal verification on the other hand means that the candidate's information, which will not lead to a job offer, is never verified. The above properties imply that we can focus on direct mechanisms $(\boldsymbol{x}, \boldsymbol{z})$ of the following form: $(\boldsymbol{x}, \boldsymbol{z})$ specifies for every combination of agent report and signal realisation (t, s):

$x_{t,s} \in [0, 1]$, the probability that the agent gets the good <u>without</u> being verified.

$z_{t,s} \in [0, 1]$, the probability that the agent receives the good <u>and is verified</u>.

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

[11] See Apendix A for details.

[12] Note that misreports are an off-equilibrium event in an incentive compatible direct mechanism so that there is no cost to applying maximal punishment.

The Agent's Problem

Agent type t's utility is given by $u(t)$ multiplied by the probability that he receives the object. The specification above implies that an agent of type t who reports truthfully faces the random allocation probability $x_{t,S} + z_{t,S}$.[13] Whether his report is verified or not, is irrelevant. If however, t reports $\hat{t} \neq t$, he expects to receive the good with probability $x_{\hat{t},S}$; only if he is not verified. We focus on Bayesian incentive compatible (BIC) mechanisms. In order to ensure truthful reports by every agent, a direct revelation mechanism needs to satisfy the following incentive constraints:

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

Note that the right hand side of the above expression changes in t not only through the value $u(t)$ but also through the expectation operator due to the fact that different types have different conditional beliefs over the distribution of S. More precisely, unlike in BDL and **?** the expected allocation probability at a certain misreport is not independent of the real type. This complicates the analysis of optimal mechanisms and requires a different approach to prove our results then the arguments used in BDL and **?**. In order to ensure incentive compatibility it is no longer sufficient to identify the type with the lowest expected allocation probability and ensure that this type is weakly worse-off after any misreport.[14] In their proofs of robustness properties of Bayes-optimal mechanisms, both **?** and **?** make use of the fact that the principal can alter any (optimal) potentially non-robust mechanism in the space of ex post allocations, making changes in $x_{t,s}$ for different values of s, while maintaining interim expected allocations and therefore all incentive constraints unchanged for t until the mechanism is incentive compatible point-wise, that is for every s.[15] While we

[13]Note that this probability depends on the agent's belief over random variable S.

[14]See BDL p. 3793 for details: Independence would allow us to define $\chi = \inf_{t \in \mathcal{T}}\{\mathbb{E}_S[x_{t,S} + z_{t,S}]\}$ and write the IC constraints as $\forall \hat{t} \in \mathcal{T}: \quad \chi \geq \mathbb{E}_S[x_{\hat{t},S}]$ where the RHS (the value of lying) does not depend on which real type considers the misreport. The expectation need not be conditional on t.

[15]Instead of a signal S, these models feature several players with independently distributed types. Think of s in our case, having the role of other players' reports t_{-i} in their settings.

also show robustness of the optimal mechanism, note that the underlying reason is different. As will become clear in the second part of the analysis, with correlation this property depends on the specific interplay the effect S has through the information and the preferences of the principal; techniques used in the two papers above cannot be applied as soon as T and S are not independently distributed.

As we assume that every type derives positive utility from the good ($u(t) > 0$), it follows that the intensity of type t's preferences can be eliminated from both sides of the IC constraint: The agent simply maximises his expected allocation probability. The utility he derives from the good is the same whether he reported truthfully or not. Furthermore, both expectation operators condition on the same event, so we can multiply both sides by $\mathbb{P}[\ T = t\ ]$ which allows for cleaner exposition in what follows. Hence, we can simplify the constraints and require for all $t, \hat{t} \in \mathcal{T}$:

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

The Principal's Problem

The principal designs a mechanism that maximises the expected allocative efficiency net of the costs of verification. If the task is assigned without verification, she gains $v(T, S)$. In case of allocation with prior verification, she additionally pays cost c. Hence, the principal's problem can be stated as the following linear programme:

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

Optimal EPIC mechanisms

Numerous results in mechanism design without monetary transfers – and more specifically those considering costly verification – show that it is without loss for the principal to implement mechanisms which are ex post (EPIC) or dominant (DIC) incentive compatible rather than only Bayesian incentive compatible (BIC).[16] This requires that the mechanism satisfy IC constraints for *any* beliefs.[17]

However, to the best of our knowledge, all results in this direction make use of the fact that the information of different players is independently distributed and therefore agent beliefs are equal at all preference types.

On the other extreme, papers on mechanism design with correlated information which allow for monetary transfers show that under certain conditions even the smallest variation of beliefs in preference types can be exploited by the principal to implement ex post efficient allocations and extract all surplus which implies that the structure of optimal mechanisms changes drastically once there is slight correlation.[18]

This paper studies the combination of these two extremes: Can the different beliefs induced by correlated information still be exploited by a principal who uses costly verification and no monetary transfers to incentivise truthful reporting? The answer we find is no. Different beliefs over a noisy signal realisation cannot be exploited. The optimal BIC mechanism presented in the next section coincides with the optimal EPIC mechanism derived in this section; i.e. it satisfies incentive constraints for every realisation s, so that it is without loss for the principal to make her signal public before the agent reports his type.

EPIC requires that after the outcome is realised, no agent type have an incentive to modify his report. In our setting this boils down to requiring that given any realisation s,

[16] See, among others, BDL, **?**, **?**.

[17] In our setting with one agent and a signal the applicable concept is that of EPIC; requiring that the mechanism is IC even if the agent knows the realisation of S.

[18] Full surplus extraction requires unbounded transfers or a high amount of correlation. Results on bounded transfers show however, that the applied technique of offering lotteries to different types which they value differently due to their diverse beliefs can still be applied even if there are bounds or non-negativity constraints on the transfers. **?** is one example.

reporting his true type is a best response. Hence, that $\forall s \in \mathcal{S},\ \forall t, \hat{t} \in \mathcal{T}$:

$$(EPIC(s)_{t,\hat{t}}): \qquad x_{t,s} + z_{t,s} - x_{\hat{t},s} \overset{!}{\geq} 0$$

If IC constraints are required to hold for any s, as the above condition states, the principal's problem can be separated into $|\mathcal{S}|$ independent problems, one for every realisation.[19]

Therefore, this section derives the optimal mechanism for any given s for the general case of $v(t,s)$ and illustrates how this result leads to optimal "global" EPIC mechanisms, over all s.

For given s, the principal's problem can be written as the following linear programme where $\boldsymbol{x}_{\cdot,s}$ denotes the vector $(x_{t,s})_{\{t \in \mathcal{T}\}}$ and similarly for $\boldsymbol{z}_{\cdot,s}$.

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

First note that for any optimal mechanism, the allocation probability $x_{t,s}$ has to be constant in the report, that is $\forall t,\ \hat{t} : x_{t,s} = x_{\hat{t},s}$.

If this were not the case, there would be different reports t and $\hat{t}$ with $x_{\hat{t},s} > x_{t,s}$. Incentive compatibility implies that for all $\tilde{t} \in \mathcal{T}$ $x_{\tilde{t},s} + z_{\tilde{t},s} \geq x_{\hat{t},s} > x_{t,s}$. I.e. there could not be a type with a binding incentive constraint regarding the report t. This in turn implies that, optimally $z_{t,s} = 0$.[20]

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 assumption. Hence, we must have that $\forall t,\ \hat{t} : x_{t,s} = x_{\hat{t},s} \equiv \chi_s$. With constant x, all incentive constraints are clearly fulfilled, so that

[19] In the original problem, the (Bayesian) IC conditions are in terms of weighted sums over different realisations s.

[20] If it were positive, $z_{t,s}$ could be lowered and $x_{t,s}$ increased at least until the strict inequality above binds. This remains the allocation probabilities unchanged but lowers verification costs.

the principal's problem now can be rewritten as:

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

$$\text{s.t.} \quad \forall t \in \mathcal{T} : \chi_s + z_{t,s} \leq 1$$

Since the incentive constraints are fulfilled for any combination of χ and z, it follows that $z_{t,s}$ will be set as high as possible, i.e. to $1 - \chi_s$ if $(v(t,s) - c)$ is positive and equal to 0 otherwise. As χ_s does not depend on t, it can be written outside the sum. Together with the characterisation of the optimal z, this allows us to further simplify the problem towards:

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

Or equivalently (dividing by $\mathbb{P}[\; S = s \;]$) as:

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

Since the problem is linear in χ_s, the optimal value is either 0 or 1, depending on which of the expectations is bigger. This yields the following characterisation of optimal EPIC mechanism for all s:

Lemma 0.1

For given s the optimal mechanism is characterised as follows: For all $t \in \mathcal{T}$:

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

The above result states that at a given signal s the principal allocates to the agent irrespective of his type t if the expected value (conditional on s) of doing so is higher than the expected value of allocating only to high types but bearing the necessary verification costs.

This ex post problem can also be seen as a special case of the model presented by BDL. The optimal mechanism is obtained with s representing a second player with verification costs of 0.

Figure **??** sketches how Lemma 1 leads to the cut-off mechanism for all s in the case of a noisy signal, that is when $v(t,\cdot)$ is constant in s:

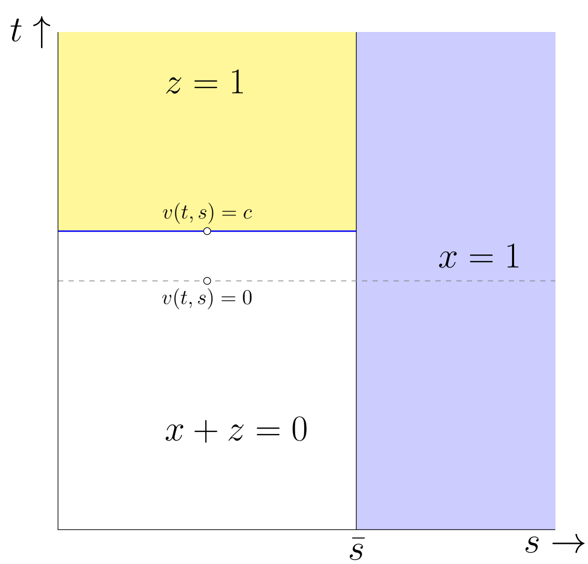
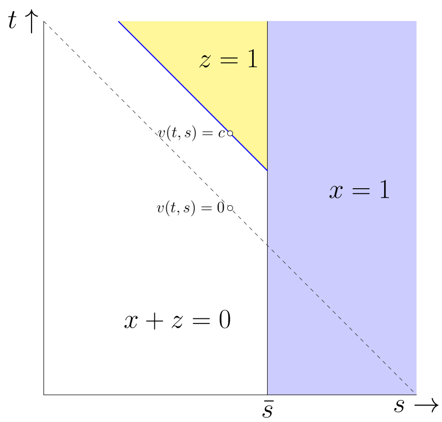


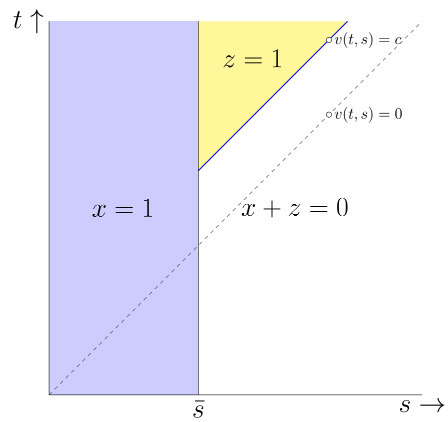
Figure 1: EPIC cut-off mechanisms with noisy signal

If the observed signal is above the cut-off, $\bar{s}$, the principal is optimistic about T and outsources the task 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 task only after his type report is verified ($z = 1$) to be above a threshold so that $v(t, s) - c$ is positive. With a noisy signal, this threshold does not depend on s (see the solid horizontal line in blue). It is easy to see that this mechanism is EPIC. Either the signal is such that the agent gets the task independent of his type (the shaded blue area to the right of $\bar{s}$) or he can only get the task after being verified (the shaded 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 realised.

Figure **??** qualitatively depicts the cut-off mechanism in the case where the principal's private information affects her preferences directly:



(a) Increasing $v(t, \cdot)$



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

Figure 2: EPIC cut-off mechanisms with preference effect

In the case of Figure **??** (a) observing a higher signal is good news, both, in terms of the more optimistic distribution over T, and because it directly increases $v(t, \cdot)$ for all t. Similar to the noisy signal case in Figure **??**, the agent is awarded without verification if the signal is high enough. However, now for signals below $\bar{s}$ the agent's threshold increases when s decreases.

Figure **??** (b) sketches the optimal EPIC mechanism when $v(t, \cdot)$ is decreasing enough. More precisely, the mechanism takes this structure if the (negative) direct effect of a higher signal dominates its information effect in the sense that $\mathbb{E}_T[\ (v(T,s) \mid S = s\ ]$ decreases in s. In this case, the principal outsources without verification for low signals. At high signals, she verifies the agent's type and allocates whenever he reports a type above the threshold which is now increasing in s. The higher the signal, the higher needs to be t so that verification and allocation is worthwhile.

The next section contains the main result of our paper: Unlike in mechanism design problems with monetary transfers, the principal cannot benefit from the types' differences in beliefs over a correlated signal. The cut-off mechanism depicted above is therefore also optimal in the less restricted class of BIC mechanisms. Without loss for the principal, the signal realisation can be made public. The subsequent section then shows how this result extends for the case (a) but not in the case of (b).

Optimal BIC mechanisms with noisy signal

If S has no preference effect so that we can denote the principal's valuation by $v(t)$, the linear programme which is solved by the optimal BIC mechanism can be written as follows:

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

The first main result of our paper shows that the cut-off mechanism presented in the previous sections solves the above problem:

Theorem 1. *If S represents a noisy signal, the following cut-off mechanism solves the principal's problem:*

$$x_{t,s} = \mathbb{1}_{\{s \geq \bar{s}\}}$$

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

with

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

Proof. In order to prove the above result we propose a relaxed version of the principal's problem and show that it is solved by the stated cut-off mechanism. Since the cut-off mechanism is clearly feasible in the original problem, this suffices to show that it is also optimal.

Define the subspace of profitable types as those t with positive allocation value:

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

and the unprofitable type accordingly as $\mathcal{T}^- \equiv \mathcal{T} \backslash \mathcal{T}^+$. Not that if either of the two subspaces is empty, the problem becomes trivial. Either the principal hires the agent at all types or never. She does not need to verify any type report.

The relaxed problem includes only those incentive constraints that prevent types in $\mathcal{T}^-$ to misreport types in $\mathcal{T}^+$. Hence, it reads.

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

We prove a series of lemmata which derive properties of and induce 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. In the proof of these lemmata and throughout the rest of the paper we make repeated use of the following notation: We denote changes in 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 $dz_{t,s}$.

$d(IC_{t,\hat{t}})$ denotes the change in surplus utility type t receives from reporting the truth rather than misreporting $\hat{t}$, which is induced by a change of the above type.[21]

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. The requirement in the following lemmata will be to propose only changes to a solution $(\boldsymbol{x}, \boldsymbol{z})$ of the relaxed probnlem which ensure that $dV \geq 0$ and that $d(IC_{t,\hat{t}}) \geq 0$ for all $t \in \mathcal{T}^-$ and all $\hat{t} \in \mathcal{T}^+$.

Lemma 1.1

The optimal mechanism in the relaxed problem features:

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

Proof. This follows from the relaxation, since the incentive constraints to misreport a type within $\mathcal{T}^-$ are ignored. Any positive $z_{t,s}$ at $t \in \mathcal{T}^-$ could therefore be set to 0 while $x_{t,s}$ is increased by the same amount. The allocation as well as the unprofitable types' incentive to misreport remain unchanged and the expected verification costs decrease. Formally, Suppose for some $t \in \mathcal{T}^-$ and $s \in \mathcal{S}$ $z_{t,s} > 0$. Apply following shift:

$$dx_{t,s} + dz_{t,s} = 0, \; dz_{ts} < 0$$

[21]Recall that the constraint $(IC_{t,\hat{t}})$ reads $\sum_s f_{t,s} \left[x_{t,s} + z_{t,s} - x_{\hat{t},s} \right] \geq 0$ so that $d(IC_{t,\hat{t}})$ denotes the change to the left hand side of this inequality

$\square$

Lemma 1.2

Without loss for the principal, we can assume that the optimal mechanism in the relaxed problem takes the following cut-off form for $\boldsymbol{x}_{\hat{t},\cdot}$:

For all $\hat{t} \in \mathcal{T}^+$:

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

Proof. Suppose, contrary to the structure in the lemma, that for a feasible, IC mechanism of the relaxed problem it holds true 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 $dx_{\hat{t},s'} > 0$. We choose these shifts in a proportion, such that for the highest unprofitable type, $\tilde{t} = \max\{t | t \in \mathcal{T}^-\}$, his incentives to misreport $\hat{t}$ remains unchanged:

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

By the monotone likelihood ratio property, allocating with higher probability at higher signals ($dx_{\hat{t},s'} > 0$) with the above proportion will make the report $\hat{t}$ weakly less attractive for all types $t < \tilde{t}$. (And more attractive for higher types whose incentive constraints are however not included in the relaxed problem.) This can be seen as follows since the term in squared brackets is positive:

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

The principal's value changes in the following way:

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

here the last inequality follows from the facts that $dx_{\hat{t},s'} > 0$ and $v(\hat{t}) \geq 0$ and $\hat{t} \in \mathcal{T}^+$.

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 $z_{\hat{t},s}$ in order to maintain $x_{\hat{t},s'} + z_{\hat{t},s'}$ and $x_{\hat{t},s} + z_{\hat{t},s}$ constant:

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

Note that this implies $dz_{\hat{t},s'} < 0$ and $dz_{\hat{t},s} > 0$, which is feasible.[22]

The above changes in x imply for z:

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

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

Further, the principal's value now changes by:

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

[22] As we assume $x_{\hat{t},s'} < 1$, $x_{\hat{t},s'} + z_{\hat{t},s'} = 1$ implies that $z_{\hat{t},s'} > 0$. Since $x_{\hat{t},s} > 0$ we further must have $z_{\hat{t},s} < 1$ for feasibility.

The second equality follows since the total allocation $(\boldsymbol{x}+\boldsymbol{z})$ remains the same with these shifts, so that only the verification cost changes. The change in verification is such that the perceived verification probability of a lower type remains unchanged while shifting verification probability to lower signals $(-dz_{\hat{t},s'} \geq 0)$. Hence, for the true type $\hat{t}$, and therefore from the principal's perspective, the expected verification cost decreases weakly (MLRP implies that the term in squared brackets is negative). $\square$

Lemma 1.3

Without loss for the principal we can assume that the optimal mechanism of the relaxed problem features $x_{\hat{t},s} = x_{\hat{t}',s}$ for all $\hat{t}, \hat{t}' \in \mathcal{T}^+$ and for all $s \in \mathcal{S}$.

This implies that the allocation probability $\boldsymbol{x}$ is constant in the report for all reports within the set of profitable types and is implied by the following argument:

Proof. This follows from the fact that, by the previous lemma, there is a cut-off signal for every report such that above the cut-off the allocation probability is 1, for lower signals it is 0, and in $[0,1)$ at the cut-off.

Since $v(\hat{t}) \geq 0$ for all profitable types the principal has an incentive to increase $x_{\hat{t},s}$ up to the maximum level without violating incentive constraints.

Given the cut-off form prescribed in the previous lemma, the incentive constraints from the relaxed problem read:

$$(IC_{t,\hat{t}}): \quad \sum_{s \in \mathcal{S}} f_{t,s}\left[x_{t,s} + z_{t,s}\right] - f_{t,\tilde{s}(\hat{t})} x_{\hat{t},\tilde{s}(\hat{t})} - \sum_{s > \tilde{s}(\hat{t})} f_{t,s} \cdot 1 \geq 0$$

The incentive to misreport is decreasing in $\tilde{s}(\hat{t})$ and increasing in $x_{\hat{t},\tilde{s}(\hat{t})}$ lexicographically while the principal's value changes in the opposite direction. Therefore, she optimally chooses the same $\boldsymbol{x}$ for all profitable types; namely, setting $\tilde{s}$ as low as possible and $x_{\hat{t},\tilde{s}}$ as high as possible without violating the IC constraints. $\square$

Lemma 1.4

It is without loss in the relaxed problem to choose $\boldsymbol{x}_{\cdot,s}$ constant in the report for all reports $t \in \mathcal{T}$ at all signals s. This implies that the incentive constraints in the relaxed problem can be fulfilled ex post, that is for every realisation s, rather than in expectation only.

Proof. Inversely to the above argument, the principal wants to choose the allocation probability of the unprofitable types as low as possible. This implies that all incentive constraints of the relaxed problem bind and without loss the principal can choose to allocate all probability mass for these types at high signals as well.[23] □

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 reduces to choose whether to allocate without verification at all reports or to allocate after verification only if the report is such that $v(t) - c \geq 0$. This concludes the proof of the theorem, showing that the cut-off mechanism solves the relaxed problem and is therefore optimal in the original problem. □

BIC mechanisms with preference effect

For this section the principal's private information may also have an effect on the preferences. We distinguish the two cases where this direct effect is positive everywhere (a) or negative everywhere (b).

Case (a): Increasing $v(t, \cdot)$

Suppose that the principal's private information is about already hired staff so that S represents their performance which cannot be observed by the new potential employee at the time of the hiring process. T and S are positively correlated, e.g. because the already employed

[23]As $v(t)$ is not affected by s providing a type $t \in \mathcal{T}^-$ with the same expected allocation probability gives the principal the same expected allocation value independent of to which signal realisation she assigns the allocation probability mass.

and the candidate share some educational background or their performance depends on the complexity of the tasks they are to carry out.

If, additionally, there are complementaries between the ability levels of existing and new staff, S may also have a direct effect. That is, for higher performance or ability levels of the existing staff members, it is more worthwhile to employ the new candidate at any *given* type t.

This case is captured with $v(t,\cdot)$ increasing for all t and this subsection shows that the optimal BIC mechanism also satisfies the EPIC constraints. That is, the results from the last section carry over.

The problem case reads:

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

And the corresponding result:

Theorem 2. *The optimal EPIC cut-off mechanism solves the principal's problem. It is characterised as follows:*

$$x_{t,s} = \mathbb{1}_{\{s\,\geq\,\bar{s}\}}$$

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

with

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

This is the characterisation of a mechanism as sketched in Figure **??** (a). Note that the above cut-off is well defined since, $s\mapsto\mathbb{E}_T[v(T,s)-(v(T,s)-c)^+|\ S=s\ ]$ is non decreasing in s.

Proof. In order to prove the theorem we again propose a relaxed version of the linear programme and establosh a series of lemmata to derive the above mentioned mechanism. However, the distinction between "profitable" and "unprofitable" types as well as the exact choice of incentive constraints to be ignored is now more complex since for given types, the value and sign of $v(t,s)$ changes in s. Similarly, the last step, arguing that, without loss, any optimal mechanism of the relaxed problem can be required to satisfy the constraints of the original problem ex post, requires some more detail.

The proofs of the different steps are similar to the previous theorem's proof and are therefore relegated to Appendix B.

We again define the subspace $\mathcal{T}^+$ but now the relaxed problem includes those incentive constraints that ensure 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. I.e. this relaxed problem ignores all downward incentive constraints and all constraints preventing misreports towards types within $\mathcal{T}^- \equiv \mathcal{T} \setminus \mathcal{T}^+$.

$$\mathcal{T}^+ = \{t \in \mathcal{T} | v(t, \bar{s}) \geq 0\}$$

where $\bar{s}$ is defined in the theorem. The relaxed problem:

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

Lemma 2.1

The optimal mechanism in the relaxed problem features:

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

The proof is identical as in the relaxed problem of the previous section and therefore omitted.

Lemma 2.2

Without loss for the principal, we can assume that the optimal mechanism in the relaxed form takes the following cut-off form for $\boldsymbol{x}_{t,\cdot}$:

For all $t \in \mathcal{T}$:

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

The proof can be found in Appendix B. The arguments are similar to the ones used in the proof of Lemma **??**.

Lemma 2.3

Without loss in the relaxed problem, the optimal mechanism also takes a cut-off form for $\boldsymbol{x}_{t,\cdot} + \boldsymbol{z}_{t,\cdot}$:

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

The proof is relegated to the appendix. As z is only relevant for the true type's value it can be freely allocated among the signals as long as the perceived probability for the true types and therefore for the principal remains unchanged.

Lemma 2.4

Without loss for the principal, the relaxed problem can be restricted, requiring the IC constraints to hold point-wise. That is $\forall t \in \mathcal{T}$, $\forall \hat{t} \in \mathcal{T}^{+}$ with $t < \hat{t}$ and $\forall s \in \mathcal{S}$:

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

The proof can be found in the appendix. Note that the cut-off in allocation probabilities implies that - in order to fulfil any given incentive constraint (in expectation) the cut-off for the true type has to lie further to the left than for the considered misreport; hence allocating to the true type at any signal at which he could receive the good with a different report.

The last lemma ensures that the incentive constraints of the relaxed problem, which ensure that no type from $\mathcal{T}$ misreports a higher type from $\mathcal{T}^+$, can be fulfilled point-wise. To conclude the proof of the theorem, it remains to show that any optimal mechanism in the class derived above satisfies the IC constraints for all types ex post: In the case of the previous theorem this follows immediately from the fact that with constant $v(t, \cdot)$

Lemma 2.5

An optimal mechanism of the above cut-off structure, satisfying $(EPIC(s)_{t,\hat{t}})$ for all $t \in \mathcal{T}$ and all $\hat{t} \in \mathcal{T}^+$ with $\hat{t} > t$, also satisfies $(EPIC(s)_{t,\hat{t}})$ for all $t, \hat{t} \in \mathcal{T}$.

The proof in the appendix rules out two possible violations of the original (ex post) incentive constraints that can arise in a solution to the relaxed problem. First, no type profits from a higher report in the set $\mathcal{T}^-$. Second, no type profits from reporting a lower value.

This concludes the proof of the second theorem as the optimality of the cut-off mechanism in the class of EPIC mechanisms is shown above. $\square$

Case (b): Decreasing $v(t, \cdot)$

Assume now that the new candidate should be hired to join the present staff members in order to fulfil the very same tasks and that their ability levels substitute each other. Whenever the existing staff is very productive (S has a high realisation), there is less need for a new staff member for any given value of t.[24] This situation is captured by decreasing $v(t, \cdot)$.

We show that in this case the principal can achieve the same total allocation $(\boldsymbol{x} + \boldsymbol{z})$ as in the optimal EPIC mechanism but at a lower verification cost, by exploiting the fact

[24] While high s is still indicative that T also realised on higher values. (Positive information effect)

that different agent types have different beliefs over the distribution of S. I.e. the principal profits strictly from keeping the signal realisation private; fulfilling the IC constraints in expectation but not ex post.

Assume that $v(t,\cdot)$ is decreasing enough so that the (negative) direct effect of a higher signal dominates its (positive) information effect in the sense that $\mathbb{E}_T[\ (v(T,s)\mid S=s\ ]$ decreases in s. The optimal EPIC mechanism for this case was depicted Figure **??** (b).

The next result shows that such a mechanism is not optimal within the class of BIC mechanisms. To gain intuition for why this is the case, figure **??** depicts a potential improvement of the EPIC mechanism of figure **??** (b).

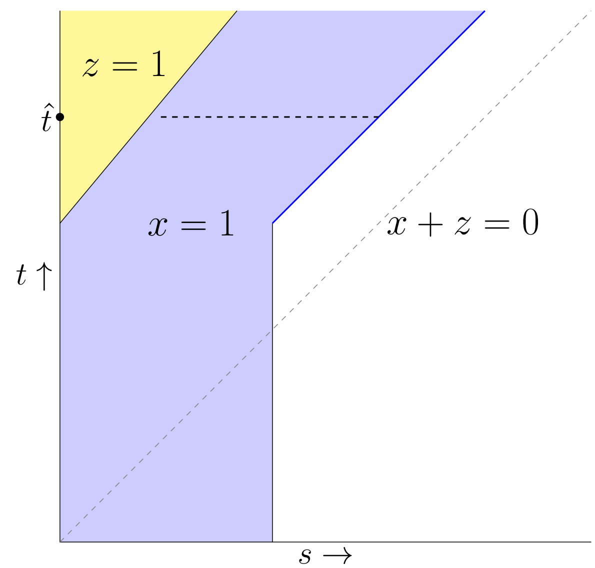


Figure 3: After verification Shift

While the overall allocation remains unchanged in both mechanisms, the second mechanism saves verification costs. In both mechanisms higher types have a higher hiring probability so that the relevant incentive constraints will be those deterring lower types t to report towards higher $\hat{t}$. A lower type t, when considering a misreport towards $\hat{t}$, can be hired only when the report is not verified. That is, when the signal falls within the blue (or darker shaded) area in the figure. Given that t is lower, his belief places relatively more weight on low signal realisations so that in Figure **??** where the blue (darker) area is placed at higher signals can be chosen larger as in Figure **??** (b) and still lead to the same perceived utility from misreporting $\hat{t}$. In other words, the verification area in yellow (or lighter shaded) can be chosen smaller and the principal saves verification costs with the latter mechanism.

To see that this mechanism is not EPIC, consider a low type who observes that the signal is relatively high so that he would not be hired after revealing his true type. By misreporting he can secure himself allocation probability of 1.

The following result formalises this idea:

Proposition 3

If $v(t, s)$ is decreasing in s and the optimal EPIC 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 strictly higher value which is BIC. Hence, the principal profits strictly from S being private.

Note that under the assumption that $\mathbb{E}_T[\ (v(T, s) \mid S = s\ ]$ is decreasing in s, the optimal EPIC mechanism has the properties stated in the theorem whenever it is non-trivial in the sense that $\boldsymbol{x}$ and $\boldsymbol{z}$ are positive at some combinations (t, s).

Proof. To prove the claim, we construct an improvement which will not violate the Bayesian incentive constraints. This suffices to show that the principal strictly profits from maintaining the realisation of S private since 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 $x_{\hat{t},s}$:

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

In order 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 minimised at $t' = \max\{t \in \mathcal{T} | t < \hat{t}\}$.[25] Setting $dx_{\hat{t},s'} = \frac{f_{t',s}}{f_{t',s'}}\left(-dx_{\hat{t},s}\right)$ ensures that the incentives to misreport towards $\hat{t}$ 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 since 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 of magnitude small enough to not violate these constraints.[26] $\square$

[25]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.

[26]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.

Concluding Remarks

This paper studies the role of correlation in a mechanism design model, where the principal may use costly verification instead of monetary transfers to incentivise the revelation of private information. The optimal mechanism is characterised and shown to be similar in structure to the case of independent information which contrasts with results on correlation in mechanism design problems with money. Also analogue to independence, the mechanism is ex post incentive compatible which implies that it is without loss for the principal to make the signal realisation public before contracting with the agent. In an extension where the principal's private information also affects her preferences we characterise the mechanism and show that the above qualities remain in the case where the information and direct effect work in the same direction. In the opposite case we show how the principal can benefit by maintaining the signal private. Interesting directions for future analysis include the optimal BIC mechanism for the case with negative direct effect, the implementation and commitment requirements for the optimal mechanisms, and the question whether, with correlation, the equivalence between BIC and EPIC mechanisms holds more generally than for optimal mechanisms.

Appendix A

Revelation Principle

The revelation principle presented here is close to the revelation principle in BDL, but takes into account possible issues arising from the correlation between the signal and the type realisation.
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 realisation s, the principal revises 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 realisation 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:[27]

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

- The principal verifies the agents 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 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 verifiaction 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 on, assuming that the game had reached this stage and it was found out 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 behaviour after this knot and the verification. The strategy s_A contains a plan for the behaviour of the agent from this stage on. The principal simulates his

[27] 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 the signal realized to s.

own behaviour as prescribed in the original mechanism.

Note that given any signal realisation 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 realisation s):

The agent of type t imitates type $\hat{t}$'s behaviour $s(\hat{t})$ until the first verification and then sticks to the behaviour 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 utility of the agent and principal, are the same in both mechanisms.

But truth-telling is optimal for the agent in the constructed mechanism, since 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 to simplify the class of possible optimal mechanisms. In short, in any optimal mechanism the principal will chose the punishment and the reward for detected truth telling as high as possible.

1. Maximal punishment: $t \notin \{\hat{t}, \emptyset\} \Rightarrow a(\hat{t}, t, s) = 0$

   Since the mechanism is direct, in equilibrium the agent will not lie, and 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 does also only increase 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 $de(\hat{t}, s) < 0$ whilst 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. $a(s, \hat{t}, \emptyset) > 0$. But in this case this allocation could be lowered $a(s, \hat{t}, \emptyset) < 0$ such that $d((1 - e(\hat{t}, s))a(s, \hat{t}, \emptyset)) = 0$, such that the incentives to misreport and the overall allocation probability would stay constant. As these procedure would save verification costs, whilst keeping all unconditional allocation probabilities constant, it can be excluded that an optimal mechanism features non maximal reward.

Thess observations fix the allocation after verification. So effectively the mechanism designer has to chose only the verification probability $e(t, s)$ and the allocation probability conditional on no verification $a(s, t, \emptyset)$. For convenience define $z_{t,s} = et, s)$, 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 mechanism 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. [28]

[28] 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.

Appendix B

Proofs not included in the paper

Lemma 2.2 (cut-off in x)

Proof. The proof works similar as the one of Lemma **??**. Suppose that for some $t \in \mathcal{T}$ $\exists s < s' \in \mathcal{S}$ such that $x_{t,s} > 0$ and $x_{t,s'} < 1$.

Modify the mechanism at two points, shifting allocation probability from $x_{t,s}$ to $x_{t,s'}$.

For this lemma, since the principal's allocation value is affected by the signal, we have to choose a different proportion for these shifts in order to ensure that the IC constraints are not violated and at the same time the principal's value does not decrease. The probabilities are changed in a way such that for the true type, t, his expected allocation probability and therefore incentives to misreport remain constant. As before, MLRP ensures that allocating with higher probability at higher signals will make the report t weakly less attractive for lower types.

Recall that t's expected allocation probability when reporting truthfully is given by $\sum_s f_{t,s}\left[x_{t,s} + z_{t,s}\right]$.

This implies, that - in order to keep his incentives constant t' - the proposed shift is chosen to satisfy:

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

For all lower types $\tilde{t}$, the surplus from truth-telling rather than misreporting t is given by $\sum_s f_{\tilde{t},s}\left[x_{\tilde{t},s} + z_{\tilde{t},s} - x_{t,s}\right]$.

To ensure that IC is not violated, the proposed shift has to increase this value for all $\tilde{t} < t$. This holds since the term in squared brackets below is positive for $\tilde{t} < t$ and $s < s'$:

$$\forall \tilde{t} < t:\ d(IC_{\tilde{t},t}) = -f_{\tilde{t},s}\ dx_{t,s} - f_{\tilde{t},s'}\ dx_{t,s'} = f_{\tilde{t},s}\left[\frac{f_{t,s'}}{f_{t,s}} - \frac{f_{\tilde{t},s'}}{f_{\tilde{t},s}}\right] dx_{t,s'} \geq 0$$

The value for the principal from the original mechanism is given by

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

Given the proposed shifts, this changes in the following way:

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

which follows from the facts that $dx_{t,s'} > 0$ and $v(t,s') > v(t,s)$.

The proposed shift is clearly feasible if in the original mechanism $x_{t,s'} + z_{t,s'} < 1$. In the case that $x_{t,s'} + z_{t,s'} = 1$, we can proceed as in the proof of **??**, shifting in addition mass from $z_{t,s'}$ to $z_{t,s}$ in order to maintain $x_{t,s'} + z_{t,s'}$ and $x_{t,s} + z_{t,s}$ constant. □

Lemma 2.3 (cut-off in x+z)

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

Suppose for $t \in \mathcal{T}^+$ that there exist $s < s' \in \mathcal{S}$ with $z_{t,s} > 0$ and $x_{t,s'} + z_{t,s'} < 1$. To rule out this possibility consider a shift of 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.

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

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 favourable since it keeps the verification probability and therefore the costs constant, whilst shifting allocation mass from (t, s) to the more favourable type-signal pair (t, s'), i.e.

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

□

Lemma 2.4 (Ex post)

Proof. By the above lemma, the (Bayesian) IC constraints in the relaxed problem can be written as:[29]

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

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

This condition clearly requires that $\underline{s}(t) \leq \tilde{s}(\hat{t})$ and, in case of equality, that furthermore $x_{t,\underline{s}(t)} + z_{t,\underline{s}(t)} \geq x_{\hat{t},\tilde{s}(t)}$. Since 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 $(EPIC(s)_{t,\hat{t}})$. □

Lemma 2.5 (From relaxed to original)

Proof. We rule out two possible violations of the original (ex post) incentive constraints that can arise in a solution to the relaxed problem. First, no type benefits from a higher report in the set $\mathcal{T}^-$. Second, no type from the entire $\mathcal{T}$ benefits from reporting a lower value.

Assume first that for some $s \in \mathcal{S}$ there are types $\tilde{\tilde{t}} < \tilde{t} \in \mathcal{T}^-$ such that the constraint $(EPIC(s)_{\tilde{\tilde{t}},\tilde{t}})$ is violated. Since $z_{\tilde{\tilde{t}},s} = 0$ for the unprofitable type, this implies $x_{\tilde{\tilde{t}},s} < x_{\tilde{t},s}$. Define $s' \equiv \min\{s \in \mathcal{S} | \exists \tilde{\tilde{t}} < \tilde{t} \in \mathcal{T}^- : x_{\tilde{\tilde{t}},s} < x_{\tilde{t},s}\}$ the lowest signal for which some type $\tilde{\tilde{t}}$ profits from higher report $\tilde{t} \in \mathcal{T}^-$. As t' is from $\mathcal{T}^-$, $x_{t',s'} > x_{\tilde{\tilde{t}},s}$ can be optimal in the relaxed problem only if $v(t', s') \geq 0$, since t''s constraints in the relaxed problem all slack under the assumption. This implies that $s' > \bar{s}$.[30] As $\tilde{t}$ is from $\mathcal{T}^-$ and $z_{\tilde{t},s} = 0$, incentive compatibility for reports within $\mathcal{T}^+$ and the cut off structure derived above imply that $x_{t,s} = 0$ for all t and for all $s < s'$. In particular $x_{t,\bar{s}} = 0$ for all t which by definition of $\bar{s}$ cannot be optimal in the relaxed problem since $\mathbb{E}_T[\ (v(T, \bar{s})\ |\ S = \bar{s}\ ] > \mathbb{E}_T[\ (v(T, \bar{s}) - c\ )^+\ |\ S = \bar{s}\ ] \geq 0$.

Second, 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 $\mathbb{E}_T[\ (v(T, s)\mathbb{1}_{\{T\leq t'\}}\ |\ S = s\ ] > 0$. Otherwise the principal would be better off by lowering x for all types below t''.[31] Monotonicity of the value in the

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

[30]Note that $\mathcal{T}^-$ is precisely defined as the set of types with $v(t, \bar{s}) < 0$.

[31]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''.

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[32]. $\square$

[32] Either by just increasing $x_{t,s}$ if $x_{t,s} + z_{t,s} < 0$ or by lowering $z_{t,s}$ at the same time to save verification costs.