Costly Verification and Money Burning

Rohit Patel[*] Can Urgun[†]

March 2025[‡]

Abstract

We consider the problem of a principal aiming to allocate an indivisible, productivity-enhancing resource—e.g., computing equipment, a grant, etc.—to one of many agents. The principal can utilize two instruments previously studied only in isolation but often used together in practice: costly verification and money burning. We identify the optimal allocation protocol, which takes one of two forms. When full separation of types is suboptimal, money burning is not employed, but there is significant pooling. If full separation is optimal, then both instruments are used. Lower types are only subject to money burning. Higher types go through verification and burn money. Notably, money burning is not substituted away.

JEL Code: C72, D73, D82

Keywords: costly verification, money burning, allocation problem

[*]Meta, Data Science, rohit@rohit-patel.com

[†]Department of Economics, Princeton University, curgun@princeton.edu

[‡]We would also like to thank various seminar and conference audiences, Rakesh Vohra, Eddie Dekel, Niko Matouschek, James Schummer, Alvaro Sandroni, Michael Powell, Navin Kartik, Alessandro Lizzeri, Slyvain Chassang, Hector Chade, Andrew Ferdowsian, Ayca Kaya, Doruk Cetemen, Christopher Sandmann, Mark Whitmeyer, Pietro Ortoleva, Vasiliki Skreta, Caroline Thomas, Axel Niermeyer, Jan Knoepfle, Ludvig Sinander, Leeat Yariv and Kemal Yildiz for helpful comments and suggestions.

1 Introduction

Efficient allocation of resources is a crucial component of any organization (Maritan and Lee (2017)). We consider allocating a scarce resource—computing equipment, personnel, a research grant, funding, etc.—to one of many agents. Importantly, the agents are heterogeneous: they vary in the productivity increase that access to this resource allows, which they know privately. In the presence of transfers, the principal's problem is trivial: she can simply auction the resource to guarantee an efficient allocation. Nonetheless, in many settings, monetary transfers are not feasible either due to legal, ethical, or organizational constraints. A manager cannot auction a new piece of office equipment, researchers cannot bribe funding agencies for grants, etc. We therefore consider the case in which the principal cannot use transfers.

Suppose agents report their value for the resource to the principal. If the principal's allocation is responsive to these reports—ideally corresponding to an efficient allocation—agents would clearly have an incentive to misreport, and embellish their assessed returns from the resource. One approach to this problem relies on costly verification of agents' reports by the principal (see Ben-Porath et al. (2014) and literature that follows). The principal could simply check, possibly probabilistically, the reports of some of the agents. Another approach is to impose bureaucratic hurdles for agents applying for the resource. This is commonly referred to as money-burning activities: costly and meaningless activities for the agents that destroy surplus (see Hartline and Roughgarden (2008) and the literature that follows). It might seem like only one of these instruments is necessary, as both are costly to the principal and serve the same purpose. Indeed, the literature has analyzed them in isolation. For instance, capital financing within large firms has been studied through the lens of costly verification by Harris et al. (1982) and through the lens of money burning by Ross (1986), albeit separately. Nonetheless, in practice, these instruments are often employed in conjunction. In this paper, we analyze the optimal usage of the two instruments in tandem.

We show that money burning and costly verification are complements. When used in conjunction, the two instruments always achieve allocative efficiency, i.e., all types are differentiated, and the agent with the highest productivity receives the resource (breaking ties uniformly). In contrast, costly verification in isolation always results in significant pooling: any two types that are below the cost of verification have the same allocation probability regardless of the difference in the associated productivities. Money burning in isolation, on the other hand, can achieve allocative efficiency albeit with a very large loss of surplus. Indeed, every potential increase in productivity requires a different and higher level of money burning. When the two instruments are both available and used in tandem, for lower types we see that only money burning is used. The required money burning increases in type until the productivity reaches the cost of verification. All the types whose productive is equal to or larger than the cost of verification burn the same amount of money. For those higher types, the fixed amount of money burning is complemented by increasing the verification probability, thus avoiding excessive loss of surplus.

Our results have implications for several applications. For example, a diversified firm might require mid-level managers to go through multiple bureaucratic hoops, resulting in wasted manager-time and hence company resources before assigning them new resources. Our findings suggest that these bureaucratic requirements will increase as the managers' claims increase, but only up to a point. After that, the firm starts to utilize external audits, but even when a manager's claim is subject to an audit, they still need to fulfill the bureaucratic requirements. Similarly, a researcher applying for a grant might have to go through a time-consuming application process, diverting their focus from research even if the funding agency solicits outside expert evaluations before granting funds. Likewise, different municipalities might engage in lobbying activities for some government facility –a clinic, school, bridge, etc.– to be built in their jurisdiction, "wasting" municipal funds instead of allocating them for the welfare of the local populace. Despite lobbying, the government might still use federal funds to conduct a costly population survey before deciding where to build such a facility.

In our setup, there are D agents, each of which has a productivity increase drawn from gaining access to a singular resource. The productivity increase is drawn independently and identically from a finite set of values. Both the agents and the principal are risk neutral. The principal has two instruments at her disposal: *costly verification* and *money burning*. Verification has a fixed cost per agent, borne exclusively by the principal, and reveals productivity perfectly (similar to Ben-Porath et al. (2014)). Money burning has no informational content beside signaling and is costly both to the principal and the agent who is burning money (similar to Hartline and Roughgarden (2008)). Each agent cares about the productivity increase he will get from the resource, net of his losses incurred through the required money burning. All agents have an outside option that provides a normalized value of 0 throughout. The principal cares about the productivity increase, net of the money burning costs and verification costs. A mechanism then corresponds to an extensive form game that the principal designs and commits to, with Bayes Nash Equilibrium as the solution concept.

It is worth noting that the set of direct mechanisms is very large, corresponding to arbitrarily large extensive form games. Therefore analogous to the analysis of Ben-Porath et al. (2014) it is worthwhile to focus on *optimal direct mechanisms*. We show that optimal direct mechanisms correspond to a straightforward procedure. First, all agents report their productivities to the principal, then the principal chooses the agent with the highest reported type (breaking ties uniformly) as a "candidate". This candidate then goes through the necessary money burning dictated by the vector of reports. Once the money burning is completed, the candidate might still be subject to verification, depending on the reported types. Finally, the resource is allocated to the candidate if there was no verification, or if the candidate was found truthful. A candidate that does not do the necessary money burning or found to be lying is punished by withholding of the object. Notably, as discussed in our examples, bureaucratic requirements precede external audits in *optimal direct mechanisms*.

In much of standard mechanism design, finding an optimal direct mechanism boils down to the optimal selection of allocation probabilities. In particular, an educated guess of the binding incentive constraints is enough to deduce the optimal usage of the instrument in question.[1] The inclusion of multiple instruments requires new techniques. To see why, consider, for instance, money burning in isolation. Suppose an incentive constraint is binding for an agent: she is indifferent between reporting her "type" to be t or t'. If the corresponding interim allocation probabilities for types t and t' are known, the difference between money burning for the two types can be readily inferred: it is simply the difference in expected payoffs due to the allocation for the agent reporting t and t'. Thus, in such settings, deducing the set of binding incentive constraints gets us a long way toward a full characterization. In contrast, with two instruments, this is not the case. A binding constraint for an agent could be generated via different combinations of the two instruments. Therefore, optimization of the instruments becomes necessary even both allocation probabilities and binding incentive constraints are known. Moreover, when dealing with two instruments, restricting attention to a particular set of constraints ex ante is also not feasible. In fact, even when the allocation probabilities are fixed, different usages of instruments result in different sets of binding incentive constraints.

The main technical contribution of the paper is to show that the three components of an optimal mechanism: the set of binding constraints, the optimal usage of instruments, and the optimal allocation probabilities, can be optimized in a "modular" fashion. In essence, an optimal mechanism simultaneously chooses both the interim allocation probabilities, as well as the usage of instruments. Clearly, in an optimal direct mechanism that is truthful, no agent will shirk from the required money burning on path. Thus, we can restrict attention to the objective and the constraints at the interim stage, where the agents know their own types and form expectations of the outcome. Due to risk neutrality, the principal's objective and the interim expected utilities of agents are linear. Our approach relies on combining linear programming and *polymatroid* optimization. Border (1991) shows that

[1] This could be ex-ante guessing that only local constraints (type t pretending to be $t+1$ or $t-1$) would be binding, or there would be cutoff type who has binding constraints.

feasible interim allocation probabilities come from a polytope called a *polymatroid*. First, suppose the usage of instruments is fixed by an outside party. Then the decision maker still needs to choose interim allocation probability for each potential productivity level while respecting the incentive constraints. However, which constraints end up binding depends on the allocation probabilities. This optimization of allocation probabilities is the *polymatroid optimization problem*, subject to incentive constraints. On the other hand, now suppose the interim allocation probability of each productivity level is fixed. Then the decision maker still needs to use the instruments optimally, while respecting incentive constraints. In this scenario the set of binding constraints depend on the usage of instruments. This optimization of instruments is the *linear optimization problem*, since the expected interim utilities are linear. In principle, these two problems have to be solved together, but we show that once some basic properties are established, then they can be solved separately without knowing the solution of the other.

The polymatroid structure enables us to identify feasible perturbations to allocations. Since the instruments are also choice variables, we can make perturbations to allocations and instruments as a whole, which in turn alters the set of binding constraints.[2] The linearity of the objective and the constraints allows us to easily assess the resulting gain or loss in the objective from any such perturbation. Thus, we can rule some constraints to be never binding in an optimal mechanism to attain some monotonicity properties: i) utilities are increasing; ii) only upward constraints can bind; and, iii) there is an endogenous pooling cutoff. The types below this endogenous cutoff are treated identically and hence are pooled. They are not subject to any instrument, but they have the lowest allocation probability. The setting also has an exogenous cutoff, which is the smallest type that has a productivity higher than the cost of verification. The characteristics of the optimal usage of instruments are completely determined by the ordering of these two cutoffs.

We use standard linear programming techniques to figure out the usage of instruments for *any* mechanism that satisfies the basic monotonicity properties mentioned above.[3] We show that if the endogenous cutoff is above the exogenous one, then money burning is completely absent. The mechanism only uses costly verification and, in fact, the mechanism corresponds to the mechanism studied in Vohra (2011) and Ben-Porath et al. (2014) with symmetric agents. On the other hand, if the endogenous cutoff is below the exogenous cutoff, then the instruments are used in a complementary manner. Specifically, the mechanism involves only money burning for the types below the exogenous cutoff. Among these lower types (below the exogenous cutoff), the level of money burning is increasing.

[2]This is in stark contrast to the case where the instruments are inferred from binding incentive constraints.

[3]Notably, these characteristics remain optimal for *any* allocation probability, with any collection of binding constraints that satisfies the basic monotonicity properties.

For these lower types verification is simply inefficient; thus, money burning is the only option to differentiate them. For the higher types, verification is introduced along with money burning. Specifically, all higher types have the same amount of money burning as the exogenous cutoff type, while the verification probability increases with the type. For higher types, a failed verification and the resulting withholding is a significant loss, thus it is more efficient to incentivize higher types to be truthful. However, verification cannot fully substitute money burning. For lower types, withholding is not a big loss, thus higher types must also have money burning to deter lower types from mimicking them. This is intuitive in light of our examples and the direct mechanisms we described; even if a report necessitates scrutiny in the form of an external audit, the bureaucratic requirements (such as an application process) do not disappear.

Finally, we turn to the allocation probabilities and the cutoffs. Our main contribution here is to translate the endogenous incentive constraints into *monotone side constraints*. This translation also applies to *any* usage of instruments as long as the basic monotonicity properties are maintained. After transforming the problem, we rely on the polymatroid optimization literature (Fujishige (1980)) to determine the optimal allocation probabilities. The allocation probabilities turn out to be *greedy*, that is, the highest reported type receives the object with ties broken uniformly. If the endogenous cutoff is above the exogenous one, then the optimal mechanism coincides with the one in Vohra (2011) where only costly verification is used. If the endogenous cutoff is below the exogenous one, then it corresponds to the lowest type. This means that if money burning is used, it is used to fully differentiate all the types. With all the instruments identified, the choice of the optimal cutoff, and therefore the entire optimal mechanism, boils down to a comparison of the two values. We also identify sufficient conditions to identify the presence of money burning without solving the problem.

Research on costly verification in mechanism design is not new. The main difficulty that arises in case of costly verification is the presence of non-local incentive constraints, which is usually resolved by the identification of a cutoff type. Townsend (1979) models optimal contracts with costly verification, while Border and Sobel (1987), Mookherjee and Png (1989), and Dunne and Loewenstein (1995) further explore problems related to costly state verification. Ben-Porath et al. (2014) and Erlanson and Kleiner (2019) study allocation problems with multiple agents in the presence of costly verification techniques, and Vohra (2012) considers a discrete version of the same problem. Li (2020) takes a different approach by considering a costly verification problem where the designer has limited punishments available. Mylovanov and Zapechelnyuk (2017) takes a slightly different approach where verification is done ex-post, with limited punishments. Notable recent contributions also include Perez-Richet and Skreta (2023) where both local and global incentive constraints appear as part of the allocation problem without transfers and Kattwinkel and Knoepfle (2023) which shows that a cutoff structure remains when the principal also has private information.

There is also substantial research on money burning in mechanism design. McAfee et al. (1992), Hartline and Roughgarden (2008), Yoon (2011), and Chakravarty and Kaplan (2008) all investigate allocation problems involving some form of money burning. Similarly, much of the literature on money burning addresses the problem in isolation. Money burning has also been used interchangeably with costly signaling in mechanism design and cheap talk literature, as seen in Austen-Smith and Banks (2000). When money burning is only costly to the agent there is clearly no use for verification. In fact the setup reduces to one of costly signalling. Therefore, we focus on cases where money burnt is costly to both parties.

2 Model

2.1 Basic Setup

There are D agents, denoted by $1, 2, \ldots, D$. Each agent without the object is assumed to have a normalized productivity of 0. We assume that each agent i has a productivity (value) v_i associated with the object drawn from an independent and identically distributed (i.i.d.) distribution F over a finite space $V = \{1, 2, \ldots, \mathcal{V} - 1, \mathcal{V}\}$.

A principal has one object to allocate among the agents. If an agent receives the object, his share of increased productivity resulting from allocation is r, while the principal receives a share R, where $r, R > 0$. This can be interpreted as both the agents and the principal receiving fixed commissions dictated by an external entity, such as the headquarters of a large company where the principal corresponds to a branch manager.

Both the agents and the principal are assumed to be risk-neutral. The principal has two instruments at her disposal: costly verification and money burning. We assume verification is perfect and the cost for perfect verification of a claim is denoted by $k \in V$, regardless of the agent's type (costly only to the principal).[4] We further assume that there exists at least one valuation $v \leq V$ such that $v > k$, indicating that verification can potentially be useful. We will further assume $k > 1$ so that verification is not so cheap as to render money burning unnecessary. We define money burning as the agent spending p units of productive time, which is interpreted as time wasted in unproductive activities. The principal incurs these costs because the time wasted by the agent directly impacts their productivity. Therefore,

[4] We assume $k \in V$ for notational simplicity; the analysis holds with more cumbersome notation for any $1 < k < \mathcal{V}$.

it is costly to both sides at their respective rates r, R.

At the outset, the principal commits to a finite extensive form game (mechanism). The principal's objective is to maximize the expected value from the allocation, minus the costs of any assessments and the costs due to money burning activities. The agents, on the other hand, aim to maximize their share of the output after accounting for the money burning they incur. The solution concept for this mechanism is the Bayes Nash Equilibrium.

Throughout the process, we assume the agents have an outside option valued at 0 available to them; they can quit at any point in the game designed by the principal. This assumption is fairly natural in many examples. If a researcher is expected to write very long and tedious applications with little to no probability of receiving a grant, they will simply quit midway through their application. Similarly, if a town is expected to spend a large amount of public funds on lobbying without receiving a hospital, they will simply stop wasting further resources on such endeavors. The direct implication here is that the ability to quit limits the liability of the agent and prevents them from facing very large punishments. In particular, the trivial mechanism that checks each claim with vanishing probability and punishes lies with almost infinite amounts of money burning is ruled out. It is also worth noting that one can consider additional timing restrictions such as in Mylovanov and Zapechelnyuk (2017), where verification can only be done ex-post but the punishment is potential partial withholding/repossession of the object. As will soon be clear, as long as the punishment cannot violate ex-post individual rationality (whether it is withholding/repossesion, money burning or a combination), such timing restrictions do not alter the results. For example, whether grant applications are checked before or after allocation doesn't matter as long as the punishment for a lie is the removal of said grant (or its worth), and not assuming that the researcher will be forced to write grant applications till the end their career without having a chance to ever receive one.

2.2 Revelation Principle

In principle, the mechanism can be quite complex, involving multiple rounds of checks and money burning. However, it turns out that optimal, truthful, and direct mechanisms take a remarkably simple form.

Proposition 1 (Direct Mechanisms). *The following describes an optimal, direct revelation mechanism:*

1. *The principal commits to a selection policy, an inspection policy, and a money burning policy;*

2. *Agents report their types to the principal;*

3. *Following the selection policy, an agent is selected for Stage 4;*

4. *The selected agent burns the required money;*

5. *The selected agent is inspected according to the inspection policy;*

6. *The selected agent receives the object if there is no inspection, or if the agent is found truthful upon inspection.*

Both the optimality of such mechanisms and the revelation principle are straightforward extensions of the argument presented in Ben-Porath et al. (2014). Thus, the formal proof of Proposition 1 is delegated to the online appendix. It is evident that we only need truthful mechanisms since the principal can simply simulate the BNE of any complicated game after asking for the types. An agent's claim is verified only if he is considered for allocation since there is no point in spending money on inspecting an agent's claim otherwise. For maximum penalty, the agents that lie must have the object withheld from them. Money burning cannot be a punishment as it is costly to the principal, and agents cannot be required to burn money without a chance of allocation; they will simply leave. Finally, money burning must precede inspection in an optimal mechanism as inspection is fully revealing.

The requirement of perfect verification is essential to maintain a direct revelation mechanism. This aspect was also acknowledged in Ben-Porath et al. (2014) concerning the problem with only costly verification, and the reasoning in this setting is identical. In the absence of perfect verification, there is value in conducting sequential checks, potentially leading to extensive form games of arbitrary length.

Money burning activities can manifest in various forms, such as imposing extensive documentation requirements or conducting multiple rounds of time-consuming meetings that consume valuable resources. We refer to these activities as bureaucracy, which incurs costs without generating any direct output. The extent of bureaucracy is directly dictated by the principal. Alternatively, agents can resort to expected lobbying or "buttering up the boss," where some amount of time on activities that are not productive is expected by the principal. Regardless of the specific form, we assume money burning involves engaging in unproductive activities that *do not provide any verifiable information*, hence it is different from (possibly imperfect) information acquisition which is costly to both parties. Albeit being an interesting extension, imperfect information acquisition through money burning presents a similar technical challenge to that of imperfect verification and is not considered.

2.3 Formal Structure and Notation

To facilitate the analysis, it will be convenient to work with a fixed distribution. Thus, we will first introduce the following change of type space.

Definition 1 (Uniformization and Type Space)**.** *For almost any F, we can duplicate types as many times as necessary to have a type space $T = \{1, \dots, n\}$ where the distribution over types is uniform, but the valuations associated with types are only weakly increasing.*[5] *We will denote the values associated with the types after uniformization as $\{v_1, \dots, v_n\}$. With a slight abuse of notation, we will use the notation $f(v)$ for $v \leq \mathcal{V}$ to denote how many copies of valuation v exist. Clearly, $\frac{f(v)}{n}$ is the probability of valuation v under F, and $\frac{f(v)}{\sum_{i=v}^{\mathcal{V}} f(i)}$ is the hazard rate.*

The symmetric environment clearly implies that it is without loss to treat duplicate types identically, as any asymmetry can be randomized over relabeling of type indices. Thus we will assume duplicate types are treated identically (pooled).

The cost of verification defines an exogenous cutoff, the smallest (index) type where verification is not a guaranteed loss of value.

Definition 2 (Efficient Verification Level)**.** *The efficient verification level, t^{EF}, is the smallest type where the value exceeds the cost of verification. Formally, $t^{EF} = \min\{t \in \{1, \dots, n\} : v_t \geq k\}$; the associated value is k.*[6]

We will call types above t^{EF} (including its duplicates with higher indices) "verification-efficient types" and types below t^{EF} "inefficient types".

A direct mechanism maps a type profile in T^D to allocation probabilities, money burning requirements, and verification probabilities. Formally, a mechanism is three functions: $p_i(t_i, t_{-i}) : T^D \mapsto \mathbb{R}_+$, $a_i(t_i, t_{-i}) : T^D \mapsto [0, 1]$, and $c_i(t_i, t_{-i}) : T^D \mapsto [0, 1]$. Here $p_i(t_i, t_{-i})$ is the amount of money burning by agent i. Without loss of generality, we state p_i unconditionally on selection to simplify the statement of incentive constraints. Similarly, $a_i(t_i, t_{-i})$ is the probability of the i^{th} agent being selected in Stage 3 for further review. Finally, we have $(1 - c_i(t_i, t_{-i}))$ as the probability that a costly inspection of the selected agent is done in Stage 5. Thus, $c_i(t_i, t_{-i})$ corresponds to the conditional (on allocation) probability of **no** verification, capturing the probability of getting away with a misreport.

[5] The "almost" qualifier is for cases when the probabilities are not rational; any such distribution can be approximated as closely as desired.

[6] As noted earlier, if k is not in V, we would simply introduce additional notation for the value of t^{EF} and proceed analogously.

Within this setting, the principal wishes to maximize:

$$E\left[\sum_{i=1}^{D}\left[R(v_{t_i}a_i(t_i,t_{-i})-p_i(t_i,t_{-i}))-\frac{k}{R}\left(1-c_i(t_i,t_{-i})\right)a_i(t_i,t_{-i})\right]\right]$$

subject to:

$$rE_{t_{-i}}\left[v_{t_i}a_i(t_i,t_{-i})-p_i(t_i,t_{-i})|t_i\right]\geq rE_{t_{-i}}\left[v_{t_i}a_i(\hat{t}_i,t_{-i})c_i(\hat{t}_i,t_{-i})-p_i(\hat{t}_i,t_{-i})|t_i\right]\forall t_i,\hat{t}_i$$

$$rE_{t_{-i}}\left[v_{t_i}a_i(t_i,t_{-i})-p_i(t_i,t_{-i})|t_i\right]\geq 0 \qquad \forall t_i$$

From the linear structure above, it is clear that different values of r, R can simply be captured by a change of k. Thus, going forward, we will assume, without loss of generality, $r = R = 1$ with k adjusted accordingly.[7] Furthermore, since the agent types are i.i.d., it is without loss to focus on symmetric mechanisms. Thus, we can focus on the interim objects:

$\mathcal{A}_t = E_{t_{-i}}\left[a_i(t_i,t_{-i})\|t_i=t\right]$	Interim allocation probability
$p_t = E_{t_{-i}}\left[p_i(t_i,t_{-i})\|t_i=t\right]$	Interim money burning
$c_t = \frac{E_{t_{-i}}\left[c_i(t_i,t_{-i})a_i(t_i,t_{-i})\|t_i=t\right]}{E_{t_{-i}}\left[a_i(t_i,t_{-i})\|t_i=t\right]}$	Conditional probability of **no** verification

Clearly, $0 \leq c_t \leq 1$ since $0 \leq c_i(t_i, t_{-i}) \leq 1$. Furthermore, we have $p_t \geq 0$ for all $t \in T$ since $p_i(t_i, t_{-i}) \geq 0$. Of course, we need to ensure the interim probabilities, $\mathcal{A}_t$, can be achieved using an allocation rule. Border (1991) provides us with a characterization of the set of feasible interim allocations. We use $\langle \mathcal{A} \rangle$ to denote the collection $\langle \mathcal{A}_t \rangle_{t\in T}$. The inequalities state that $\langle \mathcal{A} \rangle \in \mathbb{R}^n_+$ is feasible if and only if,

$$\frac{D}{n}\sum_{t\in S}\mathcal{A}_t \leq 1-\left(1-\frac{|S|}{n}\right)^D \qquad \forall S \subseteq T.$$

For any S, the inequality can be written as $\sum_{t\in S}\mathcal{A}_t \leq g(S)$ where we define $g : \mathcal{P}(T) \mapsto \mathbb{R}^+$ as $g(S) = \frac{n}{D} - \frac{n}{D}\left(1-\frac{|S|}{n}\right)^D$; $\mathcal{P}(\cdot)$ denotes the set of all subsets. $g(S)$ is essentially a function of the cardinality of the set S. One useful property of g is that it is strictly submodular, thus the feasible set of allocations is a *polymatroid*.[8]

In order to provide an intuitive understanding of the polymatroid structure notice that ex-ante each agent in a symmetric mechanism expects to receive the object at $1/D$ probability. If the agent also knows that they are the highest type, their interim expected probability cannot be 1, in a symmetric mechanism. The highest they can expect is the

[7] Without loss of generality we maintain the assumption this adjusted $k \in V$ to avoid additional notational burden.

[8] A formal proof is provided in the online appendix for completeness.

case where the highest type wins with ties broken at uniform. Say this is some number $a > 0$. Now, suppose instead the agent knows that they are the second highest type. Under the same rule as above (highest type wins with ties broken at uniform) the interim probability they face must be strictly less than a. Thus the **sum** of the interim probabilities that are given to the highest type and the second highest type is increasing, at a decreasing rate, i.e. submodularly in the set of types considered, which defines a consistency requirement. Under other rules (say the second highest type wins) the probabilities for different types will differ, but will still need to satisfy an internal consistency of submodular growth. The polymatroid structure gives a precise definition of such consistent allocation probabilities. We refer readers to Border (1991) for a complete argument.

Let us denote by the set $\mathscr{A}$ the set of all feasible intermediate allocations that satisfy Border's inequalities. Thus, we can interpret a mechanism as a tuple $M = \{\langle \mathcal{A} \rangle, \langle c \rangle, \langle p \rangle\}$. Using symmetry and uniformization, we can write the problem as:

$$\max_{\langle \mathcal{A} \rangle, \langle c \rangle, \langle p \rangle} \sum_{t \in T} (v_t \mathcal{A}_t - p_t + k\mathcal{A}_t c_t - k\mathcal{A}_t)$$

$$\text{s.t} \quad \langle \mathcal{A} \rangle \in \mathscr{A}; \quad 0 \leq c_t p_t \leq p_t \ \forall t \in T; \ c_t, p_t \in \mathbb{R}^+ \tag{F}$$

$$v_t \mathcal{A}_t - p_t \geq v_t \mathcal{A}_{\hat{t}} c_{\hat{t}} - p_{\hat{t}} \quad \forall t, \hat{t} \in T \tag{IC}$$

$$v_t \mathcal{A}_t - p_t \geq 0 \qquad \forall t \in T. \tag{IR}$$

We say that a mechanism $M = \{\langle \mathcal{A} \rangle, \langle c \rangle, \langle p \rangle\}$ is feasible if $\langle \mathcal{A} \rangle \in \mathscr{A}$ and $0 \leq c_t \leq 1 \ \forall t \in V$; $c_t, p_t \in \mathbb{R}_+$. A mechanism is incentive compatible if it satisfies all IC constraints. When referring to particular values instead of types, with a slight abuse of notation, we will use a superscript such as $\mathcal{A}^v, c^v, p^v$ to refer to the relevant objects for any t such that $v_t = v$.

2.4 Example Optimal Mechanism and Comparison

Before we proceed further let us analyze a simple example with two agents and compare it to the scenario with only verification (Ben-Porath et al. (2014)) and only money burning (Hartline and Roughgarden (2008)). Each agent has a valuation drawn i.i.d. from $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$ with $P(v = 1) = \frac{23}{30}$ and $P(v = x) = \frac{1}{30}$ for all $x \neq 1$.The cost of inspection is $k = 5$. Table 1 represents an optimal mechanism in the three respective cases.[9]

First, let us consider the optimal mechanism with only money burning. In this case, one can see that there is allocative efficiency, all types are differentiated.[10] Additionally, when there is only money burning, the only binding constraints are local constraints. Indeed, given the incentive constraints, the mechanism is expected to behave very similarly to an

[9] All numbers are truncated at the second decimal.

[10] Recall that we must have $\sum_{i=1}^{7} \mathcal{A}^i \frac{f(i)}{n} = 1/D$ by Border's Inequalities.

	Both			Only Verification		Only Money Burning	
v	$\mathcal{A}^v$	$1-c^v$	p^v	$\mathcal{A}^v$	$1-c^v$	$\mathcal{A}^v$	p^v
1	0.38	0	0	0.46	0	0.38	0
2	0.78	0	0.4	0.46	0	0.78	0.40
3	0.81	0	0.46	0.46	0	0.81	0.46
4	0.85	0	0.56	0.46	0	0.85	0.56
5=k	0.88	0	0.70	0.46	0	0.88	0.7
6	0.91	0.03	0.70	0.46	0	0.91	0.86
7	0.95	0.7	0.70	0.95	0.50	0.95	1.06
8	0.98	0.10	0.70	0.98	0.52	0.98	1.3

Table 1: An Optimal Mechanism

auction with money burning that is strictly increasing in the valuation.

Now let us consider the optimal mechanism with only costly verification. In this case, we see that there is inefficient pooling, below (inclusive) an endogenous cutoff value of 6. The reasoning for this is evident when we consider the verification probabilities: if the principal wants to lower the cutoff to 5, the verification probabilities of all the types above would need to increase, along with introducing verification to type 6. Thus, pooling turns out to be the more efficient option. When we consider the binding constraints, we see that besides constraints between the pooled types, the only binding constraints are between the cutoff type 6 and higher types. Notably, there is no binding constraint between type 7 and 8. And even for differentiating these types the principal suffers a great deal of losses as the verification probabilities are quite high to deter lower types from misreporting.

Now consider the optimal mechanism where both instruments are available. All types are differentiated. Money burning is increasing until the exogenous cutoff 5. However, after 5, the money burning plateaus. At a glance verification at a cost of 5 might seem overly expensive. But recall that $1-c$ is the conditional probability, the unconditional probability is $\frac{f(v)}{n}\mathcal{A}^v(1-c^v)$. One might also ask why it isn't completely substituted away but remains as a complement to verification. This is because verification has no cost on the agent. For higher types, verification is more effective since withholding is a costly punishment to them. For lower types, withholding is not a big deterrent, thus money burning for higher types is a cheaper alternative to deter lower types from declaring as high types. Finally, the set of binding constraints is fairly complex. Type 1 only has a local constraint to type 2, and type 2 only has a local constraint to type 3 and type 3 only has a local constraint to type 4. However, type 4 has binding constraints to all the types above it. Similarly, type 5 also has binding constraints to all the types above it. For all types that are strictly above k, there are no binding constraints; they strictly prefer being truthful. Finally notice that money burning first increases and then flattens out, whereas verification start out flat and then increases.

3 Results

3.1 Basic Monotonicities

In standard mechanism design, having binding (tight) incentive constraints often simplifies the problem by reducing it to just choosing allocation probabilities. The usage of instruments, such as required payments, can be deduced from these allocation probabilities. As the example above demonstrates, in our setting, even if we have knowledge of the allocation probabilities and the set of tight incentive constraints, determining the verification probabilities or required money burning is not straightforward; some constraints are made tight with just money burning, some constraints are made tight with both. Narrowing down the set of tight incentive constraints ex ante by intuition, such as only the local constraints to be binding, is both potentially erroneous and difficult.

Our first goal is to try to at least limit the set of binding constraints in a more methodical manner. Our approach will be to see if we can find improvements on a mechanism by replacing or generating some binding constraints. An important concept that facilitates in our analysis is the notion of *mutually binding incentive constraint set.* A set of types S has jointly mutually binding constraints if, for any pair of types s and t in S, type t is indifferent between revealing its true type and pretending to be type s, and type s is indifferent between revealing its true type and pretending to be type t. Note that this does not exclude the possibility of types outside S wanting to mimic types in S, or types in S wanting to mimic types outside S; it simply limits these behaviors to one-sided mimicry. In environments involving only money burning (or some standard scenarios like auctions), mutually binding incentive constraints are equivalent to pooling, where all types are treated identically. In cases with only costly verification, pooling is necessary but not sufficient to establish mutually binding incentive constraints. For example, it is possible to have two types with the same allocation probability and a strictly positive probability of verification, causing them to be pooled but without a desire to mimic each other. With two instruments at play, types may still want to mimic each other despite being treated differently, making pooling neither necessary nor sufficient.

Our first result demonstrates that by leveraging *optimality* and *mutually binding incentive constraints*, we can drop all the downward-facing constraints from the problem and it remains unchanged. In other words, we can assume only upward constraints bind unless the types are pooled *and* have mutually binding constraints. Furthermore, this result provides a formalization of the extent of substitution that can be inferred directly from incentive constraints. Specifically, it shows that money burning increases with allocation without verification.

Proposition 2. *In an optimal mechanism, agents are strictly better off by reporting truthfully than any strictly lower value unless the types are pooled and have mutually binding constraints. Additionally, money burning is strictly monotonic in the probability of allocation without an assessment, i.e., in an optimal mechanism,* $\mathcal{A}_t c_t > \mathcal{A}_s c_s \iff p_t > p_s$ *and* $\mathcal{A}_s c_s = \mathcal{A}_t c_t \iff p_s = p_t$*.*

The first part of the proof relies on the following fact: it is possible to "move" higher allocations to lower ones in a polymatroid. In particular if $\mathcal{A}_t \geq \mathcal{A}_{t'}$ for some t, t' in a feasible mechanism, then there is a small enough $\delta > 0$ such that $\mathcal{A}_t - \delta$, and $\mathcal{A}_{t'} + \delta$, with the other allocations unchanged, remains feasible in the polymatroid (even if $\mathcal{A}_t = \mathcal{A}_{t'}$). These kind of pairwise perturbations can then be used, along with changes to the instruments, to figure out some basic, technical, characteristics of an optimal mechanism (Lemmas 2, 3, 4, 5). When combined, these characteristics are then used to show the statement of the proposition by simply appealing to the fact that there cannot be any objective-improving perturbations while respecting these basic characteristics.

The second part of the proposition, which can be straightforwardly derived from incentive compatibility, is of particular importance. It is noteworthy that the proposition's statement does not make any assumptions regarding the presence of money burning in the mechanism. In particular, when there is no money burning, for any pair of types t and s, we must have $\mathcal{A}_s c_s = \mathcal{A}_t c_t$. In the case of costly verification alone, this observation is sufficient to determine the optimal usage of verification from allocation probabilities. Furthermore, it is worth noting that this type of monotonicity resulting from incentive compatibility alone does not provide significant information by itself, as it is difficult to ascertain whether $\mathcal{A}_t c_t$ or p_t should increase or decrease with respect to the type.

A more desirable aspect of monotonicity relates to allocation probabilities and utilities. It turns out in an optimal mechanism, higher types are indeed associated with higher allocation probabilities. This intuition aligns with the principal's interest in achieving allocative efficiency. Moreover, since we have established that only the upward constraints are binding, we can expect higher types to enjoy higher incentive rents and, consequently, higher expected utilities.

Proposition 3. *In an optimal mechanism, higher types have higher interim allocation probabilities, i.e.,* $v_t > v_{\hat{t}} \implies \mathcal{A}_t \geq \mathcal{A}_{\hat{t}}$*. Furthermore, higher types also have higher expected payoffs, i.e.,* $v_t > v_{\hat{t}} \implies v_t \mathcal{A}_t - p_t \geq v_{\hat{t}} \mathcal{A}_{\hat{t}} - p_{\hat{t}}$*.*

For proving these results, we show that if the proposition is not true, then the higher type must have a positive inspection probability, which eventually leads to a contradiction of *optimality.* It is noteworthy that monotonicity results such as Proposition 3 only require *incentive compatibility* in much of the classical mechanism design literature. However, with two instruments, incentive compatibility is not enough, and we must use optimization.

While Propositions 2 and 3 provide desirable and intuitive results, additional structure is still necessary to determine verification probabilities and money burning. Our analysis so far establishes that the binding constraints are always upward-looking, but they need not be local, so the structure is still quite loose. Our next result delivers the first structural component, the endogenous cutoff.

Proposition 4. *Each optimal mechanism adheres to a structure in which there exists a threshold value $\underline{t} \in T$ such that types with valuations v_t less than or equal to $v_{\underline{t}}$ undergo neither verification nor money burning. Conversely, types with valuations exceeding this threshold, $v_t > v_{\underline{t}}$, are subject to either a positive probability of verification, some degree of money burning, or both. Additionally, all types falling below this threshold have the lowest probability of being implemented, specifically, $\mathcal{A}_t = \min_{r \in V} \mathcal{A}_r$ for all $t \leq \underline{t}$..*

Since lower types have a lower chance of allocation, it is not expected that these very low-value types would typically undergo assessment or involve significant money burning. This intuition is formalized through the introduction of a cutoff, where types below the cutoff have a fixed probability of allocation. No money burning is required for these low-value types, and verification is also not employed.

The proof of this result relies on several auxiliary results that capture the properties of an optimal mechanism and their relation to mutually binding incentive constraints sets. In the appendix, we first establish that optimality requires any cycle of incentive constraints to be in a mutually binding incentive constraint set. For example, suppose t wanted to mimic t' and t' wanted to mimic t'' and t'' wanted to mimic t, forming a cycle. Optimality will dictate that, in fact, we must have that the desire to mimic is mutual between any pair from $\{t, t', t''\}$. (Recall this still doesn't rule out say t'' wanting to also mimic some type t''', where t''' is not part of the cycle). Moreover, optimality will again show that in a mutually binding incentive constraint set with more than one type, all the types must have the *same amount of money burning, chance of allocation, and no verification* (Lemma 8), ergo, they are pooled. Finally, in every optimal mechanism, there is a unique mutually binding incentive constraint set with no outside types wanting to mimic any type within the set (Lemma 9). No outside types wanting to mimic any inside type implies there cannot be any no money burning in this minimal set either, reinforcing the idea that these types should be grouped at the bottom, giving rise to a cutoff structure.

Remark 1. *It is important to note here the structure at this point is far from complete; we have only ruled out some sets of binding constraints but we have not identified any particular collection. Moreover, our understanding of the usage of instruments is also fairly limited. However, it is worth noting that only this amount of structure is enough to actually identify all the allocation probabilities for any endogenous cutoff, despite the incentive constraints still remaining somewhat undetermined. It is also important to note that our analysis so far uses some basic tools from graph theory in addition to standard optimization arguments as we detail in the appendix similar to the analysis in Vohra (2012) and Vohra (2011).*

3.2 Optimal usage of Instruments

To advance our analysis, we will now shift our attention to another structural element of the problem, namely the instruments. Essentially, our problem can be conceptualized as a polymatroid optimization problem with incentive constraints. Incentive constraints are inherently linked to the set of binding constraints, about which we have limited information. Therefore, our initial approach revolves around determining the optimal use of incentive instruments given any allocation. This step is not typically required in many standard mechanism design frameworks, where binding incentive constraints directly dictate the use of instruments based on allocation.

Proposition 4 identifies an endogenous cutoff type $\underline{t}$. However, it's essential to recall that in our problem setup, there's also an exogenous cutoff, t^{EF}, representing types not considered "worth" verifying due to the associated cost. As we will soon demonstrate, the interaction between $\underline{t}$ and t^{EF} significantly influences the optimal use of instruments.

Proposition 5. *Let $\underline{t}$ be the cutoff type in an optimal mechanism, and let $v_{\underline{t}} \geq k$. Then, $p_t = 0$ for all t, and $c_t = \frac{\mathcal{A}_{\underline{t}}}{\mathcal{A}_t}$ for all $t \geq \underline{t}$.*

To gain intuition, let's first observe that if there is a type associated with money burning, no type below the cutoff can mimic this type. The reasoning behind this is straightforward: if that were the case, the incentive constraint of the cutoff type would be violated. Since we are considering the scenario where the cutoff type itself is verification-efficient, the only incentive constraints exist among efficient types. However, given their already high valuation, verification-efficient types are willing to burn a significant amount of money. Therefore, it's always worthwhile to substitute money burning for verification to induce higher types to be truthful.

Proposition 5 also speaks to the environments when money burning is completely prohibited. The optimal mechanism is characterized by a cutoff type as in Proposition 4, and verification probabilities are precisely determined by the ratio of allocation probabilities, as observed in Proposition 2. This result mirrors the findings of Ben-Porath et al. (2014) and Vohra (2012).

On the other hand, when the endogenous cutoff falls below t^{EF}, money burning becomes a necessary component of an optimal mechanism. In fact, as we will demonstrate, for all types below t^{EF} but above the cutoff, money burning is the only instrument utilized. This intuitively makes sense, as attempting to verify these types would not be worth the associated cost. Additionally, the required level of money burning increases monotonically until t^{EF}. Formally, we have:

Proposition 6. *Let $\underline{t}$ be the cutoff type in an optimal mechanism, and let $v_{\underline{t}} < k$. For all $t \leq t^{EF}$, $c_t = 1$, $p_t = v_{t-1}(\mathcal{A}_t - \mathcal{A}_{t-1}) + p_{t-1}$, with $p_{\underline{t}} = 0$. Additionally, if $v_{t-r} = v_{t-1}$ for any $r \leq t-1$, then, $p_t = v_{t-r}(\mathcal{A}_t - \mathcal{A}_{t-r}) + p_{t-r}$.*

Building upon the above result, our next aim is to consider the usage of incentive instruments for types above t^{EF}. For these types, verification is not a guaranteed loss as the cost of verification is below their value. Indeed, the principal starts using verification for these types. One might potentially consider whether the principal would like to substitute away from money burning completely for these types; after all, verification is perfect. However, as we show below, the types above t^{EF} have both money burning and costly verification.

Proposition 7. *Let $\underline{t}$ be the cutoff type in an optimal mechanism, and let $v_{\underline{t}} < k$. Then for all $t > t^{EF}$, $p_t = p_{t^{EF}}$ (including the duplicates of t^{EF} with higher indices), and $c_t = \frac{\mathcal{A}_{t^{EF}}}{\mathcal{A}_t}$.*

In addition to providing a comprehensive understanding of the instrument usage for a sufficiently low cutoff value, the aforementioned result reveals intriguing monotonicity properties. Money burning reaches its peak at t^{EF} and remains constant thereafter, while the verification probabilities initially remain constant at 0 but then increase monotonically. For verification-efficient types, money burning is relatively insignificant compared to the potential loss resulting from withholding due to failed verification. Therefore, it emerges as the more efficient option for maintaining incentive constraints among these types. However, lower types still need deterrence from misrepresenting themselves as verification-efficient types; hence money burning remains a necessary component. Notably, instead of completely replacing money burning, verification acts as a complement to it.

The proofs of Propositions 6 and 7 rely on standard linear programming techniques. Using duality, we show that the type immediately above the cutoff only needs to have a positive payment and no verification, and by induction, we extend it to all the types below t^{EF}. For the types above t^{EF}, the upper bounding technique for the simplex algorithm (Eiselt and Sandblom (2007), pp. 212-218) delivers that these types must be subject to verification, but the optimal way to introduce verification turns out to be fixing the money burning for these types.

3.3 Allocation

At this juncture, having fully understood the utilization of the instruments for any set of allocation probabilities, our next objective is to determine these allocations precisely. In an auction where transfers from the agents to the principal take place, withholding or pooling, thus, allocative inefficiency is not required. However, the findings presented above indicate that a combination of costly verification and money burning alone cannot fully compensate for the absence of transfers unless the cutoff is the lowest type. As a result, allocative inefficiency via pooling is observed, but it is confined solely to the types below the cutoff. Additionally, as we will soon demonstrate, the allocation above the cutoff is greedy, similar to an auction, thus the highest announced type receives the object, breaking ties uniformly.

Proposition 8. *Let V be the space be the space of let $f(v)$ denote the number of duplicates in T. Let $\mathcal{A}^v = \mathcal{A}_t$ denote the allocation probability of for any type $t \in T$ with $v_t = v$. In any optimal mechanism, for a given cutoff $\underline{t}$ with value $\underline{v}$, the probability of allocation for any type t with value v_t is given by*

$$
\begin{aligned}
\mathcal{A}^{\mathcal{V}} &= \frac{g(f(\mathcal{V}))}{f(\mathcal{V})} \\
\mathcal{A}^{\mathcal{V}-1} &= \frac{g\left(f(\mathcal{V}) + f(\mathcal{V}-1)\right) - g(f(\mathcal{V}))}{f(\mathcal{V}-1)} \\
&\vdots \\
\mathcal{A}^{v} &= \frac{g\left(\sum_{i=v}^{\mathcal{V}} f(i)\right) - g\left(\sum_{i=v+1}^{\mathcal{V}} f(i)\right)}{f(v)} \\
\bar{\mathcal{A}}^{\underline{v}} &:= \frac{g\left(\sum_{i=1}^{\mathcal{V}} f(i)\right) - g\left(\sum_{i=\underline{v}+1}^{\mathcal{V}} f(i)\right)}{\sum_{i=1}^{\underline{v}} f(i)}
\end{aligned}
$$

Where $\bar{\mathcal{A}}^{\underline{v}}$ denotes the probability of allocation for any given type that is below or equal to the cutoff type $\underline{v}$.

Many mechanism design problems beyond this setup fall into a category of problems called polymatroid optimization with incentive constraints (see Border (1991), Vohra (2012)), and the methodology of tackling them and how they relate to polymatroid optimization with monotone side constraints might be of independent interest. Thus, let us briefly sketch the argument for the proof of Proposition 8. Proposition 2 shows that an optimal mechanism has an endogenous cutoff type $\underline{t}$. Given a cutoff, we first take $\langle \mathcal{A}c, p\rangle$ (neither being necessarily the optimal ones we derived above) as given and ask what the optimal $\langle \mathcal{A}\rangle$ is. Notice changing $\langle \mathcal{A} \rangle$ while taking $\langle \mathcal{A}c \rangle$ as given is equivalent to changing c (equivalent to solving a relaxed problem where $c > 1$ is allowed). Due to Proposition 2, for each potentially binding IC regarding $t' \geq t$, we can let $d_{t,t'} = \mathcal{A}_{t'}c_{t'} - \frac{p_{t'}}{v_t} + \frac{p_t}{v_t}$ and let $d_t = \max_{t' \geq t}\{d_{t,t'}\}$. Define the associated polytope $\mathcal{P}(\mathcal{A}c, p) = \{\sum_{i \in S} \mathcal{A}_i \leq g(S), S \subseteq T; \mathcal{A}_i \geq d_i,\ i \in T\}$. The optimal $\mathcal{A}'s$ given $\mathcal{A}c, p$'s solve the following problem:

$$\max_{\langle \mathcal{A} \rangle \in P(\mathcal{A}c, p)} \sum_{t=1}^{n} v_t \mathcal{A}_t$$

This is a linear optimization over a polymatroid identified by the set function $g(S)$ and **side constraints** d_i. At this point, we don't know the values of individual $d_i's$, but by Proposition 3 and feasibility, we can restrict attention to d_i's that are **weakly monotone**. Clearly, if $(d_1, \ldots, d_n) \notin \mathcal{P}(\mathcal{A}c, p)$ then the problem is not feasible. If the problem is feasible, by proposition 2.2 of Fujishige (1980), we can consider *the contraction of the polymatroid with respect to the constraints* d_i, which in turn also defines a polymatroid without constraints. This polymatroid consists of all the vectors $(\mathcal{B}_i)_{i \in \{1,\ldots,n\}}$ such that $(\mathcal{B}_i + d_i)_{i \in \{1,\ldots,n\}}$ is feasible in the original polymatroid. In particular the function $g(S)$ is replaced by a function $\rho(S) = \min_{S' \supseteq S}(g(S') - \sum_{i \in S'} d_i)$. The weak monotonicity of the $d_i's$ turns out to be enough to identify the structure of this new polymatroid. This new polymatroid problem then can be simply solved by a greedy algorithm starting from the highest type since there are no more side constraints. Given the form of the contraction from $\mathcal{B}$'s, we can back out $\mathcal{A}$'s for the given $\mathcal{A}c, p$. We then show that if the resulting $\mathcal{A}$'s are not of the stated form $\mathcal{A}c, p$ cannot be optimal.

Remark 2. *The allocation problem is tackled by solely considering the weak monotonicity properties (section 3.1). This analysis does not necessitate specific knowledge of instrument usage or the set of binding constraints. In fact, the usage of the instruments can be calculated by standard linear programming methods algorithmically after finding the allocations here. Furthermore, up to this point, the analysis does not depend on any properties of the distribution, such as the structure of the set V or the hazard rate assumption.*

3.4 Optimal Cutoff

With the mechanism fully determined based on the cutoff, the final step is to identify the optimal cutoff itself. Unlike the characterization up to this point, the distribution now starts to play an important role. Thus, from this point onward, we will directly use the values instead of the duplicated types.

The presence of money burning is closely tied to the cutoff, making it essential to explore both scenarios: with and without money burning. In particular, we can try to start from 1 as the cutoff, where there is allocative efficiency, and slowly increase the cutoff to calculate the effect on the total payoff. The only difficulty we would face is that when the cutoff reaches k, there will necessarily be a discontinuity since money burning is going to vanish. Letting $\Pi(\underline{v})$ denote the payoff of the principal when the cutoff is $\underline{v}$, we provide a discrete analogue of the derivative with respect to the cutoff in the following two lemmas. First, we consider an increase with the cutoff still remaining below k.

Lemma A. *If the cutoff is below the efficient verification level, $\underline{v} < k - 1$,*

$$\Pi(\underline{v}+1) - \Pi(\underline{v}) \propto \sum_{i=1}^{\underline{v}} f(i)(i - \frac{1}{h(\underline{v}+1)})$$
$$\propto E(v - \frac{1}{h(\underline{v}+1)} | v \leq \underline{v}).$$

Where $\frac{1}{h(\underline{v}+1)}$ is the inverse hazard rate at $\underline{v}+1$.

There are a few noteworthy points regarding the above. First, the characterization continues to hold for $\Pi(k) - \Pi(k-1)$. Second, under the monotone hazard rate assumption, the inverse hazard rate is non-increasing, thus, this derivative is necessarily increasing. This has a direct implication: if there is money burning it comes with full separation of types and thus allocative efficiency. To see this suppose the cutoff below the efficient verification level is $k > \hat{v} > 1$. This means that it was worthwhile to increase the cutoff from $\hat{v} - 1$ to $\hat{v}$, but the above lemma directly implies that in that case it would be optimal to it all the way to k. Thus if money burning exists then the endogenous cutoff is a trivial one, i.e. $\underline{v} = 1$.

Now let us consider the analogue of the discrete derivative when the cutoff considered is above k.

Lemma B. *If the cutoff is above the efficient verification level, $\underline{v} \geq k$,*

$$\Pi(\underline{v}+1) - \Pi(\underline{v}) \propto \sum_{i=1}^{\underline{v}+1} (i - (\underline{v}+1)) f(i) + k \sum_{i=1}^{\mathcal{V}} f(i)$$
$$\propto E(v - \underline{v} + 1 | v \leq \underline{v}) + \frac{k}{P(v \leq \underline{v})}.$$

In this case, it is easy to see that the derivative is decreasing as the cutoff increases, since $\sum_{i=1}^{\underline{v}+1} (i - (\underline{v}+1)) f(i)$ is decreasing in $\underline{v}$. Therefore, a candidate for the cutoff when there is no money burning is the level where the derivative first becomes negative. Clearly, if the derivative is negative throughout, then k is the only candidate when the mechanism doesn't have money burning. Conversely, if the derivative is positive throughout, then $\mathcal{V}$ is the cutoff without money burning, and all the types are pooled. Unlike the previous lemma, it turns out the above characterization of the derivative is robust to changes in the distribution.

With the two preceding Lemmas, the cutoff simply boils down to checking level where the derivative becomes negative for the no-money-burning case and comparing it to the cutoff of 1.

Theorem 1. *Let* $v^{NoMB} = \max\{\mathcal{V} > \hat{v} \geq k : E(v - \hat{v} + 1 | v \leq \hat{v}) + \frac{k}{P(v \leq \hat{v})} \leq 0\}$ *if the set is non-empty, and* $\mathcal{V}$ *otherwise. Then the optimal cutoff solves*

$$\arg \max_{v \in \{1, v^{NoMB}\}} \Pi(v)$$

Since we did not restrict $\mathcal{V}$ or k, and the distribution beside the monotone hazard rate assumption, ex-ante, necessary and sufficient statements about the presence of money burning or not are difficult to make. Nonetheless, the two lemmas preceding the theorem are sharp characterizations of the cutoffs in either case. Thus, combined with the complete characterization of the rest of the mechanism, identifying the cutoffs boils down to comparing easy-to-calculate payoffs.

We can also identify easy sufficient conditions to check for the presence of money burning using Lemmas A and B.

Corollary 1. *If* $E(v - \frac{1}{h(k)} | v < k) < 0$ *and* $E(v - k + 1 | v \leq k) + \frac{k}{P(v \leq k)} < 0$*, then the optimal mechanism has money burning.*

The first condition of the corollary simply guarantees that it is worth it to lower the cutoff from k to $k - 1$, introducing money burning. Then, by Lemma A we know that lowering the cutoff all the way down is worth it. This is a necessary condition for money burning to exist. However, by itself it is not sufficient since it is possible that increasing cutoff above the level k might also lead to an increase in the objective and which one is larger depends on the specifics of the distribution, values and k. The second condition guarantees that increasing the cutoff from k to $k + 1$ results in a loss. Then, by Lemma B we know that increasing the cutoff will only accrue more losses.

We can also determine sufficient conditions to check when money burning is not part of the optimal mechanism.

Corollary 2. *If the inverse hazard rate of type* 2 *is low enough,* $1 - \frac{1}{h(2)} > 0$*, then the optimal mechanism has* ***no*** *money burning.*

Recall that $E(1 - \frac{1}{h(2)} | v \leq 1) \propto 1 - \frac{1}{h(2)}$, that is the condition is simply capturing the marginal gain to increasing the cutoff. The corollary follows from the fact that the inverse hazard rate is non-increasing under monotone hazard rate assumption, guaranteeing the marginal gains to increasing the cutoff increases until k. It is only sufficient, because it is possible that $E(1-\frac{1}{h(2)}|v<2)<0$ but it becomes positive enough before k eliminating the need for money burning.

It is also worth noting that the analysis up to the determination of the cutoff is completely independent of both the set of values (i.e., it can have holes) or whether the distribution has monotone hazard rate or not. The difficulty without the structural assumptions is that the analogue of the discrete derivatives stops being monotone when those two assumptions are dispensed. In principle, this means that the optimal cutoff (which can now also be larger than 1 with money burning) can still be found by simply comparing associated Π for each v and picking the largest one. This is easy to do in small applications, but quickly becomes cumbersome as the value space gets larger.

3.4.1 Comparative Statics

As suggested by Theorem 1, conducting broad comparative statics without imposing stringent assumptions is challenging within this framework. Two main difficulties arise, limiting the extent of the analysis. Firstly, the presence of the hazard rate of the cutoff, as opposed to the virtual valuation, indicates that simple first-order or second-order stochastic dominance shifts generally do not provide conclusive results. Secondly, the polymatroid base introduces another obstacle. It can be shown, for instance, that the $\mathcal{A}_t$ terms are all decreasing in D, but this decrease follows a non-monotonic pattern.

However, the cost of verification is amenable to analysis without additional assumptions. It is worth noting that the relationship between R and r can be captured by a change in the cost of verification k. Hence, if the choice of R and r (with some tradeoff) is also considered, the overall impact can be comprehended through the effect of the cost of verification.

Proposition 9. *The optimal cutoff $\underline{t}$ is increasing in the cost of verification k. Consequently, the probability of verification, $1-c_t$, is weakly decreasing in k for all t.*

The insight behind this result is somewhat nuanced. As the cost of verification increases, the principal faces the option to rely more on money burning and decrease dependence on verification, while maintaining allocation probabilities unchanged. However, the changing value of t^{EF} adds complexity to the relationship.

If the original mechanism had no money burning, an increase in the cost of verification resulting in an increase of t^{EF} might imply that there are now more types below t^{EF} (even the previously optimal cutoff might now be pooled), and the principal might want to screen some of them at the cost of introducing money burning. On the other hand, the required money burning to separate the types also changes with an increase in t^{EF}. Specifically, an increase in t^{EF} leads to an increase in the required money burning for types above the new t^{EF}, while the money burning for other types remains the same. Thus, the principal also simultaneously faces a larger money burning bill as a result of the increase in the cost of verification. It turns out this increase in money burning is enough to outweigh the potential gain of screening types that fell below t^{EF}, leading to a weak increase in the cutoff.

The weakly increasing cutoff implies that verification probabilities decrease, as shown in Propositions 5 and 7. Therefore, as the cost of verification increases, the amount of verification and the associated costs in the mechanism weakly decrease.

When examining the relationship between money burning and the cost of verification, Propositions 6, 7, and 9 may initially suggest that money burning increases since t^{EF} increases. If money burning does remain present it indeed increases. On the other hand, if the cutoff does increase, our theorem 1 indicates that money burning vanishes altogether. However, much like theorem 1, whether money burning vanishes or not depends on the specifics of the distribution and valuations.

4 Conclusions and Discussion

We have developed a model with minimal assumptions. The principal is allowed to use costly inspection and observable money burning activities for optimal allocation. From a mechanism design perspective, we observe that when money burning activities are not available, there is significant pooling. Money burning activities enable the principal to discourage low types from mimicking high types, thereby avoiding unnecessary pooling or eliminate high verification probabilities. Contrary to the perception that money burning activities merely destroy surplus, they can serve as a useful instrument for reducing incentive rents, which cannot be directly substituted.

On the technical front, we approach the mechanism design problem with non-local incentive constraints and multiple instruments. The potential applicability of a similar approach to other mechanism design problems is of interest. Traditionally, mechanism design does not require complex optimizations to handle the use of incentive instruments, thanks to the simplifications provided by an implicit understanding of binding constraints. However, our work demonstrates that in the absence of such an understanding, studying previously unexplored objects such as mutually binding incentive constraints and their interaction with maximization can be a fruitful endeavor. Additionally, the translation of polymatroids incentive constraints into associated polymatroids with side constraints is a potentially valuable technique that can be applied to a broader range of problems. Notably, the allocation of multiple objects with incentive constraints shares a similar polymatroid base, making our techniques applicable in such scenarios. We aim to address this in future research.

5 Appendix

Definition 3 (Transform of a Mechanism)**.** *M^* is a transform of the mechanism M denoted as $M^* = T_{M,r,s}(\mathcal{A}'_r, \mathcal{A}'_s, c'_r, c'_s, p'_r, p'_s)$. If $M^* = \{\langle \mathcal{A}^* \rangle, \langle c^* \rangle, \langle p^* \rangle\}$ then $\mathcal{A}^*_r = \mathcal{A}'_r$; $\mathcal{A}^*_s = \mathcal{A}'_s$; $c^*_r = c'_r$; $c^*_s = c'_s$; $p^*_r = p'_r$; $p^*_s = p'_s$; $\mathcal{A}^*_t = \mathcal{A}_t \ \forall t \notin \{r, s\}$; $c^*_t = c_t \ \forall t \notin \{r, s\}$; $p^*_t = p_t \ \forall t \notin \{r, s\}$.*

In order to simplify the optimization of the binding constraints we propose interpreting a mechanism M as a directed graph. In particular, we will be using the edges to represent the binding constraints. This provides us a convenient language to analyze effect of changing the binding constraints. Let G_M be the directed graph that represents the mechanism M. Let $E(G_M)$ be the set of edges where $(s, t) \in E(G_M)$ means that there is a directed edge originating at s and pointing towards t in G_M. Since all graphs we consider have the same vertex set, we simply use T to denote the set of vertices. We say that $(s, t) \in E(G_M)$ if the incentive constraint for type s pretending to be type t is tight. That is $(s, t) \in E(G_M)$ if and only if

$$v_s \mathcal{A}_s - p_s = v_s \mathcal{A}_t c_t - p_t$$

We call a graph G_M associated with an optimal mechanism M an optimal graph.

Definition 4 (In-Out Sets)**.** *For any $t \in V$ let $deg^-_M(t)$ denote the indegree of t and let $deg^+_M(t)$ denote the outdegree of t in G_M. The in-set of type t is $i_M(t) = \{r : (r, t) \in E(G_M)\}$ out-set of type t is $o_M(t) = \{r : (t, r) \in E(G_M)\}$. For any $S \subseteq V$ in-set of S is $i_M(S) = \{t \in V \backslash S : (t, s) \in E(G_M)$ for some $s \in S\}$.*

The edges originating from a node t are the types that type t wants to mimic, and the edges pointing to a node t are the types that want to mimic t. A **set of types with mutually binding incentive constraints** corresponds to a special structure on a graph known as a *web*.

Definition 5 (Web)**.** *$W \subseteq T$ is a* web *in G_M iff $(s, t) \in E(G_M)$ $\forall s, t \in W$. A web W is* non-trivial *if $|W| > 1$.*

In essence, a web is a set of types where every node points to every other node in W, and no more nodes can be added to it while still having every node pointing to every other node. A web necessarily contains directed cycles, but directed cycles need not contain webs. There can be outside types pointing to a web or types in a web pointing outside. For example if the set of edges is $\{(t, t'), (t', t), (t, t''), (t''', t)\}$, the collection $\{t, t'\}$ is a web, since the edges are mutual, t'' or t''' does not belong in a web despite the former being pointed to by t and the latter pointing to t.

Lemma 1. *Let $\langle \mathcal{A} \rangle \in \mathscr{A}$ and $\mathcal{A}_t \geq \mathcal{A}_s$ for some s, t, then $\exists\ \delta > 0$ such that $\langle \mathcal{A}' \rangle \in \mathscr{A}$, where $\mathcal{A}'_t = \mathcal{A}_t - \delta$, $\mathcal{A}'_s = \mathcal{A}_s + \delta$ and $\mathcal{A}'_r = \mathcal{A}_r\ \forall r \notin \{s, t\}$.*

Proof. Consider an arbitrary $S \subseteq V \backslash \{s, t\}$. Let $S_t \equiv S \cup \{t\}$, $S_s \equiv S \cup \{s\}$. Since $\mathcal{A}_s \leq \mathcal{A}_t$ we can write

$$\sum_{r \in S_s} \mathcal{A}_r \leq \frac{1}{2} \left[\sum_{r \in S} \mathcal{A}_r + \sum_{r \in S_t \cup S_s} \mathcal{A}_r \right]$$

We need to show,

$$\sum_{r \in S'} \mathcal{A}'_r \leq g(S') \qquad \forall S' \subseteq V$$

For the sets that do not contain s, or contain both s, t the inequality holds trivially. Since S is arbitrary and δ can be arbitrarily small, we need to show:

$$\sum_{r \in S_s} \mathcal{A}_r < g(S_s)$$

Since g is strictly submodular

$$\begin{aligned}
& g(S_s \cup S_t) + g(S) < g(S_s) + g(S_t) \\
\implies \quad & \sum_{r \in S} \mathcal{A}_r + \sum_{r \in S_t \cup S_s} \mathcal{A}_r < g(S_s) + g(S_t) \\
\implies \quad & 2 \sum_{r \in S_s} \mathcal{A}_r < g(S_s) + g(S_t) \\
\implies \quad & \sum_{r \in S_s} \mathcal{A}_r < g(S_s) \qquad \because g(S_s) = g(S_t)
\end{aligned}$$

□

Lemma 2. *In an optimal M if $deg^-_M(t) = 0$, then $p_t = 0$; $c_t = 1$ and $\mathcal{A}_t = \min_{r \in V} \mathcal{A}_r$*

Proof. If $p_t > 0$ then define $M' = T_{M,s,t}(\mathcal{A}_s, \mathcal{A}_t, c_s, c_t, p_s, p_t - \delta)$ for small $\delta > 0$ then M' is feasible and a δ improvement. If $c_t < 1$ then define $M' = T_{M,s,t}(\mathcal{A}_s, \mathcal{A}_t, c_s, c_t + \delta, p_s, p_t)$ for small $\delta > 0$ then M' is feasible and a $k\mathcal{A}_t\delta$ improvement. If $\mathcal{A}_t > \mathcal{A}_s$ for some s we then must have $v_s\mathcal{A}_s - p_s < v_s\mathcal{A}_t$. Since $p_s \geq 0$, IC of s is violated. □

Lemma 3. *In an optimal M if $i_M(t) = \{t' \in T \setminus \{t\} : v_{t'} = v_t\}$, then $p_t = 0$; $c_t = 1$ and $\mathcal{A}_t = \min_{r \in V} \mathcal{A}_r$*

Proof. Recall that all duplicates are treated identically. If $p_t > 0$ then simply lower p_t for all duplicates for small $\delta > 0$ this is feasible and a $\delta f(v_t)$ improvement. If $c_t < 1$ then similarly just increase c_t for all duplicates for small $\delta > 0$ resulting $k\mathcal{A}_t\delta f(t_t)$ improvement. If $\mathcal{A}_t > \mathcal{A}_s$ for some s we then must have $v_s\mathcal{A}_s - p_s < v_s\mathcal{A}_t$. Since $p_s \geq 0$, IC of s is violated. □

Lemma 4. *In an optimal, incentive compatible mechanism M if $(t, \hat{t}) \in E(G_M)$ then $\mathcal{A}_{\hat{t}}c_{\hat{t}} \geq \mathcal{A}_t$*

Proof. Suppose not, and let $\mathcal{A}_{\hat{t}}c_{\hat{t}} < \mathcal{A}_t$, then it must be that $p_t > p_{\hat{t}} \geq 0$. Consider

$$M' = T_{M,t,\hat{t}}(\mathcal{A}_{\hat{t}}c_{\hat{t}}, \mathcal{A}_{\hat{t}}, c_t, c_{\hat{t}}, p_{\hat{t}}, p_{\hat{t}})$$

M' is feasible. Incentive constraints not involving the type t remain unaffected. For $r \in o_M(t)$, we have

$$\begin{aligned}
v_t\mathcal{A}'_t - p'_t &= v_t\mathcal{A}_{\hat{t}}c_{\hat{t}} - p_{\hat{t}} \\
&= v_t\mathcal{A}_t - p_t & \because (t, \hat{t}) \in E(G_M) \\
&\geq v_t\mathcal{A}_r c_r - p_r \quad \forall r \neq t \\
&= v_t\mathcal{A}'_r c'_r - p'_r
\end{aligned}$$

For all $r \in T$, thus also all $r \in i_M(t)$ we have

$$v_r\mathcal{A}'_r - p'_r = v_r\mathcal{A}_r - p_r \geq v_r\mathcal{A}_{\hat{t}}c_{\hat{t}} - p_{\hat{t}} \geq v_r\mathcal{A}_{\hat{t}}c_{\hat{t}}c_t - p_{\hat{t}} = v_r\mathcal{A}'_t c'_t - p'_t$$

Thus, all incentive constraints are satisfied. The objective is increased by,

$$Obj(M') - Obj(M) = k(1 - c_t)(\mathcal{A}_t - \mathcal{A}_{\hat{t}}c_{\hat{t}})$$

This is positive if $c_t < 1$ giving a contradiction. If $c_t = 1$ then either $i_M(t) = \{t' \in T \setminus \{t\} : v_{t'} = v_t\}$ or $i_M(t) = \emptyset$ or there exists some r in $i_M(t)$ with $v_r \neq v_t$. In the first two cases we can simply lower p_t for all duplicates by a small $\delta > 0$ and this is a $\delta f(v_t)$ improvement giving the desired contradiction. So we must have the last case to be true. let $r' \in i_M(t)$ and $r' < t$. If $v_{r'} < v_t$ then

$$\begin{aligned}
& v_t\mathcal{A}_t - p_t = v_t\mathcal{A}_{\hat{t}}c_{\hat{t}} - p_{\hat{t}} \\
\implies \quad & v_{r'}(\mathcal{A}_t - \mathcal{A}_{\hat{t}}c_{\hat{t}}) < p_t - p_{\hat{t}} & \because \mathcal{A}_t > \mathcal{A}_{\hat{t}}c_{\hat{t}}; \\
\implies \quad & v_{r'}\mathcal{A}_{r'} - p_{r'} < v_{r'}\mathcal{A}_{\hat{t}}c_{\hat{t}} - p_{\hat{t}} & \because c_t = 1;\ (r', t) \in E(G_M)
\end{aligned}$$

Again, giving us a contradiction. So we must have $v_{r'} \geq v_t$ for all r' such that $(r', t) \in E(G_M)$ and $c_t = 1$. Also recall that all duplicates are pooled thus all share the same in-set. Now let $\tilde{t}$ be the smallest non-duplicate of t in $i_M(t)$. We must have $v_{\tilde{t}} > v_t$ and $v_{t'} \geq v_{\tilde{t}}$ for all $t' \in i_M(t)$ such that $v_{t'} \neq v_t$.

Given above consider lowering p_t for all duplicates for small $\delta v_{\tilde{t}}$ since $p_t > 0$ and also lowering $\mathcal{A}_t$ for all duplicates by δ. For any $r' \in i_M(t)$ such that $v_{r'} \neq v_t$

$$
\begin{aligned}
& v_{r'}\mathcal{A}_{r'} - p_{r'} = v_{r'}\mathcal{A}_t - p_t \\
\implies \quad & v_{r'}\mathcal{A}_{r'} - p_{r'} \geq v_{r'}(\mathcal{A}_t - \delta) - p_t + \delta v_{\tilde{t}} \because v_{r'} \geq v_{\tilde{t}}
\end{aligned}
$$

Therefore IC continues to hold for all $r' \in i_M(t)$. Notice for all types that were not in the in-set of t originally, their incentive constraint regarding t and its duplicates were strict. Thus a small enough δ maintains those strict incentives. The objective has increased by $f(v_t)\delta(v_{\tilde{t}} - v_t)$ giving us a contradiction. □

Lemma 5. *If M is optimal, $(s, t) \in E(G_M)$ and $\mathcal{A}_s = \mathcal{A}_t c_t$ then $c_s = 1$.*

Proof. If s has indegree zero then Lemma 3 delivers the result. Thus, assume $\exists r$ such that (r, s) bind, and let $c_s < 1$. Since $\mathcal{A}_s = \mathcal{A}_t c_t$ and $v_s\mathcal{A}_s - p_s = v_s\mathcal{A}_t c_t - p_t$ we have $p_s = p_t$. Since (r, s) bind,

$$
\begin{aligned}
v_r\mathcal{A}_r - p_r &= v_r\mathcal{A}_s c_s - p_s \\
&= v_r\mathcal{A}_t c_t c_s - p_s \\
&< v_r\mathcal{A}_t c_t - p_t
\end{aligned}
$$

contradicting the IC constraint. □

Proof of Proposition 2. The second part of the proposition is proved separately as Lemma 6. For the first part, suppose not and let $v_s > v_t$ but s points to t. Note that Lemma 4 gives us $\mathcal{A}_t \geq \mathcal{A}_s$. Also note that we cannot have $\mathcal{A}_t = 0$ as that would imply both s, t are outside the mechanism with zero allocation no money burning and no verification. We consider four cases,

Case I: $c_t < 1$; $o_M(t) \neq \emptyset$

Let $r \in o_M(t)$. We have,

$$\begin{aligned}
& v_t\mathcal{A}_t - p_t = v_t\mathcal{A}_r c_r - p_r \\
\implies \quad & p_r - p_t \leq v_s\left(\mathcal{A}_r c_r - \mathcal{A}_t\right) && \because v_s > v_t;\ \mathcal{A}_r c_r \geq \mathcal{A}_t \text{ due to Lemma 4} \\
\implies \quad & p_r - p_t < v_s\left(\mathcal{A}_r c_r - \mathcal{A}_t c_t\right) && \because c_t < 1 \\
\implies \quad & v_s\mathcal{A}_s - p_s < v_s\mathcal{A}_r c_r - p_r && \because (s,t) \in E(G_M)
\end{aligned}$$

Violating the IC constraint.

Case II: $c_t < 1;\ o_M(t) = \emptyset$
Consider the mechanism

$$M' = T_{M,s,t}\left(\mathcal{A}_s + \delta, \mathcal{A}_t - \delta, \frac{\mathcal{A}_s c_s}{\mathcal{A}_s + \delta}, \frac{\mathcal{A}_t c_t}{\mathcal{A}_t - \delta}, p_s, p_t\right)$$

then M' is feasible due to Lemma 1 and the fact that $c_t < 1$. IC's not involving s, t remain unchanged. $\mathcal{A}'_t c'_t = \mathcal{A}_t c_t$ and $\mathcal{A}'_s c'_s = \mathcal{A}_s c_s$ and so no other type has an incentive to pretend to be s or t. Surplus of s has increased so it has no incentive to deviate in M' and since t has outdegree zero, for small enough δ, t will not deviate. M', however, is an improvement of $\delta(v_s - v_t)$ in the objective function.

Case III: $c_t = 1;\ \mathcal{A}_t = \mathcal{A}_s$
From Lemma 5 we have that $c_s = 1$. Since $v_s\mathcal{A}_s - p_s = v_s\mathcal{A}_t - p_t$ we get $p_s = p_t$ and so $s \in i_M(t);\ t \in i_M(s)$ and they must belong in a web. Since they are treated the same they are pooled.

Case IV: $c_t = 1;\ \mathcal{A}_t > \mathcal{A}_s$
We have,

$$\begin{aligned}
& p_t - p_s = v_s(\mathcal{A}_t - \mathcal{A}_s) \\
\implies \quad & p_t - p_s > v_t(\mathcal{A}_t - \mathcal{A}_s) && \because v_t < v_s \text{ and } \mathcal{A}_s < \mathcal{A}_t \\
\implies \quad & v_t\mathcal{A}_s - p_s > v_t\mathcal{A}_t - p_t
\end{aligned}$$

Consider $M' = T_{M,s,t}\left(\mathcal{A}_t, \mathcal{A}_s, c_t, c_s, p_t, p_s\right)$. s, t have no incentives to deviate in M' since their surplus is not decreasing. No other type has any incentive to pretend to be s or t either since they would otherwise have done that in M. M' improves the objective by $(v_t\mathcal{A}_s - p_s) - (v_t\mathcal{A}_t - p_t) > 0$.

Lemma 6. *In an optimal mechanism, the amount of money burning is strictly monotonic in the probability of allocation without an assessment, i.e., in an optimal M, $\mathcal{A}_t c_t > \mathcal{A}_s c_s \iff p_t > p_s$ and $\mathcal{A}_s c_s = \mathcal{A}_t c_t \iff p_s = p_t$.*

Proof. If either type has an in-degree 0 Lemma 3 that type has no verification, no money burning and has minimal allocation probability. Wlog assume this is type s. By Lemma **??** $\mathcal{A}_t c_t \geq \mathcal{A}_s$. If $\mathcal{A}_t c_t = \mathcal{A}_s$, then p_t cannot be positive in an optimal mechanism as simply lowering p_t will yield an improvement. Now, let s', t' be such that $(s', s) \in E(G_M)$ and $(t', t) \in E(G_M)$ then we have

$$v_{s'} \mathcal{A}_s c_s - p_s \geq v_{s'} \mathcal{A}_t c_t - p_t$$
$$v_{t'} \mathcal{A}_t c_t - p_t \geq v_{t'} \mathcal{A}_s c_s - p_s$$

Combining the two we get

$$v_{t'}(\mathcal{A}_t c_t - \mathcal{A}_s c_s) \geq p_t - p_s \geq v_{s'}(\mathcal{A}_t c_t - \mathcal{A}_s c_s)$$

Thus concluding the proof. □

□

Proof of Proposition 3. The second part of the proposition is proved separately as a Lemma 7 below. Let $\hat{t}, t$ be such that $v_t > v_{\hat{t}}$ and $\mathcal{A}_t < \mathcal{A}_{\hat{t}}$ in an optimal mechanism M. If $deg^-_M(\hat{t}) = 0$ then the proposition is trivially true due to Lemma 3, so $i_M(\hat{t}) \neq \emptyset$. The proof consists of two parts. First, we show that for this mechanism M we must have $c_t < 1$ and $p_{\hat{t}} > 0$.

Suppose $c_t = 1$ or $p_{\hat{t}} = 0$. If $c_{\hat{t}} = c_t = 1$ then from the IC constraints of $t, \hat{t}$ we have

$$v_{\hat{t}}(\mathcal{A}_{\hat{t}} - \mathcal{A}_t) \geq p_{\hat{t}} - p_t \geq v_t(\mathcal{A}_{\hat{t}} - \mathcal{A}_t)$$

which is not possible since $\mathcal{A}_{\hat{t}} > \mathcal{A}_t$ and $v_t > v_{\hat{t}}$, so:

$$c_{\hat{t}} c_t < 1 \tag{1}$$

If $deg^+_M(\hat{t}) = 0$ and $c_{\hat{t}} < 1$ then consider the mechanism M'' with

$$M' = T_{M,\hat{t},t}\left(\mathcal{A}_{\hat{t}} - \delta, \mathcal{A}_t + \delta, \frac{\mathcal{A}_{\hat{t}} c_{\hat{t}}}{\mathcal{A}_{\hat{t}} - \delta}, \frac{\mathcal{A}_t c_t}{\mathcal{A}_t + \delta}, p_{\hat{t}}, p_t\right)$$

By Lemma 1, for small δ this is feasible. All IC constraints can be verified to be satisfied for small δ and the objective function improvement is $\delta(v_t - v_{\hat{t}})$. And so if $deg^+_M(\hat{t}) = 0 \implies c_{\hat{t}} = 1$. However, this implies that $p_{\hat{t}} > 0$, since otherwise $\hat{t}$ must have lowest allocation probability and $\mathcal{A}_{\hat{t}} \leq \mathcal{A}_t$ contradicting the supposition. It also implies $c_t < 1$ by (1). Thus, we have $\neg(c_t < 1 \& p_{\hat{t}} > 0) \implies deg^+_M(\hat{t}) > 0$.

Now, since $deg^+_M(\hat{t}) > 0$ (i.e. $o_M(\hat{t}) \neq \emptyset$), let $\hat{t}' \in o_M(\hat{t})$. We have:

$$
\begin{aligned}
& v_t\mathcal{A}_t - p_t \geq v_t\mathcal{A}_{\hat{t}'}c_{\hat{t}'} - p_{\hat{t}'} \\
\implies \quad & p_{\hat{t}'} - p_t \geq v_t(\mathcal{A}_{\hat{t}'}c_{\hat{t}'} - \mathcal{A}_t) \\
\implies \quad & p_{\hat{t}'} - p_t > v_{\hat{t}}(\mathcal{A}_{\hat{t}'}c_{\hat{t}'} - \mathcal{A}_t) \qquad \because \text{Lemma 4} : \mathcal{A}_{\hat{t}'}c_{\hat{t}'} \geq \mathcal{A}_{\hat{t}} > \mathcal{A}_t \\
\implies \quad & v_{\hat{t}}\mathcal{A}_t - p_t > v_{\hat{t}}\mathcal{A}_{\hat{t}'}c_{\hat{t}'} - p_{\hat{t}'}
\end{aligned}
$$

$$
\implies \quad v_{\hat{t}}\mathcal{A}_t - p_t > v_{\hat{t}}\mathcal{A}_{\hat{t}} - p_{\hat{t}} \tag{2}
$$

Consider $M' = T_{M,\hat{t},t}(\mathcal{A}_t, \mathcal{A}_t, c_t, c_t, p_t, p_t)$, which is IC by (2). Objective changes by

$$
\begin{aligned}
& v_{\hat{t}}\mathcal{A}_t - p_t + k\mathcal{A}_t c_t - k\mathcal{A}_t - [v_{\hat{t}}\mathcal{A}_{\hat{t}} - p_{\hat{t}} + k\mathcal{A}_{\hat{t}}c_{\hat{t}} - k\mathcal{A}_{\hat{t}}] \\
= & (v_{\hat{t}}\mathcal{A}_t - p_t - v_{\hat{t}}\mathcal{A}_{\hat{t}} + p_{\hat{t}}) + k(\mathcal{A}_{\hat{t}} - \mathcal{A}_t) + k(\mathcal{A}_t c_t - \mathcal{A}_{\hat{t}}c_{\hat{t}})
\end{aligned}
$$

Which is strictly positive if $c_t = 1$ since the k terms will reduce to $k(\mathcal{A}_{\hat{t}} - \mathcal{A}_{\hat{t}}c_{\hat{t}})$ and the first term is positive by equation 2. It also strictly positive if $\mathcal{A}_t c_t \geq \mathcal{A}_{\hat{t}}c_{\hat{t}}$. Thus $c_t < 1$ and $\mathcal{A}_{\hat{t}}c_{\hat{t}} > \mathcal{A}_t c_t$ implying $p_{\hat{t}} > p_t \geq 0$ due to lemma 6.

Now given $c_t < 1$ & $p_{\hat{t}} > 0$, consider $M^* = T_{M,\hat{t},t}(\mathcal{A}^*_{\hat{t}}, \mathcal{A}^*_t, c^*_{\hat{t}}, c^*_t, p^*_{\hat{t}}, p^*_t)$ where,

$$
\begin{aligned}
\mathcal{A}^*_{\hat{t}} = \mathcal{A}_{\hat{t}} - \varepsilon;\ c^*_{\hat{t}} = c_{\hat{t}} - \varepsilon\left(\frac{v_t(\mathcal{A}_{\hat{t}}c_{\hat{t}} - \mathcal{A}_t c_t) - c_{\hat{t}}(p_{\hat{t}} - p_t)}{(p_{\hat{t}} - p_t)(\mathcal{A}_{\hat{t}} - \varepsilon)}\right);\ p^*_{\hat{t}} = p_{\hat{t}} - v_t\varepsilon \\
\mathcal{A}^*_t = \mathcal{A}_t + \varepsilon;\ c^*_t = c_t + \varepsilon\left(\frac{v_t(\mathcal{A}_{\hat{t}}c_{\hat{t}} - \mathcal{A}_t c_t) - c_t(p_{\hat{t}} - p_t)}{(p_{\hat{t}} - p_t)(\mathcal{A}_t + \varepsilon)}\right);\ p^*_t = p_t + v_t\varepsilon
\end{aligned}
$$

This part of the proof proceeds in 3 steps.

Step 1 - M^* is feasible: Lemma 1 shows that $\mathcal{A}^*_{\hat{t}}$ and $\mathcal{A}^*_t$ are feasible for small ε. $p_{\hat{t}} > 0$ and so $p_{\hat{t}}, p_t$ are feasible. c^*_t is feasible for small enough ε, the previous part established $c_t < 1$. Also notice that $c_{\hat{t}} > 0$ as otherwise no type would ever want to mimic it which would in turn $p_{\hat{t}} = 0$ in an optimal mechanism. Let us show $0 < c^*_{\hat{t}} \leq c_{\hat{t}}$. The first inequality holds for ε sufficiently small. For the second inequality we must show

$$
\left(\frac{v_t(\mathcal{A}_{\hat{t}}c_{\hat{t}} - \mathcal{A}_t c_t) - c_{\hat{t}}(p_{\hat{t}} - p_t)}{(p_{\hat{t}} - p_t)(\mathcal{A}_{\hat{t}} - \varepsilon)}\right) \geq 0 \tag{3}
$$

Consider $\underline{\hat{t}} = \inf i_M(\hat{t})$, then $t > \hat{t} > \underline{\hat{t}}$ due to Proposition 2. (Notice duplicate types cannot belong to the same web when $c < 1$ hence they do not point to each other) Thus:

$$\begin{aligned} & v_{\underline{\hat{t}}}\mathcal{A}_{\hat{t}}c_{\hat{t}} - p_{\hat{t}} \geq v_{\underline{\hat{t}}}\mathcal{A}_t c_t - p_t \\ \implies \quad & v_{\underline{\hat{t}}}(\mathcal{A}_{\hat{t}}c_{\hat{t}} - \mathcal{A}_t c_t) \geq p_{\hat{t}} - p_t \\ \implies \quad & v_t(\mathcal{A}_{\hat{t}}c_{\hat{t}} - \mathcal{A}_t c_t) \geq p_{\hat{t}} - p_t \qquad \because \mathcal{A}_{\hat{t}}c_{\hat{t}} > \mathcal{A}_t c_t \\ \implies \quad & v_t(\mathcal{A}_{\hat{t}}c_{\hat{t}} - \mathcal{A}_t c_t) - c_{\hat{t}}(p_{\hat{t}} - p_t) \geq 0 \end{aligned}$$

Proving (3).

Step 2 - M^* is IC: The only changes concern type $\hat{t}$ and t. We have $v_{\hat{t}}\mathcal{A}^*_{\hat{t}} - p^*_{\hat{t}} \geq v_{\hat{t}}\mathcal{A}_{\hat{t}} - p_{\hat{t}}$ and $v_t\mathcal{A}^*_t - p^*_t \geq v_t\mathcal{A}_t - p_t$ and so $\hat{t}, t$ have no incentive to lie. Define

$$\delta = \frac{v_t \varepsilon}{p_{\hat{t}} - p_t}$$

Then we see

$$\begin{aligned} p^*_{\hat{t}} &= (1-\delta)p_{\hat{t}} + \delta p_t \\ p^*_t &= (1-\delta)p_t + \delta p_{\hat{t}} \\ \mathcal{A}^*_{\hat{t}}c^*_{\hat{t}} &= (1-\delta)\mathcal{A}_{\hat{t}}c_{\hat{t}} + \delta\mathcal{A}_t c_t \\ \mathcal{A}^*_t c^*_t &= (1-\delta)\mathcal{A}_t c_t + \delta\mathcal{A}_{\hat{t}}c_{\hat{t}} \end{aligned}$$

Consider any arbitrary type r:

$$\begin{aligned} v_r\mathcal{A}^*_t c^*_t - p^*_t &= v_r\left((1-\delta)\mathcal{A}_t c_t + \delta\mathcal{A}_{\hat{t}}c_{\hat{t}}\right) - \left((1-\delta)p_t + \delta p_{\hat{t}}\right) \\ &= (1-\delta)\left[v_r\mathcal{A}_t c_t - p_t\right] + \delta\left[v_r\mathcal{A}_{\hat{t}}c_{\hat{t}} - p_{\hat{t}}\right] \\ &\leq v_r\mathcal{A}_r - p_r \end{aligned}$$

And so r doesn't mimic t. Similar argument shows that r doesn't mimic $\hat{t}$ either.

Step 3 - M^* is an improvement:

$$\begin{aligned} Obj(M^*) - Obj(M) &= v_{\hat{t}}\mathcal{A}^*_{\hat{t}} - p^*_{\hat{t}} + k\mathcal{A}^*_{\hat{t}}c^*_{\hat{t}} - k\mathcal{A}^*_{\hat{t}} - (v_{\hat{t}}\mathcal{A}_{\hat{t}} - p_{\hat{t}} + k\mathcal{A}_{\hat{t}}c_{\hat{t}} - k\mathcal{A}_{\hat{t}}) \\ &\quad + v_t\mathcal{A}^*_t - p^*_t + k\mathcal{A}^*_t c^*_t - k\mathcal{A}^*_t - (v_t\mathcal{A}_t - p_t + k\mathcal{A}_t c_t - k\mathcal{A}_t) \\ &= (v_t - v_{\hat{t}})\varepsilon - k\delta(\mathcal{A}_{\hat{t}}c_{\hat{t}} - \mathcal{A}_t c_t) + k\varepsilon \\ &\quad + k\delta(\mathcal{A}_{\hat{t}}c_{\hat{t}} - \mathcal{A}_t c_t) - k\varepsilon \\ &= (v_t - v_{\hat{t}})\varepsilon > 0 \end{aligned}$$

Contradicting the optimality of M. $\square$

Lemma 7. *$v_t\mathcal{A}_t - p_t \geq v_{\hat{t}}\mathcal{A}_{\hat{t}} - p_{\hat{t}}$ if $v_t \geq v_{\hat{t}}$.*

Proof. Suppose not and let $v_{\hat{t}} < v_t$ with $v_{\hat{t}}\mathcal{A}_{\hat{t}} - p_{\hat{t}} > v_t\mathcal{A}_t - p_t$ then we must have $c_{\hat{t}} < 1$ otherwise $v_t\mathcal{A}_t - p_t < v_t\mathcal{A}_{\hat{t}} - p_{\hat{t}}$ violating the IC constraint. We must also have $o_M(\hat{t}) = \emptyset$ otherwise if $r \in o_M(\hat{t})$, $v_t\mathcal{A}_t - p_t < v_{\hat{t}}\mathcal{A}_{\hat{t}} - p_{\hat{t}} = v_{\hat{t}}\mathcal{A}_r c_r - p_r < v_t\mathcal{A}_r c_r - p_r$ violating the IC constraint. Now, consider

$$M' = T_{M,\hat{t},t}\left(\mathcal{A}_{\hat{t}} - \delta, \mathcal{A}_t + \delta, \frac{\mathcal{A}_{\hat{t}}c_{\hat{t}}}{\mathcal{A}_{\hat{t}} - \delta}, \frac{\mathcal{A}_t c_t}{\mathcal{A}_t + \delta}, p_{\hat{t}}, p_t\right)$$

M' is IC and is feasible due to Lemma 1 and the fact that $c_{\hat{t}} < 1$. The improvement in the objective is $\delta(v_t - v_{\hat{t}})$. $\square$

Lemma 8. *In an optimal M, if W' is a directed cycle then all the types in W' must have same probability of allocation, and the same amount of money burning with no chance of assessment. Equivalently, $W' \subseteq W$ for some web W.*

Proof. Let $W' = \{t_1, t_2, \ldots, t_n\}$ be the directed cycle then from Lemma 4 we have $\mathcal{A}_{t_1} \leq \mathcal{A}_{t_2}c_{t_2} \leq \mathcal{A}_{t_2} \leq \mathcal{A}_{t_3}c_{t_3} \leq \cdots \leq \mathcal{A}_{t_1}c_{t_1}$ thus giving us an equality throughout. From IC, we have $v_{t_1}\mathcal{A}_{t_1} - p_{t_1} = v_{t_1}\mathcal{A}_{t_2}c_{t_2} - p_{t_2} \implies p_{t_1} - p_{t_2} = 0$, hence all the agents in the cycle must have the same requirements. Since $\mathcal{A}_{t_i} = \mathcal{A}_{t_j};\ p_{t_i} = p_{t_j};\ c_{t_i} = c_{t_j} = 1\ \forall t_i, t_j \in W'$ we have $t_i \in i_M(t_j)\ \forall t_i, t_j \in W'$ and W' must be part of a web. $\square$

Lemma 9. *Every optimal G_M is weakly connected and contains exactly one web W^* with $i_M(W^*) = \emptyset$.*

Proof. Let G_M be optimal, we will first show that there must exist a web W in G_M with $i_M(W) = \emptyset$. If $\exists t \in V$ such that $i_M(t) = \emptyset$ then $\{t\}$ is the desired web. So, $deg^-_M(t) > 0\ \forall t \in V$, and G_M must contain directed cycles. Due to Lemma 8 these cycles must belong in a web thus G_M must contain webs. Let $\mathscr{W}$ be the set of webs in G_M.

Let $W^* = \arg\min_{W\in\mathscr{W}}\{\mathcal{A}_t : t \in W\}$, thus for any $t^* \in W^*$ and any $t \in W \in \mathscr{W}$, $W \neq W^*$, $\mathcal{A}_{t^*} < \mathcal{A}_t$.

It must be that $i_M(W^*) = \emptyset$, for otherwise let $(t_1, t^*) \in E(G_M);\ t^* \in W^*;\ t_1 \in V\backslash W^*$ then $\mathcal{A}_{t^*} \geq \mathcal{A}_{t_1}$ due to Lemma 4. Now, since $deg^-_M(t_1) > 0$ we can find a t_2 such that $\mathcal{A}_{t_1} \geq \mathcal{A}_{t_2}$ and so on until we end up in another cycle with $\mathcal{A}_{t_k} \leq \mathcal{A}_{t^*}$ which is a contradiction to $W^* = \arg\min_{W\in\mathscr{W}}\{\mathcal{A}_t : t \in W\}$.

W^* is the only web with $i_M(W^*) = \emptyset$. Suppose not, let W be another such web. $deg^-_M(W) = 0 \implies p_t = 0\ \forall t \in W$, for if not then we can decrease money burning for all the types in the web by small $\delta > 0$ and improve the objective function without violating any IC constraints. Similarly, $deg^-_M(W^*) = 0$ mean that there is no money burning in W^*. However, let $t \in W, t^* \in W^*$, since $p_t = p_{t^*} = 0$ and $\mathcal{A}_t > \mathcal{A}_{t^*}$, $v_{t^*}\mathcal{A}_{t^*} < v_{t^*}\mathcal{A}_t$ thus violating the IC constraint and giving a contradiction.

If the graph is not connected, then repeating the steps above every connected component must contain a web W with $deg^-_M(W^*) = 0$. However, this gives us a contradiction since there can only be one such web. □

Proof of Proposition 4. Let M^* be an optimal mechanism and consider the web W^* with $i_{M^*}(W^*) = \emptyset$. Let $\underline{t} = maxW^*$ (this implies $\underline{t}$ is the highest index among copies if there are multiple types with same valuation). Since $i_{M^*}(W^*) = \emptyset$ we must have $p_t = 0\ \forall t \in W^*$, otherwise the mechanism M' defined by decreasing money burning for all the projects in W^* by small ε is an improvement over M^* and is feasible and incentive compatible. Since $p_t = 0;\ c_t = 1\ \forall t \in W^*$ we must have $\mathcal{A}_t = \min_{r\in V}\mathcal{A}_r\ \forall t \in V$ for if not and let $\mathcal{A}_{\hat{t}} > \mathcal{A}_t$ then $v_{\hat{t}}\mathcal{A}_{\hat{t}}c_{\hat{t}} - p_{\hat{t}} < v_{\hat{t}}\mathcal{A}_t$, thus violating the IC constraint. Due to Proposition 3 we must have $t \in W^*\ \forall t < \underline{t}$, since for all such $t, \mathcal{A}_t = \mathcal{A}_{\underline{t}}$ thus implying $c_t = 1;\ p_t = 0$ and $\{(t,\underline{t}),(\underline{t},t)\} \subseteq E(G_M)$. Moreover, for any $t > \underline{t}$ we must have $\mathcal{A}_t > \mathcal{A}_{\underline{t}}$. Thus by IC every $t > \underline{t}$ must either have $c_t < 1, p_t > 0$ or both. □

Lemma 10. *Let $\underline{t}$ be the cutoff, let $\tilde{t}$ be such that $p_{\tilde{t}} > 0$. Then $\nexists t' < \underline{t}$ with $v_{t'} < v_{\underline{t}}$ and $t' \in i_M(\tilde{t})$.*

Proof. Suppose not, $\exists t' < \underline{t}$ with $t' \in i_M(\tilde{t})$. By proposition 4 we have $\mathcal{A}_{\underline{t}} = \mathcal{A}_{t'}$ Since $t' \in i_M(\tilde{t})$ and $p_{\tilde{t}} > 0$ we have:

$$p_{\tilde{t}} = (\mathcal{A}_{\tilde{t}}c_{\tilde{t}} - \mathcal{A}_{t'})\, v_{t'}$$

But since $v_{\underline{t}} > v_{t'}$.

$$p_{\tilde{t}} < \left(\mathcal{A}_{\tilde{t}}c_{\tilde{t}} - \mathcal{A}_{\underline{t}}\right) v_{\underline{t}}$$

Contradicting the IC of type $\underline{t}$. □

Proof of Proposition 5. Suppose not, let t be such that $p_t > 0$ in the optimal mechanism. By lemma 10 any type in $i_M(t)$ must be weakly larger than $\underline{t}$. Let $t' = \min i_M(t)$, and reduce p_t by ε and c_t by $\frac{\varepsilon}{\mathcal{A}_t v_{t'}}$. This leads to $i_M(t') = \{t'\}$ and the objective is also increased $\frac{1}{n}\varepsilon\mathcal{A}_t(t' - k) > 0$ delivering the desired contradiction. Since there is no money burning in the mechanism the c_t's simply follow from Lemma 6. □

Proof of Proposition 6. We are going to prove the proposition by induction.

Basis Step: Let $\underline{t}$ be the cutoff. If $v_{\underline{t}} < k$ then $p_{\underline{t}+1} = v_{\underline{t}}(\mathcal{A}_{\underline{t}+1} - \mathcal{A}_{\underline{t}})$ and $c_{\underline{t}+1} = 1$.

Proof of Basis Step. By definition p, c must be optimal for the optimal $\mathcal{A}$. Let $\mathcal{A}$ be given by the optimal mechanism, then using Proposition 3 we can restrict attention to only upwards constraints (recall duplicates are treated identically) and 10 ensures that only the cutoff type can point outward from the bottom web. Finally we also know $c_t > 0$ in any optimal mechanism since $c_t = 0$ would imply certain verification, which in turn means no type will want to mimic that type, which can be improved by increasing c_t slightly for that type. Thus, the problem of choosing c and p is as follows:

$$\max_{\langle c\rangle, \langle p\rangle} \sum_{t=\underline{t}+1}^{n} [-p_t + k\mathcal{A}_t c_t]$$
$$s.t.$$
$$c_t \leq 1 \ \forall t > \underline{t};$$
$$p_t \geq 0 \ \forall t > \underline{t};$$
$$v_t \mathcal{A}_{\hat{t}} c_{\hat{t}} - p_{\hat{t}} + p_t \leq v_t \mathcal{A}_t \quad \forall t, \hat{t} \in \{\underline{t}, \ldots n\}, \hat{t} > t$$

Consider the dual of this LP problem, letting $y_{i,j} \geq 0$ denote the coefficient of i pretending to be j. Similarly let $r_j \geq 0$ denote the coefficient for the constraint $c_j \leq 1$. The coefficients are all weakly positive since the constraints are all required to be less than equal to their respective bounds. Since our constraint matrix is relatively simple we will directly interpret the inequality constraints instead of introducing and keeping track of slack variables. This also means that we do not need to introduce a separate coefficient for non-negativity of p_t. We remark that Vohra (2004) chapter 4 is an excellent and compact reference to keep track of how weak primal constraints translate to dual constraints. The dual is:

$$\min_{\langle y_{i,j}\rangle, \langle r\rangle} \sum_{i \in \{\underline{t}, \ldots n\}} \sum_{j > i} y_{i,j} v_i \mathcal{A}_i + \sum_{j \in \{\underline{t}+1, \ldots, n\}} r_j$$
$$\sum_{i=\underline{t}}^{j} v_i \mathcal{A}_j y_{i,j} + r_j \geq k\mathcal{A}_j, \quad \forall j \in \{\underline{t}+1, \ldots, n\}$$
$$-\sum_{i=\underline{t}}^{j} y_{i,j} + \sum_{l=j+1}^{n} y_{j,l} \geq -1 \quad \forall j \in \{\underline{t}+1, \ldots, n\}$$

Towards a contradiction, assume $p_{\underline{t}+1} = 0$ that is the non-negativity constraint is binding. By complementary slackness the constraint $-y_{\underline{t},\underline{t}+1} + \sum_{l=\underline{t}+2}^{n} y_{\underline{t}+1,l} \geq -1$ is slack (see Vohra (2004), chapter 4, pp 64). But then we must have $v_{\underline{t}} \mathcal{A}_{\underline{t}+1} y_{\underline{t},\underline{t}+1} = k\mathcal{A}_{\underline{t}+1}$ and $r_{\underline{t}+1} = 0$ as otherwise we would have $c_{\underline{t}+1} = 1$, which contradicts $\underline{t}$ being the cutoff. This in turn implies $\sum_{l=\underline{t}+2}^{n} y_{\underline{t}+1,l} > \frac{k}{v_{\underline{t}}} - 1 > 0$. Which further implies there is at least one l with $y_{\underline{t}+1,l} > 0$. Since $-y_{\underline{t},\underline{t}+1} + \sum_{l=\underline{t}+2}^{n} y_{\underline{t}+1,l} > -1$ we can reduce $y_{\underline{t}+1,l}$ by $\varepsilon > 0$ that is small enough and increase $y_{\underline{t},l}$ by $\delta = \frac{\varepsilon v_{\underline{t}+1}}{v_{\underline{t}}}$. Then the dual decreases by $\varepsilon v_{\underline{t}+1}(\mathcal{A}_{\underline{t}+1} - \frac{\mathcal{A}_{\underline{t}}}{v_{\underline{t}}}) > 0$ contradicting $p_{\underline{t}+1} = 0$ minimizing the dual (equivalently maximizing the primal). Hence we must have $-y_{\underline{t},\underline{t}+1} + \sum_{l=\underline{t}+2}^{n} y_{\underline{t}+1,l} = -1$, which by complementary slackness implies $p_{\underline{t}+1} > 0$.

To see $c_{\underline{t}+1} = 1$, for a contradiction assume $0 < c_{\underline{t}+1} < 1$ which implies $r_{\underline{t}+1} = 0$. But then we must have $v_{\underline{t}}\mathcal{A}_{\underline{t}+1}y_{\underline{t},\underline{t}+1} = k\mathcal{A}_{\underline{t}+1}$ as otherwise complementary slackness would imply $c_{\underline{t}+1} = 0$, which in turn would imply that there can be no types that want to mimic $\underline{t}+1$ (since all claims are checked with certainty). But then increasing $c_{\underline{t}+1}$ by a small ε would be a strict improvement of the objective so we must have $v_{\underline{t}}\mathcal{A}_{\underline{t}+1}y_{\underline{t},\underline{t}+1} = k\mathcal{A}_{\underline{t}+1}$. Consider decreasing $y_{\underline{t},\underline{t}+1} = \frac{k}{v_{\underline{t}}} > 1$ by $\epsilon > 0$ such that $y_{\underline{t},\underline{t}+1} - \epsilon > 1$ and setting $r_{\underline{t}+1} = \epsilon v_{\underline{t}}\mathcal{A}_{\underline{t}+1}$. Then the constraint regarding $p_{\underline{t}+1}$, $-y_{\underline{t},\underline{t}+1} + \sum_{l=\underline{t}+2}^{n} y_{\underline{t}+1,l} = -1$ is now relaxed but the objective in the dual is unchanged. Thus we can find at least one l with $y_{\underline{t}+1,l} > 0$ reduce $y_{\underline{t}+1,l}$ by ε and increase $y_{\underline{t},l}$ by $\delta = \frac{\varepsilon v_{\underline{t}+1}}{v_{\underline{t}}}$. Then the dual decreases by by $\varepsilon v_{\underline{t}+1}(\mathcal{A}_{\underline{t}+1} - \frac{\mathcal{A}_{\underline{t}}}{v_{\underline{t}}}) > 0$ contradicting the optimality of $c_{\underline{t}+1} < 1$.

To see that $p_{\underline{t}+1} = v_{\underline{t}}(\mathcal{A}_{\underline{t}+1} - \mathcal{A}_{\underline{t}})$, observe that the constraint regarding $\underline{t}$ mimicking $\underline{t}+1$ must bind as there has to be at least one type that wants to mimic $\underline{t}+1$, but since $c_{\underline{t}+1} = 1$ we must have $p_{\underline{t}+1} = v_{\underline{t}}(\mathcal{A}_{\underline{t}+1} - \mathcal{A}_{\underline{t}})$. □

Inductive Step: Let $\underline{t}$ be the cutoff and let $v_{t'} < k$ and $t' > \underline{t}$. By the inductive hypothesis we have $p_{t'-l} - p_{t'-l-1} = v_{t'-l-1}(\mathcal{A}_{t'-l} - \mathcal{A}_{t'-l-1})$ and $c_{t'-l} = 1$ for all $t' - \underline{t} > l \geq 0$. We need to show $p_{t'+1} - p_{t'} = v_{t'}(\mathcal{A}_{t'+1} - \mathcal{A}_{t'})$ and $c_{t'+1} = 1$.

Proof of Inductive Step. By the inductive hypothesis the money burning and verification probabilities of all types including type t' is known and fixed. Similar to the basis step we can consider the LP that pins down the rest of the money burning and verification for the optimal $\mathcal{A}$. The problem is given by

$$\max_{\langle c\rangle,\langle p\rangle} \sum_{t=t'+1}^{n} [-p_t + k\mathcal{A}_t c_t]$$

$$s.t.$$

$$c_t \leq 1 \ \forall t > t';$$

$$p_t \geq 0 \ \forall t > \underline{t};$$

$$v_t\mathcal{A}_{\hat{t}}c_{\hat{t}} - p_{\hat{t}} + p_t \leq v_t\mathcal{A}_t \quad \forall \hat{t} \in \{t', \dots n\}, t \in \{\underline{t}, \dots n\}, \hat{t} > t$$

The only difference between this LP and the one in the basis step is that we need to keep track of the incentive constraints regarding types below t' mimicking types above t' as well, since a choice of money burning and verification probability for some type $t > t'$ still needs to satisfy those constraints. Similar to the basis step let us consider the dual letting $y_{i,j} \geq 0$ denote the coefficient of i pretending to be j. Similarly let $r_j \geq 0$ denote the coefficient for the constraint $c_j \leq 1$. The dual is:

$$\min_{\langle y_{i,j}\rangle, \langle r \rangle} \sum_{i \in \{\underline{t}, \ldots n\}} \sum_{j \in \{t'+1, \ldots, n\}, j > i} y_{i,j} v_i \mathcal{A}_i + \sum_{j \in \{t'+1, , \ldots n\}} r_j$$

$$\sum_{i=\underline{t}}^{j} v_i \mathcal{A}_j y_{i,j} + r_j \geq k\mathcal{A}_j, \quad \forall j \in \{t'+1, \ldots, n\}$$

$$-\sum_{i=\underline{t}}^{j} y_{i,j} + \sum_{l=j+1}^{n} y_{j,l} \geq -1 \quad \forall j \in \{t'+1, \ldots, n\}$$

Towards a contradiction, assume $p_{t'+1} = 0$ that is the non-negativity constraint is binding. Similar to the basis step complementary slackness implies the constraint $-\sum_{i=\underline{t}}^{t'} y_{i,t'+1} + \sum_{l=t'+2}^{n} y_{t'+1,l} \geq -1$ is slack. But then we must have $\sum_{i=\underline{t}}^{t'} v_i \mathcal{A}_{t'+1} y_{i,t'+1} = k\mathcal{A}_{t'+1}$ and $r_{t'+1} = 0$ as otherwise we would have $c_{t'+1} = 1$, which contradicts t' being above the cutoff. We can cancel the $A_{t'+1}$'s and reorganize the left handside term $\sum_{i=\underline{t}}^{t'} v_i y_{i,t'+1} = \sum_{i=\underline{t}}^{t'} v_{t'} y_{i,t'+1} - \sum_{i=\underline{t}}^{t'} (v_{t'} - v_i) y_{i,t'+1} = k$. Since all $(v_{t'} - v_i) y_{i,t'+1}$'s are positive we thus must have $\sum_{i=\underline{t}}^{t'} y_{i,t'+1} > \frac{k}{v_{t'}} > 1$. This in turn implies $\sum_{l=t'+2}^{n} y_{t'+1,l} > 0$. Which further implies there is at least one l with $y_{t'+1,l} > 0$. Since $-\sum_{i=\underline{t}}^{t'} y_{i,t'+1} + \sum_{l=t'+2}^{n} y_{t'+1,l} > -1$ we can reduce $y_{t'+1,l}$ by $\varepsilon > 0$ that is small enough and increase $y_{t,l}$ by $\delta = \frac{\varepsilon v_{t'+1}}{v_{t'}}$. Then the dual decreases by $\varepsilon v_{t'+1}(\mathcal{A}_{t'+1} - \frac{\mathcal{A}_{t'}}{v_{t'}}) > 0$ contradicting $p_{t'+1} = 0$ minimizing the dual (equivalently maximizing the primal). Hence we must have $p_{t'+1} > 0$.

To see $c_{t'+1} = 1$, for a contradiction assume $0 < c_{t'+1} < 1$ which implies $r_{t'+1} = 0$. But then we must have $\sum_{i=\underline{t}}^{t'} v_i \mathcal{A}_{t'+1} y_{i,t'+1} = k\mathcal{A}_{t'+1}$ as otherwise complementary slackness would imply $c_{t'+1} = 0$ which cannot be optimal similar to the basis case. Consider decreasing $y_{\underline{t},t'+1}$ by $\epsilon > 0$ such that $\sum_{i=\underline{t}}^{t'} y_{i,t'+1} - \epsilon > 1$ and setting $r_{t'+1} = \epsilon v_{\underline{t}} \mathcal{A}_{t'+1}$. Then similar to the basis step the constraint regarding $p_{t'+1}$, is now relaxed but the objective in the dual is unchanged. Thus similar to above we can find at least one l with $y_{t'+1,l} > 0$ reduce $y_{t'+1,l}$ by $\varepsilon > 0$ that is small enough and increase $y_{t,l}$ by $\delta = \frac{\varepsilon v_{t'+1}}{v_{t'}}$. Then the dual decreases by $\varepsilon v_{t'+1}(\mathcal{A}_{t'+1} - \frac{\mathcal{A}_{t'}}{v_{t'}}) > 0$ contradicting the optimality of $c_{t'+1} < 1$.

Finally, to see that $p_{t'+1} = v_{t'}(\mathcal{A}_{t'+1} - \mathcal{A}_{t'}) + p_{t'}$ observe that again there has to be at least one type that wants to mimic $t'+1$. For a contradiction suppose some type l, $\underline{t} \leq l < t'$ has a binding incentive constraint regarding $t'+1$. Then we must have $p_{t'+1} = v_l(\mathcal{A}_{t'+1} - \mathcal{A}_l) + p_l$ and p_l is fixed by the inductive hypothesis. Let us now consider the type $l+1$. By the inductive hypothesis we also must have $p_{l+1} = v_l(\mathcal{A}_{l+1} - \mathcal{A}_l) + p_l$. But then, incentive constraint for type $l+1$ regarding $t'+1$ would imply

$$\begin{aligned}
v_{l+1}\mathcal{A}_{l+1} - p_{l+1} &\geq v_{l+1}\mathcal{A}_{t'+1} - p_{t+1}, \\
v_{l+1}\mathcal{A}_{l+1} - p_{l+1} &\geq v_{l+1}\mathcal{A}_{t'+1} - v_l(\mathcal{A}_{t'+1} - \mathcal{A}) - p_l, \\
v_l(\mathcal{A}_{t'+1} - \mathcal{A}_l) &\geq v_{l+1}(\mathcal{A}_{t'+1} - \mathcal{A}_{l+1}) + p_{l+1} - p_l, \\
v_l(\mathcal{A}_{t'+1} - \mathcal{A}_l) &\geq v_{l+1}(\mathcal{A}_{t'+1} - \mathcal{A}_{l+1}) + v_l(\mathcal{A}_{l+1} - \mathcal{A}_l), \\
v_l(\mathcal{A}_{t'+1} - \mathcal{A}_{l+1}) &\geq v_{l+1}(\mathcal{A}_{t'+1} - \mathcal{A}_{l+1});
\end{aligned}$$

Now, if $v_l < v_{l+1}$ then we reach the desired contradiction as the IC of type v_{l+1} will be violated. If not, then we must have $v_l = v_{l+1}$, and v_{l+1} also has a binding constraint $p_{t'+1} = v_{l+1}(\mathcal{A}_{t'+1} - \mathcal{A}_{l+1}) + p_{l+1}$. If $l+1 = t'$ then we are done. If not we consider the IC for type $l+2$. Again, either we would have a violation of IC for type $l+2$ if $v_{l+1} < v_{l+2}$ or we must have $v_{l+2} = v_{l+1} = v_l$ with $l+2$ having a binding constraint. Proceeding inductively we either will reach a contradiction or all types starting from l has the same value until t' and the IC of t' gives the desired identity $p_{t'+1} = v_{t'}(\mathcal{A}_{t'+1} - \mathcal{A}_{t'}) + p_{t'}$. Notice, in the case that all types have the same value, that is $v_l = v_{l+1} \ldots v_{t'}$ then we have $\{(l, t'+1), (l+1, t'+1), \ldots (t', t'+1)\} \subset E(G_M)$. □

□

Lemma 11. *Let $\underline{t}$ be the cutoff and let $t^{EF} > \underline{t}$. Then for all t such hat $v_t > k$, $c_t < 1$.*

Proof. Let $\mathcal{A}$'s be fixed for a given cutoff $\underline{t}$. And by proposition 6 we know that for all $\hat{t} \leq t^{EF}$, $c_{\hat{t}} = 1$ and $p_{\hat{t}} = v_{\hat{t}-1}(\mathcal{A}_{\hat{t}} - \mathcal{A}_{\hat{t}-1}) + p_{\hat{t}-1}$ starting with $p_{\underline{t}} = 0$. Then the principal's problem can be written as a standard LP

$$\begin{aligned}
&\max_{\langle c\rangle, \langle p\rangle} \sum_{t=t^k+1}^{n} [-p_t + k\mathcal{A}_t c_t] \\
&s.t. \\
&c_t \leq 1 \; \forall t \in \{t^{EF}+1, \ldots, n\}; \; c_t, p_t \in \mathbb{R}^+ \\
&v_{\hat{t}}\mathcal{A}_{\hat{t}} \geq v_{\hat{t}}\mathcal{A}_t c_t - p_t + p_{\hat{t}} \quad \forall t \in \{t^{EF}+1, \ldots, n\}, \hat{t} \in \{\underline{t}, \ldots, n\}, \hat{t} < t
\end{aligned}$$

Notice in this formulation we still need to account for lower types (whose c and p is fixed) potentially mimicking higher types. The lemma concerns c_t's so instead of directly using Simplex method, we apply the upper bounding technique detailed in Eiselt and Sandblom (2007) (pp 212-218). That is we first seperate the upper bounds on c_t, and consider the constraint matrix, defined by just the IC constraints. The constraint matrix including the slack variables for the IC constraints has the following representation

$$\left[\begin{array}{cccccccccc|c} \mathrm{V}_{t^{EF}+1} & 0 & \cdots & 0 & 0 & -\mathbf{1}^{t^{EF}+1-\underline{t}+1} & 0 & \cdots & 0 & 0 & \\ 0 & \mathrm{V}_{t^{EF}+2} & \cdots & 0 & 0 & e_1^{t'+2-\underline{t}+1} & -\mathbf{1}^{t^{EF}+2-\underline{t}+1} & \cdots & 0 & 0 & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & I \\ 0 & 0 & \cdots & \mathrm{V}_{n-1} & 0 & e_1^{n-1-\underline{t}+1} & e_2^{n-1-\underline{t}+1} & \cdots & -\mathbf{1}^{n-1-\underline{t}+1} & 0 & \\ 0 & 0 & \cdots & 0 & \mathrm{V}_n & e_1^{n-\underline{t}+1} & e_2^{n-\underline{t}+1} & \cdots & e_{n-\underline{t}+1}^{n-\underline{t}+1} & -\mathbf{1}^{n-\underline{t}+1} & \end{array}\right]$$

V_j is the column vector of size $j-\underline{t}+1$ with entries $(v_{\underline{t}}\mathcal{A}_j, v_{\underline{t}+1}\mathcal{A}_j, \ldots, v_{j-1}\mathcal{A}_j)$. $-\mathbf{1}^j$ is the column vector of size j with all entries -1 and e_l^j is the unit column vector of size j, that is, it has zero in every entry except the lth entry which has a value 1. Finally I denotes the identity matrix of size $\sum_{r=t'+1}^{n}(r-\underline{t}+1)$. In words each column vector V_j corresponds to all the types that might potentially want to mimic type j, which are all the types starting from $\underline{t}$ and ending in $j-1$. Any type that wants to mimic type j would need to pay p_j hence there are $j-\underline{t}+1$ constraints which has p_j with coefficient -1. If type $j > t'$ wants to mimic other, higher types, then we account for those with the entries e since p_j is still a choice variable. For types less than or equal to t' the p's are fixed by proposition 6. For each j we construct the following column vector $\nu_j = (v_{\underline{t}}\mathcal{A}_t, \ldots v_{t^{EF}}\mathcal{A}_{t^{EF}} - p_{t^{EF}}, v_{t^{EF}+1}\mathcal{A}_{t^{EF}+1}, \ldots, v_{j-1}\mathcal{A}_{j-1})$. And the column vector of size $\sum_{r=t^{EF}+1}^{n}(r-\underline{t}+1)$, $\nu = (\nu_{t^{EF}+1}, \ldots, \nu_n)$ which corresponds to the RHS of all the IC's.

Given that the p's have a negative value in the objective, we need to start pivoting from any of the c variables first. Suppose we start from the variable c_j, since all the entries in the column $(0, \ldots, 0, \mathrm{V}_j, 0, \ldots, 0)$ is weakly positive and all the entries in the vector ν are positive upper bounding yields $v_{\underline{t}\mathcal{A}_j}$ as the pivot variable. On the other hand if we applied standard simplex for our constraint matrix will have $n-\underline{t}+1$ additional rows and columns, rows corresponding to the upper bound constraints for c variables, and columns corresponding to their associated slack variables. In particular since the objective hasn't changed we would start pivoting from any of the c variables first again. Suppose we start from the column corresponding to the variable c_j, the associated column is now longer by $n-\underline{t}+1$ entries where all the new entries are 0 except one 1 corresponding to the equation $c_j + s_{cj} = 1$ where s_{cj} is the slack variable introduced to capture the constraint $c_j \leq 1$. That is the column will now be $(0 \ldots 0, 1, 0 \ldots 0; 0, \ldots, 0, \mathrm{V}_j, 0, \ldots, 0)$. However, similar to before the pivot variable would be $v_{\underline{t}}\mathcal{A}_j$ since the quotient for the new constraint is just 1 and we have $\mathcal{A}_{\underline{t}}/\mathcal{A}_j < 1$. But since both algorithms pivot the same way with the same quotient for the departing variable c_j, the constraint $c_j \leq 1$ cannot be binding. Proceeding with the rest of the c variables in an analogous manner, we conclude for all $t > t^{EF}$, $c_t < 1$. $\square$

Proof of Proposition 7. Suppose $\mathcal{A}$'s are all given and fixed for a given cutoff $\underline{v}$. And by proposition 6 we know that for all $\hat{t} \leq t^{EF}$, $c_{\hat{t}} = 1$ and $p_{\hat{t}} = v_{\hat{t}-1}(\mathcal{A}_{\hat{t}} - \mathcal{A}_{\hat{t}-1}) + p_{\hat{t}-1}$ starting with $p_{\underline{t}} = 0$. Finally by Lemma 11 we know that for all $t > t^{EF}$, $c_t < 1$.

We first show $p_{t^{EF}+1} = p_{t^{EF}}$. For a contradiction, suppose not. There are two cases to consider

Case I: $p_{t^{EF}+1} < p_{t^{EF}}$

Recall that $\mathcal{A}_t c_t = \mathcal{A}_t$ for all $t \leq t^{EF}$ by Lemma 4. Proposition 2 implies that the only types that can have a binding constraint regarding type $t^{EF}+1$ are the types $\hat{t}$ such that $p_{\hat{t}} \leq p_{t^{EF}+1}$. First, suppose that multiple such types with binding constraints exist and let $\tilde{t}$ be the minimal type. From the IC of type $\tilde{t}$ we have

$$v_{\tilde{t}}(\mathcal{A}_{t^{EF}+1}c_{t^{EF}+1} - \mathcal{A}_{\tilde{t}}) = p_{t^{EF}+1} - p_{\tilde{t}}$$

On the other hand by Proposition 6 we also have

$$v_{\tilde{t}}\mathcal{A}_{\tilde{t}} - p_{\tilde{t}} = v_{\tilde{t}}\mathcal{A}_{\tilde{t}+1} - p_{\tilde{t}+1}$$

which together imply that

$$v_{\tilde{t}+1}(\mathcal{A}_{t^{EF}+1}c_{t^{EF}+1} - \mathcal{A}_{\tilde{t}}) > p_{t^{EF}+1} - p_{\tilde{t}}$$

A violation of the IC constraint for type $\tilde{t}+1$. A similar violation also occurs if there is a single type $\tilde{t}$ but $p_{\tilde{t}+1} \leq p_{t^{EF}+1}$. Thus, the only plausible case is if the only type that is mimicking type $t^{EF}+1$ is type $t^{EF}-1$ so that we have $p_{t^{EF}-1} \leq p_{t^{EF}+1} < p_{t^{EF}}$. But then consider increasing $p_{t^{EF}+1}$ by δ and $c_{t^{EF}+1}$ by ϵ where $\mathcal{A}_{t^{EF}+1}\epsilon v_{t^{EF}-1} = \delta$. Such an increase means $t^{EF}-1$ is still indifferent whereas all the other constraints remain unaffected, but the increase in the objective is $\epsilon\mathcal{A}_{t^{EF}+1}(k - v_{t^{EF}-1}) > 0$, a contradiction.

Case II: $p_{t^{EF}+1} > p_{t^{EF}}$

In this case first suppose some type $\tilde{t} < t^{EF}$ has a binding constraint with type $t^{EF}+1$. Then similar to above we would have

$$v_{\tilde{t}}(\mathcal{A}_{t^{EF}+1}c_{t^{EF}+1} - \mathcal{A}_{\tilde{t}}) = p_{t^{EF}+1} - p_{\tilde{t}} \quad \text{and; } v_{\tilde{t}}\mathcal{A}_{\tilde{t}} - p_{\tilde{t}} = v_{\tilde{t}}\mathcal{A}_{\tilde{t}+1} - p_{\tilde{t}+1}$$

which together imply that

$$v_{\tilde{t}+1}(\mathcal{A}_{t^{EF}+1}c_{t^{EF}+1} - \mathcal{A}_{\tilde{t}}) > p_{t^{EF}+1} - p_{\tilde{t}}$$

A violation of the IC constraint for type $\tilde{t}+1$. Therefore the only types that can mimic type $t^{EF}+1$ are types above t^{EF}. But then consider decreasing $p_{t^{EF}+1}$ by δ and $c_{t^{EF}+1}$ by ϵ where $\mathcal{A}_{t^{EF}+1}\epsilon v_{t^{EF}} = \delta$. Then any IC pertaining to type $t^{EF}+1$ from any type weakly above t^{EF} will continue to hold, potentially strictly. The increase in the objective is $\epsilon \mathcal{A}_{t^{EF}+1}(v_{t^{EF}} - k) > 0$, a contradiction.

Proceeding inductively with identical cases we must have $p_{t^{EF}+2} = p_{t^{EF}}$, which in turn implies $p_{t^{EF}+3} = p_{t^{EF}}$ and so on until we have $p_n = p_{n-1} \ldots = p_{t^{EF}}$. Since the payments are the same by Lemma 6 we have

$$\mathcal{A}_n c_n = \mathcal{A}_{n-1} c_{n-1} = \ldots = \mathcal{A}_{t^{EF}+1} c_{t^{EF}+1} = \mathcal{A}_{t^{EF}}.$$

□

Proof of Proposition 8. If $v_{\underline{t}} > k$, then Proposition 5 implies there is no money burning at all in the mechanism and the result is the one that coincides with Vohra (2012) which satisfies the proposition. Consider the case where $v_{\underline{t}} < k$. As described in the main text the polymatroid problem is:

$$\begin{aligned} &\max_{\langle \mathcal{B} \rangle} \sum_{t=1}^{n} v_t(\mathcal{B}_t + d_t) \\ &\text{s.t.} \sum_{s \in S} \mathcal{B}_s \leq \rho(S) \quad \forall S \subset V \end{aligned}$$

Since g is submodular and $d_i's$ are increasing, the ρ function for any set $\{n, n-1, n-2, \ldots t\}$ with $t \geq \underline{t}+1$ must be of the form $\rho(\{n, n-1, \ldots, t\}) = g(n-r+1) - \sum_{i=r}^{n} d_i$ for some $t \geq r \geq \underline{t}$. We also know that if i, j are duplicates types then we must have $d_i = d_j$. The only way to "skip" some terms is by skipping a duplicate type and replacing it with another. However, recall that g is submodular. Since every duplicate has the same d_i a minimizing superset must contain all duplicates together. It is not possible to consider subsets that "skip" non-duplicate terms t and r as the term $\sum_{i=t}^{n} d_t$ is maximized by choosing consecutive elements due to weakly increasing d_i's. If $\rho(\{n, n-1, \ldots, t\}) = g(n-t+1) - \sum_{i=t}^{n} d_i$ for all sets of the form $\{n, n-1, n-2, \ldots t\}$ then the greedy algorithm proceeds downward from the highest value $\mathcal{V}$. That is $\mathcal{B}_n = \mathcal{B}_{n-1} = \mathcal{B}_{n-f(\mathcal{V})} = \frac{g(f(\mathcal{V}))}{f(\mathcal{V})} - f(\mathcal{V}) d_n$ which results in $\mathcal{A}_n = \ldots = \mathcal{A}_{n-f(\mathcal{V})} = \frac{g(f(\mathcal{V}))}{f(\mathcal{V})}$. Proceeding inductively from the top we get that $\mathcal{A}_{n-f(\mathcal{V})-1} = \ldots = \mathcal{A}_{n-f(\mathcal{V})-f(\mathcal{V}-1)} = \frac{g(f(\mathcal{V})+f(\mathcal{V}-1)) - g(f(\mathcal{V}))}{f(\mathcal{V}-1)}$ and so on resulting in the claim of the proposition with types below cutoff sharing the leftover allocation.

Towards a contradiction, suppose there exists some $\tilde{t}$ (WLOG assume $\tilde{t}$ is the lowest index among the duplicates) such that the set $\rho(\{n, n-1, \ldots, \tilde{t}\}) = g(n-r+1) - \sum_{i=r}^{n} d_i < g(n-\tilde{t}+1) - \sum_{i=\tilde{t}}^{n} d_i$ for some $\tilde{t} > r \geq \underline{t}$. Then the greedy algorithm necessarily assigns $\mathcal{B}_{\tilde{t}} = g(n-r+1) - g(n-\tilde{t}+1) - \sum_{j=r}^{\tilde{t}} d_j > d_{\tilde{t}}$, and sets $\mathcal{B}_{t'} = 0$ for any t' such that $\tilde{t} > t' \geq r$, hence those types are at their constraint level. The first inequality is strict since $\rho(\{n, n-1, \ldots, \tilde{t}\}) = g(n-r+1) - \sum_{i=r}^{n} d_i < g(n-\tilde{t}+1) - \sum_{i=\tilde{t}}^{n} d_i$. But $\mathcal{B}_{\tilde{t}} > d_{\tilde{t}}$

implies that type $\tilde{t}$ does not want to mimic any type, their allocated payoff is strictly larger than even the best type of lie they can tell. This in turn means that $\tilde{t}$ cannot be subject to only money burning since Proposition 6 implies that these types have a type that they do want to mimic. Thus we must have $o_M(\tilde{t}) = \emptyset$, and $c_{\tilde{t}} < 1$. Furthermore, the resulting $\mathcal{A}_{\tilde{t}} = \mathcal{B}_{\tilde{t}} + d_{\tilde{t}} = g(n-r+1) - g(n-\tilde{t}+1) - \sum_{i=r}^{\tilde{t}-1} d_i < g(n-\tilde{t}) - g(n-\tilde{t}+1)$, that is the constraints arising from the lower types necessitates some withholding of type $\tilde{t}$. Let's construct an improvement. By proposition 6 in $\mathcal{A}c, p$ we must have $p_{\underline{t}+1} > 0$ and $c_{\underline{t}+1} = 1$ so $\tilde{t} > \underline{t}+1$. Consider $M^* = T_{M,\underline{t}+1,t}(\mathcal{A}^*_{\underline{t}+1}, \mathcal{A}^*_{\tilde{t}}, c^*_{\underline{t}+1}, c^*_{\tilde{t}}, p^*_{\underline{t}+1}, p^*_{\tilde{t}})$ where,

$$\begin{aligned}
\mathcal{A}^*_{\underline{t}+1} &= \mathcal{A}_{\underline{t}+1} - \varepsilon;\ c^*_{\underline{t}+1} = c_{\underline{t}} - \varepsilon \left(\frac{v_{\tilde{t}}(\mathcal{A}_{\tilde{t}} c_{\tilde{t}} - \mathcal{A}_{\underline{t}+1} c_{\underline{t}+1}) - c_{\underline{t}+1}(p_{\tilde{t}} - p_{\underline{t}+1})}{(p_{\tilde{t}} - p_{\underline{t}+1})(\mathcal{A}_{\underline{t}+1} - \varepsilon)} \right);\\
p^*_{\underline{t}+1} &= p_{\underline{t}+1} - v_{\tilde{t}}\varepsilon\\
\mathcal{A}^*_{\tilde{t}} &= \mathcal{A}_{\tilde{t}} + \varepsilon;\ c^*_{\tilde{t}} = c_{\tilde{t}} + \varepsilon \left(\frac{v_{\tilde{t}}(\mathcal{A}_{\tilde{t}} c_{\tilde{t}} - \mathcal{A}_{\underline{t}+1} c_{\underline{t}+1}) - c_t(p_t - p_{\underline{t}+1})}{(p_t - p_{\underline{t}+1})(\mathcal{A}_t + \varepsilon)} \right);\\
p^*_{\tilde{t}} &= p_{\tilde{t}} + v_{\tilde{t}}\varepsilon
\end{aligned}$$

Improving $\mathcal{A}_{\tilde{t}}$ this way is feasible since $\mathcal{A}_{\tilde{t}} < g(N-\tilde{t}) - \sum_{j=\tilde{t}+1}^{n} \mathcal{A}_j$ so there is slack for small enough ε. Thus, noting $c_{\tilde{t}} < 1$ and replicating the steps 1-2-3 of the proof of proposition 3 shows that M^* is feasible, IC and an improvement. □

Proof of Theorem 1. The proof of theorem simply follows from the proof of the two Lemmas about the discrete derivative with regards to optimal cutoffs with and without money burning, which we will prove below. Let $\Pi(\underline{v})$ denote the total payoff from the mechanism when the cutoff type has value $\underline{v}$ and the associated type is $\underline{t}$. Let us introduce the notation $\bar{\mathcal{A}}_t$ to denote the allocation probability of the type t when $\underline{t}$ is the cutoff (analogously $\bar{\mathcal{A}}^{\underline{v}}$ for any type below cutoff) and we use $\mathcal{A}_t$ to denote the allocation probability of type t when the cutoff is below t (analogously $\mathcal{A}^v$ for any valuation above the cutoff). Recall $h(\underline{v}) = \frac{f(\underline{v})}{\sum_{i=\underline{v}+1}^{V} f(i)}$ is the hazard rate.

Proof of Lemma A. Let us consider a change of the cutoff from $\underline{v}$ to $\underline{v}+1$. The only change to the types above $\underline{v}+2$ is their money burning requirements since they were already getting the greedy allocations and verification probabilities don't change. For the types below and including $\underline{v}+1$ the allocation probabilities change, and type $\underline{v}+1$ is no longer required to burn money. The change in money burning for $\underline{v}+2$ and all the types above is given by:

$$\Delta p_{\underline{v}+2} = (\underline{v}+1)(\mathcal{A}^{\underline{v}+1} - \bar{\mathcal{A}}^{\underline{v}+1}) - \underline{v}(\mathcal{A}^{\underline{v}+1} - \bar{\mathcal{A}}^{\underline{v}})$$

The money burning that vanishes from $\underline{v}+1$ is $p_{\underline{v}+1} = \underline{v}(\mathcal{A}^{\underline{v}+1} - \bar{\mathcal{A}}^{\underline{v}})$.Therefore the total change in money burning is equal to

$$(\mathcal{A}^{\underline{v}+1} - \bar{\mathcal{A}}^{\underline{v}})\underline{v} \sum_{i=\underline{v}+1}^{\mathcal{V}} f(i) - (\mathcal{A}^{\underline{v}+1} - \bar{\mathcal{A}}^{\underline{v}+1})(\underline{v}+1) \sum_{i=\underline{v}+2}^{\mathcal{V}} f(i)$$

The change in allocation can be calculated again using the discrete product rule to reduce to

$$\left(\sum_{i=1}^{\underline{v}} i f(i) (\bar{\mathcal{A}}^{\underline{v}+1} - \bar{\mathcal{A}}^{\underline{v}}) \right) - (\mathcal{A}^{\underline{v}+1} - \bar{\mathcal{A}}^{\underline{v}+1})(\underline{v}+1) f(\underline{v}+1)$$

Therefore the total change in payoff $F(\underline{v}+1) - F(\underline{v})$ is given by

$$\begin{aligned}
\Pi(\underline{v}+1) - \Pi(\underline{v}) = & \left(\sum_{i=1}^{\underline{v}} i f(i) (\bar{\mathcal{A}}^{\underline{v}+1} - \bar{\mathcal{A}}^{\underline{v}}) \right) - (\mathcal{A}^{\underline{v}+1} - \bar{\mathcal{A}}^{\underline{v}+1})(\underline{v}+1) f(\underline{v}+1) \\
& + (\mathcal{A}^{\underline{v}+1} - \bar{\mathcal{A}}^{\underline{v}})\underline{v} \sum_{i=\underline{v}+1}^{\mathcal{V}} f(i) \\
& - (\mathcal{A}^{\underline{v}+1} - \bar{\mathcal{A}}^{\underline{v}+1})(\underline{v}+1) \sum_{i=\underline{v}+2}^{\mathcal{V}} f(i)
\end{aligned}$$

We are only interested in the sign of this derivative thus we can divide the entire equation by $(\bar{\mathcal{A}}^{\underline{v}+1} - \bar{\mathcal{A}}^{\underline{v}})$ and simplify to get

$$\begin{aligned}
\Pi(\underline{v}+1) - \Pi(\underline{v}) \propto & \sum_{i=1}^{\underline{v}} i f(i) - \frac{\sum_{i=1}^{\underline{v}} f(i)}{f(\underline{v}+1)} (\underline{v}+1) f(\underline{v}+1) \\
& + \underline{v} \frac{\sum_{i=1}^{\underline{v}+1} f(i)}{f(\underline{v}+1)} \sum_{i=\underline{v}+1}^{\mathcal{V}} f(i) \\
& - (\underline{v}+1) \sum_{i=\underline{v}+2}^{\mathcal{V}} f(i) \frac{\sum_{i=1}^{\underline{v}} f(i)}{f(\underline{v}+1)}
\end{aligned}$$

Reorganizing the terms above we get

$$\Pi(\underline{v}+1) - \Pi(\underline{v}) \propto \sum_{i=1}^{\underline{v}} f(i) \left(i - (\underline{v}+1) \frac{\sum_{i=\underline{v}+1}^{\mathcal{V}} f(i)}{f(\underline{v}+1)} + \underline{v} \frac{\sum_{i=\underline{v}+1}^{\mathcal{V}} f(i)}{f(\underline{v}+1)} \right)$$

Simplifying further and recalling that $\frac{\sum_{i=\underline{v}+1}^{\mathcal{V}} f(i)}{f(\underline{v}+1)}$ is the inverse hazard rate we have

$$\Pi(\underline{v}+1) - \Pi(\underline{v}) \propto \sum_{i=1}^{\underline{v}} f(i)(i - \frac{1}{h(\underline{v}+1)})$$

Recall that under monotone hazard rate assumption the inverse hazard rate $\frac{1}{h(\underline{v}+1)}$ is non-increasing. Thus the derivative is increasing. This in turn implies that if it is ever worth it to increase the cutoff with money burning, then it is also worth it to increase it all the way till $k-1$. Therefore we now consider the case of increasing the cutoff from $k-1$ to k, which would actually remove money burning from the mechanism altogether. If the cutoff is k the payoff is

$$\Pi(k) = \bar{\mathcal{A}}_k \sum_{i=1}^{k} if(i) + \sum_{i=k+1}^{\mathcal{V}} (i-k)\mathcal{A}_i f(i) + k\bar{\mathcal{A}}_k \sum_{i=k+1}^{\mathcal{V}} f(i)$$

If the cutoff is $k-1$ the payments will be equal to $(k-1)\sum_{i=k}^{\mathcal{V}} f(i)(\mathcal{A}_k - \bar{\mathcal{A}}_{k-1})$. Type k is still not subject to verification, but the verification probability of all the types above it changes. Thus the total payoff is

$$\Pi(k-1) = \bar{\mathcal{A}}_{k-1} \sum_{i=1}^{k-1} if(i) + \sum_{i=k+1}^{\mathcal{V}} (i-k)\mathcal{A}_i f(i) + k\mathcal{A}_k \sum_{i=k+1}^{\mathcal{V}} f(i)$$
$$-(k-1)\sum_{i=k}^{\mathcal{V}} f(i)(\mathcal{A}_k - \bar{\mathcal{A}}_{k-1}) + \mathcal{A}_k k f(k)$$

The difference is thus

$$\Pi(k) - \Pi(k-1) = (\bar{\mathcal{A}}_k - \bar{\mathcal{A}}_{k-1}) \sum_{i=1}^{k-1} if(i) + (\bar{\mathcal{A}}_k - \mathcal{A}_k) k f(k)$$
$$+k(\bar{\mathcal{A}}_k - \mathcal{A}_k) \sum_{i=k+1}^{\mathcal{V}} f(i) + (k-1)(\mathcal{A}_k - \bar{\mathcal{A}}_{k-1}) \sum_{i=k}^{\mathcal{V}} f(i)$$

Dividing everything by $(\bar{\mathcal{A}}_k - \bar{\mathcal{A}}_{k-1})$ and after a bit of algebra we get

$$\Pi(k) - \Pi(k-1) \propto \sum_{i=1}^{k-1} f(i)(i - \frac{\sum_{i=k}^{\mathcal{V}} f(i)}{f(k)})$$
$$\sum_{i=1}^{k-1} f(i)(i - \frac{1}{h(k)})$$

Thus the discrete derivative is in a sense continuous from the left and increases up to k as we had seen above. Dividing the equation by n yields the probabilites and dividing further by $\sum_{i=1}^{\underline{v}} f(i)/n = P(v \leq \underline{v})$ gives the conditioning probabilities present in the statement of the lemma. A direct implication is that if money burning is present in the mechanism, then the cutoff has to be 1, thus there is perfect allocative efficiency.

□

Proof of Lemma B. In this case there is no money burning in the mechanism

$$\begin{aligned}
\Pi(\underline{v}) &= \bar{\mathcal{A}}_{\underline{t}} \sum_{t=1}^{\underline{t}} v_t + \sum_{t=\underline{t}+1}^{n} \mathcal{A}_t (v_t - k) + \sum_{t=\underline{t}+1}^{n} \mathcal{A}_t c_t k \\
&= \bar{\mathcal{A}}_{\underline{t}} \sum_{t=1}^{\underline{t}} v_t + \sum_{\underline{t}+1}^{n} \mathcal{A}_t (v_t - k) + \bar{\mathcal{A}}_{\underline{t}} \sum_{t=\underline{t}+1}^{n} k \\
&= \bar{\mathcal{A}}^{\underline{v}} \sum_{i=1}^{\underline{v}} i f(i) + \sum_{i=\underline{v}+1}^{\mathcal{V}} (i-k) \mathcal{A}^i f(i) + k \bar{\mathcal{A}}^{\underline{v}} \sum_{i=\underline{v}+1}^{\mathcal{V}} f(i) \\
&= \bar{\mathcal{A}}^{\underline{v}} \left(\sum_{i=1}^{\underline{v}} i f(i) + k \sum_{i=\underline{v}+1}^{\mathcal{V}} f(i) \right) + \sum_{i=\underline{v}+1}^{\mathcal{V}} (i-k) \mathcal{A}^i f(i)
\end{aligned}$$

Now let us try calculate $F(\underline{v}+1) - F(\underline{v})$. Using the discrete product rule on the first term after a little algebra we get

$$\begin{aligned}
&\Pi(\underline{v}+1) - \Pi(\underline{v}) \\
&= (\bar{\mathcal{A}}^{\underline{v}+1} - \bar{\mathcal{A}}^{\underline{v}}) \left(\sum_{i=1}^{\underline{v}+1} i f(i) + k \sum_{i=\underline{v}+2}^{\mathcal{V}} f(i) \right) - (\mathcal{A}^{\underline{v}+1} - \bar{\mathcal{A}}^{\underline{v}}) f(\underline{v}+1)(\underline{v}+1-k)
\end{aligned}$$

Since we are only interested in the sign of the discrete derivative we can divide the equation by $(\bar{\mathcal{A}}_{\underline{t}}^{\underline{v}+1} - \bar{\mathcal{A}}_{\underline{t}})$ to get

$$\begin{aligned}
\Pi(\underline{v}+1) - \Pi(\underline{v}) &\propto \left(\sum_{i=1}^{\underline{v}+1} i f(i) + k \sum_{\underline{v}+2}^{\mathcal{V}} f(i) \right) - \sum_{i=1}^{\underline{v}+1} f(i)(\underline{v}+1-k) \\
&\propto \sum_{i=1}^{\underline{v}+1} (i - (\underline{v}+1)) f(i) + k \sum_{i=1}^{\mathcal{V}} f(i)
\end{aligned}$$

Notice the last summand in the first term is 0 by definition. Dividing everything n we recover the terms $f(i)/n$ which correspond to the probability of value i. Dividing and multiplying the first term by $\sum_{i=1}^{\underline{v}} f(i)/n = P(v \leq \underline{v})$ we recover $\frac{\sum_{i=1}^{\underline{v}} if(i)/n}{\sum_{i=1}^{\underline{v}} f(i)/n} = E(v|v \leq \underline{v})$. We recover the more familiar terms

$$\Pi(\underline{v}+1) - \Pi(\underline{v}) \propto P(v \leq \underline{v})E(v - \underline{v} + 1|v \leq \underline{v}) + k$$

□

□

Proof of Proposition 9. Notice that a change in k also changes t^{EF}, so with a slight abuse of notation we will define the function $t^{EF}(k)$, where $t^{EF}(k) = \min\{t \in \{1, \ldots, n\} : v_t \geq k\}$. In order to prove the proposition let us introduce the function $\Pi(\underline{t}, k, k')$ capturing the value of the problem when the cutoff is $\underline{v}$, cost of verification k, the endogenous cutoff is given by $t^{EF}(k')$ and allocation probabilities are given by Proposition 8. Formally:

$$\begin{aligned}\Pi(\underline{t}, k, k') = {} & I_{\underline{t} \geq t^{EF}(k')} \left(\sum_{t=1}^{n} \mathcal{A}_t v_t - \sum_{l=\underline{t}+1}^{n} k \left(\mathcal{A}_l - \mathcal{A}_{\underline{t}} \right) \right) \\ & + I_{\underline{t} < t^{EF}(k')} \left(\sum_{t=1}^{t^{EF}(k')} (n - t + 1)\mathcal{A}_t(v_t - v_{t-1}) + VK^{EF}(k, k') \right)\end{aligned}$$

Where $VK^{EF}(k,k') = \sum_{t=t^{EF}(k')+1}^{n} (\mathcal{A}_t v_t - \mathcal{A}_{t^{EF}(k')} v_{t^{EF}(k')}) - \sum_{t=t^{EF}(k')+1}^{n} k(\mathcal{A}_t - \mathcal{A}_{t^{EF}(k')})$. Our goal is to show that when parametrized by k, the function $F(\cdot, k, k)$ dominates $F(\cdot, k', k')$ for $k > k'$ in the interval order (see Quah and Strulovici (2009)). For a contradiction suppose that this is not true, then there must exist a pair $\underline{t}'' > \underline{t}'$ and for all $\underline{t} \in \{\underline{t}', \ldots, \underline{t}''\}$ we have $\Pi(\underline{t}'', k', k') \geq \Pi(\underline{t}, k', k')$ but $\Pi(\underline{t}'', k, k) < \Pi(\underline{t}', k, k)$. There are four cases we need to consider.

Case I: $\underline{t}'' > \underline{t}' \geq t^{EF}(k)$

In this case we necessarily also have $\underline{t}'', \underline{t}' > t^{EF}(k')$ and there is no money burning in both mechanisms either with k or k'. But then $\Pi(\underline{t}'', k', k') \geq \Pi(\underline{t}', k', k')$ implies that

$$\sum_{t=1}^{\underline{t}''} \mathcal{A}_{\underline{t}''} v_t - \sum_{t=1}^{\underline{t}'} \mathcal{A}_{\underline{t}'} v_t - \sum_{\underline{t}'+1}^{\underline{t}''} \mathcal{A}_t v_t \geq k' \left(\sum_{l=\underline{t}''+1}^{n} (\mathcal{A}_l - \mathcal{A}_{\underline{t}''}) - \sum_{l=\underline{t}'+1}^{n} (\mathcal{A}_l - \mathcal{A}_{\underline{t}'}) \right)$$

Notice that the term in parenthesis is negative as both $\mathcal{A}_{\underline{t}''} > \mathcal{A}_{\underline{t}'}$ and the first summation is over fewer terms. But then the inequality continues to hold when we replace k' with k thus implying $\Pi(\underline{t}'', k, k) \geq \Pi(\underline{t}', k, k)$ delivering a contradiction.

Case II: $t^{EF}(k') > \underline{t}'' > \underline{t}'$

In this case we necessarily have money burning in both mechanisms with k or k'. In particular, costs associated with k only corresponds to $VK^{EF}(\cdot,\cdot)$ in both mechanisms and hence cancel each other out. So we have $\Pi(\underline{t}'', k', k') \geq \Pi(\underline{t}', k', k')$ directly implying $\Pi(\underline{t}'', k, k) \geq \Pi(\underline{t}', k, k)$ delivering the desired contradiction.

Case III: $\underline{t}'' \geq t^{EF}(k) > \underline{t}' > t^{EF}(k')$

By definition we have $F(\underline{t}'', k', k') = \Pi(\underline{t}'', k', k) = \Pi(\underline{t}'', k, k) + (k - k') \sum_{\underline{t}''+1}^{n} \mathcal{A}_t - \mathcal{A}_{\underline{t}''}$. On the other hand we have

$$\Pi(\underline{t}', k', k') \geq \Pi(\underline{t}', k', k) = \Pi(\underline{t}', k, k) + (k - k')(\sum_{t=t^{EF}(k)+1}^{n} \mathcal{A}_t - \mathcal{A}_{t^{EF}(k)})$$

But we already know $\Pi(\underline{t}'', k', k') \geq \Pi(\underline{t}', k', k')$ and hence we have

$$\Pi(\underline{t}'', k, k) + (k - k') \sum_{\underline{t}''+1}^{n} \mathcal{A}_t - \mathcal{A}_{\underline{t}''} \geq \Pi(\underline{t}', k, k) + (k - k')(\sum_{t=t^{EF}(k')+1}^{n} \mathcal{A}_t - \mathcal{A}_{t^{EF}(k')})$$

But, clearly we have $(k-k') \sum_{\underline{t}''+1}^{n} \mathcal{A}_t - \mathcal{A}_{\underline{t}''} \leq (k-k')(\sum_{t=t^{EF}(k')+1}^{n} \mathcal{A}_t - \mathcal{A}_{t^{EF}(k')})$ implying $\Pi(\underline{t}'', k, k) > \Pi(\underline{t}', k, k)$, the desired contradiction.

Case IV: $t^{EF}(k) > \underline{t}'' \geq t^{EF}(k') > \underline{t}'$

By definition we have $\Pi(\underline{t}'', k', k') = \Pi(\underline{t}'', k, k') + (k - k') \sum_{\underline{t}''+1}^{n} \mathcal{A}_t - \mathcal{A}_{\underline{t}''}$. Furthermore by optimality we have $\Pi(\underline{t}'', k, k') \leq \Pi(\underline{t}'', k, k)$ thus we have $\Pi(\underline{t}'', k', k') \leq F(\underline{t}'', k, k) + (k - k') \sum_{\underline{t}''+1}^{n} \mathcal{A}_t - \mathcal{A}_{\underline{t}''}$ On the other hand we have

$$\Pi(\underline{t}', k', k') \geq \Pi(\underline{t}', k', k) = \Pi(\underline{t}', k, k) + (k - k')(\sum_{t=t^{EF}(k)+1}^{n} \mathcal{A}_t - \mathcal{A}_{t^{EF}(k)})$$

But we already know $\Pi(\underline{t}'', k', k') \geq \Pi(\underline{t}', k', k')$ and hence we have

$$\Pi(\underline{t}'', k, k) + (k - k') \sum_{\underline{t}''+1}^{n} \mathcal{A}_t - \mathcal{A}_{\underline{t}''} \geq \Pi(\underline{t}', k, k) + (k - k')(\sum_{t=t^{EF}(k)+1}^{n} \mathcal{A}_t - \mathcal{A}_{t^{EF}(k)})$$

But, clearly we have $(k-k') \sum_{\underline{t}''+1}^{n} \mathcal{A}_t - \mathcal{A}_{\underline{t}''} \leq (k-k')(\sum_{t=t^{EF}(k')+1}^{n} \mathcal{A}_t - \mathcal{A}_{t^{EF}(k')})$ implying $\Pi(\underline{t}'', k, k) > \Pi(\underline{t}', k, k)$, the desired contradiction.

Since all cases lead to a contradiction we must have that $\Pi(\cdot, k, k)$ dominates $\Pi(\cdot, k', k')$ for $k > k'$ in the interval order. Hence by Theorem 1 of Quah and Strulovici (2009), we have that the optimal cutoff is weakly increasing in k. Now, with a slight abuse of notation let $\underline{t}(k)$ denote the cutoff level for cost k. By the above argument we know that $\underline{t}(k)$ is weakly increasing. Recall that $t^{EF}(k)$ is also increasing. Then the types that are subject to verification are all the types $t > \max\{\underline{t}(k), t^{EF}(k)\}$. Thus as k increases, the set of types that are subject to verification weakly decreases. Furthermore and their verification probabilities are given by $c_t = \frac{\max\{\mathcal{A}_{t^{EF}(k)}, \mathcal{A}_{\underline{t}(k)}\}}{\mathcal{A}_t}$. Hence the probability of verification $1 - c_t$ for all the types that are subject to verification is also increasing in k. □

References

Austen-Smith, D. and J. S. Banks (2000). Cheap talk and burned money. *Journal of Economic Theory 91*(1), 1–16.

Ben-Porath, E., E. Dekel, and B. L. Lipman (2014). Optimal allocation with costly verification. *The American Economic Review 104*(12), 3779–3813.

Border, K. C. (1991). Implementation of reduced form auctions: A geometric approach. *Econometrica: Journal of the Econometric Society 59*(4), 1175–1187.

Border, K. C. and J. Sobel (1987). Samurai accountant: A theory of auditing and plunder. *The Review of economic studies 54*(4), 525–540.

Chakravarty, S. and T. R. Kaplan (2008). Manna from Heaven or Forty Years in the Desert : Optimal Allocation without Transfer Payments. (March).

Dunne, S. A. and M. A. Loewenstein (1995). Costly verification of cost performance and the competition for incentive contracts. *The RAND Journal of Economics 26*(4), 690–703.

Eiselt, H. A. and C.-L. Sandblom (2007). *Linear programming and its applications.* Springer Science & Business Media.

Erlanson, A. and A. Kleiner (2019). A note on optimal allocation with costly verification. *Journal of Mathematical Economics 84*, 56–62.

Fujishige, S. (1980). Lexicographically optimal base of a polymatroid with respect to a weight vector. *Mathematics of Operations Research 5*(2), 186–196.

Harris, M., C. H. Kriebel, and A. Raviv (1982). Asymmetric information, incentives and intrafirm resource allocation. *Management Science*.

Hartline, J. J. D. and T. Roughgarden (2008). Optimal mechanism design and money burning. *Proceedings of the fourtieth annual ACM symposium on Theory of computing - STOC 08* (April), 1–23.

Kattwinkel, D. and J. Knoepfle (2023). Costless information and costly verification: A case for transparency. *Journal of Political Economy 131*(2), 504–548.

Li, Y. (2020). Mechanism design with costly verification and limited punishments. *Journal of Economic Theory 186*, 105000.

Maritan, C. A. and G. K. Lee (2017). Resource allocation and strategy.

McAfee, P. R., J. McMillan, and R. P. McAfee (1992, February). Bidding Rings. *American Economic Review 82*(3), 579–599.

Mookherjee, D. and I. Png (1989). Optimal auditing, insurance, and redistribution. *The Quarterly Journal of Economics 104*(2), 399–415.

Mylovanov, T. and A. Zapechelnyuk (2017). Optimal allocation with ex post verification and limited penalties. *American Economic Review 107*(9), 2666–2694.

Perez-Richet, E. and V. Skreta (2023). *Fraud-proof non-market allocation mechanisms.* Centre for Economic Policy Research.

Quah, J. K.-H. and B. Strulovici (2009). Comparative statics, informativeness, and the interval dominance order. *Econometrica 77*(6), 1949–1992.

Ross, M. (1986). Capital budgeting practices of twelve large manufacturers. *Financial Management 15*(4), 15–22.

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

Vohra, R. V. (2004). *Advanced mathematical economics.* Routledge.

Vohra, R. V. (2011). *Mechanism design: a linear programming approach*, Volume 47. Cambridge University Press.

Vohra, R. V. (2012). Optimization and mechanism design. *Mathematical programming 134*(1), 283–303.

Yoon, K. (2011, December). Optimal mechanism design when both allocative inefficiency and expenditure inefficiency matter. *Journal of Mathematical Economics 47*(6), 670–676.