ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Banerjee, Soumen; Chen, Yi-chun; Sun, Yifei

Article

Direct implementation with evidence

Theoretical Economics

Provided in Cooperation with:
The Econometric Society

Suggested Citation: Banerjee, Soumen; Chen, Yi-chun; Sun, Yifei (2024) : Direct implementation with evidence, Theoretical Economics, ISSN 1555-7561, The Econometric Society, New Haven, CT, Vol. 19, Iss. 2, pp. 783-822, https://doi.org/10.3982/TE5015

This Version is available at:
https://hdl.handle.net/10419/320253

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.


https://creativecommons.org/licenses/by-nc/4.0/

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Check for updates

Theoretical Economics 19 (2024), 783–822 1555-7561/20240783

Direct implementation with evidence

Soumen Banerjee
China Center for Behavioural Economics and Finance, Southwestern University of Finance and Economics

Yi-Chun Chen
Department of Economics and Risk Management Institute, National University of Singapore

Yifei Sun
School of International Trade and Economics, University of International Business and Economics

We study full implementation with evidence in an environment with bounded utilities. We show that a social choice function is Nash implementable in a direct revelation mechanism if and only if it satisfies the measurability condition proposed by Ben-Porath and Lipman (2012). Building on a novel classification of lies according to their refutability with evidence, the mechanism requires only two agents, accounts for mixed-strategy equilibria, and accommodates evidentiary costs. While monetary transfers are used, they are off the equilibrium and can be balanced with three or more agents. In a richer model of evidence due to Kartik and Tercieux (2012a), we establish pure-strategy implementation with two or more agents in a direct revelation mechanism. We also obtain a necessary and sufficient condition on the evidence structure for renegotiation-proof bilateral contracts, based on the classification of lies.

Keywords. Full implementation, hard evidence, mechanism design.

JEL classification. C72, D02, D71.

1. Introduction

Consider a government which aims to balance infrastructure development with environmental protection. It consults infrastructure development firms and environmental protection organizations to do a cost-benefit analysis. Without knowing the exact state of the environment, the government needs to factor in inputs from these consultants. However, the consultants have their own incentives, which are not necessarily aligned with that of the government. For instance, infrastructure development firms will always

Soumen Banerjee: soumen08@gmail.com
Yi-Chun Chen: ecsycc@nus.edu.sg
Yifei Sun: sunyifei@uibe.edu.cn

We are grateful to two anonymous referees for their insightful comments, which led to major improvements to the paper. We also thank Bart Lipman, Hamid Sabourian, and Satoru Takahashi for helpful comments and suggestions. Chen gratefully acknowledges the financial support from the Singapore Ministry of Education Academic Research Fund Tier 1 (R-122-000-328-115). Sun gratefully acknowledges the financial support from the National Natural Science Foundation of China (NSFC72273029) and the Fundamental Research Funds for the Central Universities in UIBE (CXTD13-03).

© 2024 The Authors. Licensed under the Creative Commons Attribution-NonCommercial License 4.0. Available at https://econtheory.org. https://doi.org/10.3982/TE5015

prefer to build rather than to not build, whereas environmental protection organizations will always push to err on the side of caution in protecting nature. How can the government glean useful information from these agents whose incentives are conceivably misaligned?

Alternatively, consider a litigation scenario in which a firm is suing a supplier for providing defective parts while the supplier argues that the parts meet the specifications in the initial contract. The judge wants to impose a financial penalty commensurate to the offence, assuming one is proven. The plaintiff always prefers higher penalties while the defendant always prefers smaller ones. How, then, is the judge to decide on the scale of the penalty, given that those in the know do not have the incentive to truthfully reveal whether the parts were defective or the contract unclear?

A common thread that links these scenarios is that the preferences of the agents do not change across states. Other scenarios which share this feature include budget allocation (where agents prioritize obtaining larger shares, irrespective of their requirements), and lobbying (where groups prioritize their own interests which are independent of the state). In all of these situations, classical results on full implementation such as Maskin (1999) and Moore and Repullo (1988) cannot be used, as they rely on preference variation across states. In this paper, we pursue the idea of enriching the mechanism with the use of evidence, so that agents can no longer misreport the state arbitrarily. For instance, the infrastructure development firms or the environmental protection organizations may be able to partially prove the state of the environment by submitting their environmental research reports. Likewise, the defendant in the litigation scenario may be able to prove that the supplied product meets the specifications in the contract or the plaintiff may be able to prove that it does not.

Specifically, we study the full implementation problem with evidence due to Ben-Porath and Lipman (2012). There is a state of the world, which is common knowledge among a set of agents but unknown to a designer. At each state, an agent is endowed with some articles of evidence, which may vary from one state to another. Each article of evidence can be identified with a subset of the state space, and it refutes the possibility that the true state is outside this subset. In other words, we work with hard evidence, which differs in its availability across states, and hence can be used by the agents to partially prove the state.[1] In designing a mechanism, the planner can request evidence presentation as well as cheap talk messages (which are available in every state). In such a setting, Ben-Porath and Lipman (2012) propose a condition called *measurability*. A social choice function is said to be *measurable* with respect to the underlying evidence structure (hereafter, measurable) if whenever the desirable social outcomes differ across two states, at least one agent has a variation in the set of available evidence. To wit, if neither the preferences nor the evidence varies between two states, then any (direct or indirect) mechanism will have the same outcome in both, regardless of the solution concept to which the planner subscribes.

We study an environment in which there are two or more agents with bounded utilities and the designer can impose monetary transfers off the equilibrium. We consider

[1]The hard evidence setting is a special case of the costly evidence setting due to Kartik and Tercieux (2012a), which we study in Section 4.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

a *normal* evidence structure under which each agent can present (in one message) all the articles of evidence, which he is endowed with.[2] In this setting, we show that a social choice function is implementable in mixed-strategy Nash equilibria regardless of the agents' preferences if and only if it is measurable (Theorem 1). Moreover, we obtain the implementation result using a *direct revelation mechanism* in the sense of Bull and Watson (2007), wherein agents only report a state and present an article of evidence. Further, the mechanism achieves budget balance when there are three or more agents; in addition, if we allow the designer to randomize, then the off-the-equilibrium transfers can be made arbitrarily small (Theorem 2).

Our implementation results possesses a number of desirable features. First, we recognize, following the critique due to Jackson (1992), that implementation has long relied on invoking *integer/modulo games* to eliminate unwanted equilibria. While such devices are useful in achieving positive results in general settings, they admit no pure-strategy equilibrium. This problem is exacerbated under mixed strategies: in an integer game, an agent has no best response to an opponent's strategy, which places positive probability on all integers, whereas a modulo game possesses an unwanted mixed-strategy equilibrium. The hope has been that more realistic mechanisms may suffice in more specific settings. Our mechanism makes use of neither integer/modulo games nor sequential moves.[3] Rather, we use a direct revelation mechanism, which has the simplest possible message structure in implementing arbitrary measurable social choice functions. The design of our direct revelation mechanism is based on a novel classification of lies (i.e., inaccurate state claims). Specifically, there are lies, which can be refuted by evidence, possessed by other agents (other-refutable lies), lies which can be refuted only by the agent who is reporting them (self-refutable lies), and lies which cannot be refuted by any evidence available under the true state (nonrefutable lies).[4] We design transfer rules, which eliminate the lies successively in any mixed-strategy equilibrium.

Second, our result extends to settings where evidence presentation is costly. Specifically, as long as the designer knows the bound of the agents' evidentiary cost, every measurable social choice function remains directly implementable in mixed-strategy Nash equilibria regardless of the cost structure (Theorem 3).[5] In contrast, when there is a fixed evidentiary cost structure, which is common knowledge, the designer is able to distinguish between states by exploiting the cost variation and measurability is no longer necessary for implementation. In such a setting, Kartik and Tercieux (2012a) propose a condition called *evidence monotonicity*, and show that it is necessary for implementation where only the cheapest evidence is submitted in equilibrium. We adapt the notion

[2]Normality is also imposed in Theorem 2 of Ben-Porath and Lipman (2012), which achieves Nash implementation by invoking integer games and ε transfers. For a detailed comparison between our results and the results of Ben-Porath and Lipman (2012), see Section 6.

[3]Implementation with equililbrium refinements, which do not possess the closed-graph property (e.g., subgame-perfect Nash equilibrium) need not be robust to a "small amount of incomplete information about the state"; see Chung and Ely (2003) and Aghion et al. (2012).

[4]It follows from measurability that if state s' is not refutable at state s and induces a different social outcome, then s must be refutable at s'.

[5]It is well recognized in the literature that there may be material or psychological costs associated with the presentation of evidence. See Bull and Watson (2007), Ben-Porath and Lipman (2012), and Kartik and Tercieux (2012a) for instance.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of evidence monotonicity to our environment and establish that evidence monotonicity is sufficient for pure-strategy Nash implementation with two or more agents via a direct revelation mechanism (Theorem 4).[6]

Third, as our results hold even when there are only two agents, they allow for applications to a classical issue in the incomplete contract literature. The difficulty in this regard arises when a desirable contractual outcome (e.g., an efficient trade) needs to be conditioned on some state variables (e.g., the a buyer's valuation of some good), which are observable to both contractual parties and yet not verifiable by a third party such as a court. A well-known solution is to invoke implementation theory to design a mechanism, which has the agents announce the observed state as a verifiable equilibrium message. However, such mechanisms often involve off-the-equilibrium transfers, which penalize both agents so that such mechanisms are susceptible to renegotiation.[7] Measurability with respect to an evidence structure may be considerably easier to satisfy in practice than verifiability. Indeed, even when a state is not verifiable, the agents may still be able to provide evidence to refute certain states. In Section 5, we establish a necessary and sufficient condition on the evidence structure for the existence of renegotiation proof bilateral contracts. In particular, such a contractual outcome must lie on the Pareto frontier of the agents' utility possibility set and thereby must achieve budget balance.

The rest of the paper is organized as follows. Section 2 provides a formal description of the model, the implementing condition, and further details the classification of lies on which the mechanism is based. Section 3 presents the main implementing mechanism and a formal proof of implementation. Following this, we also establish budget balance with three or more agents (Section 3.8) and implementation with small transfers (Section 3.9). In Section 4, we extend our results to settings where evidence presentation is costly. Section 5 details our treatment of renegotiation-proof contracting and we conclude by comparing our results to the existing literature.

2. Model

2.1 *Environment*

Let $\mathcal{I}=\{1,\dots,I\}$ $(I\geq 2)$ be a set of agents, A, a set of social outcomes, and S a set of states. Suppose that S is finite. Agents have quasilinear utilities in transfers, so that $u_i(a,s,\tau)=v_i(a,s)+\tau$ where τ is the transfer to the agent. We assume that v_i is bounded and without loss of generality, we set $v_i:A\times S\to[0,1]$ (in dollars). As a result, an agent can be induced to accept any outcome if the alternative were to be any other outcome with a penalty of 1 dollar. A social planner would like to implement a social choice function (SCF) $f:S\to A$. We assume that while the true state is common knowledge among the agents, it is unknown to the social planner.

[6]We also obtain mixed-strategy implementation under a stronger version of evidence monotonicity in Banerjee, Chen, and Sun (2023).

[7]For an example of the issues posed by renegotiation, we refer the reader to Maskin and Moore (1999) where they demonstrate a setting in which only a null contract is renegotiation-proof even though it is efficient to trade with verifiable states. Maskin and Tirole (1999) attempt to circumvent this issue by using lotteries, but their result depends crucially on at least one agent being strictly risk averse.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2 *Evidence*

We assume that each agent i is endowed with a (state-dependent) collection of articles of evidence $\mathcal{E}_i : S \rightrightarrows 2^S$. In particular, by providing a message $E_i \in \mathcal{E}_i(s)$ in a state s, agent i establishes that the true state lies within E_i. At state s, we say that an article of evidence $E_i \in \mathcal{E}_i(s)$ *refutes* state s' if $s' \notin E_i$.

Following Ben-Porath and Lipman (2012), we introduce the following definition.

DEFINITION 1. An evidence structure satisfies the following conditions:

(e1) it is impossible to refute the truth, i.e., $\forall s \in S$, $E_i \in \mathcal{E}_i(s)$ only if $s \in E_i$;

(e2) if an agent can prove the event E in some state, then they must be able to do so in all the states in E, i.e., $E \in \mathcal{E}_i(s)$ only if $E \in \mathcal{E}_i(s')$ for every $s' \in E$.

We stress here that this is not an assumption. Rather, if articles of evidence differ in their availability across states, then without loss of generality, we can *name* the article in terms of the subset of states in which it is available. With such names, the above properties must be satisfied.

We say that a setting involves *hard* evidence if the set of evidence available to an agent can change from state to state. Furthermore, we say that the evidence structure is *normal* if, for every agent i, and every state s,

$$E_i^*(s) \equiv \bigcap_{E \in \mathcal{E}_i(s)} E \in \mathcal{E}_i(s).$$

The idea behind normality is that it is feasible for agents to present all their evidence at once, suggesting an idealization in which there are no time or other constraints on doing so. We refer to this message containing all the evidence that an agent has as the *tightest* evidence of the agent in the given state. For a start, we prove our first main result (Theorem 1) under the assumption of normality. We will discuss in Section 3.7 how this result should be modified when the normality assumption does not hold. We define a social choice environment as a tuple $\Psi = (\mathcal{I}, A, S, \{\mathcal{E}_i\}_{i \in \mathcal{I}}, f)$, which is assumed to be common knowledge among the designer and agents. We write $s \sim s'$ (and say that s is equivalent to s') if $\mathcal{E}_i(s) = \mathcal{E}_i(s')$ for all i.

2.3 *Illustrating example*

To illustrate the above ideas, we consider a situation wherein a government (acting as the social planner) is considering whether to approve a development project, which has an adverse environmental impact. The agents concerned are an infrastructure development firm (F) and an environmental protection organization (O). Suppose the three actions available to the planner are to allow the project (P_l), ask for it to be made smaller (P_m), or to scrap it entirely (P_h). It wishes to make these decisions if the threat to the environment is low (s^l), medium (s^m), or high (s^h), respectively. The government does not know the degree of the threat to the environment, while both F and O being more familiar with the scenario, know that the true threat level is medium (s^m).

Profit-making firm (F) is solely concerned with implementing the project, and thus has a preference ordering $P_l \succ P_m \succ P_h$. The environmental protection organization (O)

1555761, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

has the preference ordering $P_h \succ P_m \succ P_l$ since it strictly prioritizes protecting the environment. These preference orderings are not dependent on the state and this is where this scenario departs from the classical/Maskin-type implementation problem.

The social planner now requests the two parties to submit evidence for and against the project. Suppose that articles of evidence are of the form of discoveries of environmental degradation, which refute lower-impact states. Moreover, the state s^h is assumed to be associated with a very serious degree of environmental impact, so that if the true state were s^h, it would be apparent to (and provable by) all the agents. To sum up these ideas, we present below the evidence structure for each agent in all possible states:

Agent/State	s^l	s^m	s^h
Firm (F)	$\{\{s^l, s^m, s^h\}\}$	$\{\{s^l, s^m, s^h\}, \{s^m, s^h\}\}$	$\{\{s^l, s^m, s^h\}, \{s^m, s^h\}, \{s^h\}\}$
Organization (O)	$\{\{s^l, s^m, s^h\}\}$	$\{\{s^l, s^m, s^h\}\}$	$\{\{s^l, s^m, s^h\}, \{s^h\}\}$

Note that articles of evidence are subsets of the state space so that an agent presenting $\{s^m, s^h\}$ informs the designer only that the state is not s^l. We also say that this article of evidence refutes s^l. Thus, in state s^m, reporting the state s^l is a lie that only F can refute, and reporting the state s^h is a lie that no one can refute. It is interesting to note that in the true state (s^m), O would like to convince the planner that the true state is s^h (a claim that no one can refute in this scenario) and F would like to convince the planner that the true state is s^l (a claim that only F can refute).

2.4 *Measurability and implementation*

When the agents' preferences do not vary across states (i.e., v does not depend on s), evidence is the only way to differentiate two states. Indeed, if two states induce the same preference profile and evidence endowments, then irrespective of the solution concept in use, in any mechanism they must be associated with the same set of equilibria.[8] Measurability of an SCF entails that when the planner wants to implement different outcomes from one state to another, there must be at least one agent who can differentiate between the states in terms of their evidence set. We now state the formal definition.

DEFINITION 2. Given a social environment Ψ, an SCF f satisfies measurability if $f(s) = f(s')$ whenever $s \sim s'$.

In other words, if f is measurable, then in implementing the desirable social outcome, the designer needs to identify only the equivalence class of states, which contains the truth.

To fix ideas, consider the example in Section 2.3 with the modification that in state s^m, F is endowed only with $\{\{s^l, s^m, s^h\}\}$. In this case, no agent can differentiate between s^l and s^m, so that any implementable SCF must be constant between these two states. More generally, if there were to be no evidence in the model, so that $\mathcal{E}_i(s) = \{S\}$ for

[8]Unless otherwise specified, we assume that articles of evidence do not have costs associated with them so that the designer cannot exploit cost variation.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

all agents in all states, then only constant SCFs are implementable. Further, even if each agent were endowed only with $\{\{s^l, s^m, s^h\}\}$ in each state with the exception of only one agent being endowed with $\{s^h\}$ in state s^h, it (we shall show later) would still be possible to implement different outcomes in s^h relative to s^l and s^m, so that even the slightest variation in evidence is sufficient for implementing different outcomes.

A mechanism $\mathcal{M}$ in this social choice environment is defined as a tuple $\mathcal{M} = (M, g, (\tau_i)_{i\in\mathcal{I}})$ where $M = \Pi_{i\in\mathcal{I}} M_i$ is a finite set of message profiles, $g : M \to A$ is the outcome function, and $\tau_i : M \to \mathbb{R}$ is the payment rule for agent i. A mechanism $\mathcal{M}$ together with a profile of utility functions $v = (v_i)_{i\in\mathcal{I}}$ with $v_i : A \times S \to [0, 1]$ induces a complete-information game $G(\mathcal{M}, v, s)$ at state s. We call a mechanism a *direct revelation mechanism* (Bull and Watson (2007)) when $M_i = S \times \mathcal{E}_i$, i.e., every agent submits only one claim of state and one article of evidence.

A (mixed) *strategy* of agent i in the game $G(\mathcal{M}, v, s)$ is a probability distribution σ_i over M_i, which we also denote by $\sigma_i \in \Delta M_i$. A strategy profile $\sigma = (\sigma_1, \ldots, \sigma_I) \in \times_{i\in\mathcal{I}} \Delta M_i$ is said to be a (mixed-strategy) *Nash equilibrium* of the game $G(\mathcal{M}, v, s)$ if, for any agent $i \in \mathcal{I}$ and for any messages $m_i \in \operatorname{supp}(\sigma_i)$, we have

$$\sum_{m_{-i}\in M_{-i}} \sigma_{-i}(m_{-i})\big[v_i\big(g(m_i, m_{-i}), s\big) + \tau_i(m_i, m_{-i})\big]$$
$$\geq \sum_{m_{-i}\in M_{-i}} \sigma_{-i}(m_{-i})\big[v_i\big(g\big(m_i', m_{-i}\big), s\big) + \tau_i\big(m_i', m_{-i}\big)\big], \quad \forall m_i' \in M_i$$

where $\sigma_{-i}(m_{-i}) = \Pi_{j\neq i}\sigma_j(m_j)$. A pure-strategy Nash equilibrium is a Nash equilibrium σ, which assigns probability one to some message profile m.

Unlike the classical implementation problem, and in recognition of practical situations in which preferences do not vary across states, we seek to implement the SCF by relying on evidence instead of preference reversal. To stress this difference, we require in the following definition that implementation is achieved regardless of the profile of utility functions. This also has the effect of strengthening our result from the point of view that we no longer require that preferences vary between states for implementation.[9]

DEFINITION 3. An SCF f is Nash-implementable if there is a mechanism $\mathcal{M} = (M, g, (\tau_i)_{i\in\mathcal{I}})$ such that for any profile of bounded utility functions $v = (v_i)_{i\in\mathcal{I}}$, any state s, and any mixed-strategy Nash equilibrium σ of the game $G(\mathcal{M}, v, s)$, $g(m) = f(s)$, and $\tau_i(m) = 0$ for each message profile $m \in \operatorname{supp}\sigma(s)$.

That is, a mechanism implements a social choice function if at any state, and with any profile of bounded utility functions, any mixed-strategy Nash equilibrium outcome coincides with the outcome of the SCF. That is, we ask for full implementation as opposed to partial implementation that requires only one equilibrium achieving the outcome of the SCF. Note that partial implementation is trivial in such a setting, as all that

[9]Evidence variation, however, continues to be necessary. Kartik and Tercieux (2012a) provide the minimum necessary condition in this context, when both preference and evidence variation are combined—evidence monotonicity. For the purposes of this paper, we set aside preference variation and focus on evidence variation alone.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

would be required to achieve it would be to heavily penalize all agents for disagreeing with each other about the state, so that there would be an equilibrium where each agent tells the truth and there are no transfers.

In what follows, we consider two states to be different only if they induce different tightest evidence for at least one agent. Otherwise, they are treated as the same state. In other words, we identify S with its quotient space $S/\sim$ induced by the equivalence relation $\sim$ where each point corresponds to an equivalent class. Owing to the necessity of measurability, this is without loss of generality for our implementation exercise.

2.5 *A classification of lies*

We will construct a direct revelation mechanism, which leverages two ways to use evidence in cross-checking any claim of state. First, an article of evidence may be able to *refute* a state claim. That is, it establishes that the state claim is definitely not true. Second, it is possible for the designer to pick out state claims, which have not been fully *supported* by agents, i.e., states for which agents have not provided all the evidence they ought to have if the state were true. If a state claim is not supported by an agent even though he is incentivized to do so, then it signals to the designer that the state claim is false. These two ideas underlie the mechanism, which we will present later.

Formally, agent i is said to have *supported* a state claim s in an evidence message E_i if $E_i \subseteq E_i^*(s)$. Based upon the notion of refutation, we distinguish three different types of lies wherein a lie is a claim of state s', which is different from s^* (under the relation $\sim$ defined previously). First, an *other-refutable* lie for agent i is a lie that at least one agent other than i has the evidence to refute in the true state. For instance, for organization O, the lie s^l is an other-refutable lie in state s^m since it can be refuted by the article $\{s^m, s^h\}$ possessed by Firm F. Note that it is straightforward to construct a transfer rule so that no agent will tell other-refutable lies; this is done by just requiring an agent to pay a large penalty to whoever refutes his state claim.

Second, a *self-refutable* lie for agent i is a lie that only agent i has the evidence to refute in the true state. For instance, for firm F, the lie s^l is a self-refutable lie in state s^m since it can only be refuted by the article $\{s^m, s^h\}$ possessed by the firm itself.

Finally, a *nonrefutable lie* is a lie that cannot be refuted by any evidence that is possessed by any agent. For instance, the lie s^h is a nonrefutable lie in state s^m.

From an agent's perspective, the truth, other-refutable, self-refutable, and nonrefutable lies partition the entire state space. To see this, notice that given a lie, it can either be refuted at the true state, or not. If it can be refuted, it can either be refuted by other agents, or only by the agent in question.

We now prove the following observations, which will be exploited in proving our main result.

Observation 1. *If an agent i cannot refute s' at s^*, then every article of evidence available to him at s^* is also available to him at s'.*

Proof. If $s' \notin RL_i(s^*)$, then every article of evidence available to i at s^* contains s'. Then, from Property (e2) of Definition 1, every such article is available to i at s', so that $\mathcal{E}_i(s^*) \subseteq \mathcal{E}_i(s')$. □

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

OBSERVATION 2. *If s' is a nonrefutable lie at s^*, then some agent must have an article of evidence at state s', which refutes s^*.*

PROOF. As s' is nonrefutable for i at s^*, Observation 1 yields that $\mathcal{E}_i(s^*) \subseteq \mathcal{E}_i(s')$ for every i. Since s' is a lie, $s' \nsim s^*$. Hence, $\mathcal{E}_i(s^*) \subset \mathcal{E}_i(s')$ for some i, and from (e2) of Definition 1 any member of $\mathcal{E}_i(s') \backslash \mathcal{E}_i(s^*)$ must refute s^*. In other words, any article of evidence that is available to agent i under s' and not available under s^* must refute s^*. □

Observation 1 establishes that supporting either a self-refutable lie of another agent, or a nonrefutable lie requires the presentation of at least as much evidence as supporting the truth. Observation 2 entails that it is impossible to have every agent support a nonrefutable lie. Since nonrefutable lies cannot be directly refuted by evidence, this inability to support them forms the only way to eliminate them in equilibrium.

3. IMPLEMENTING MECHANISM

Fix the environment $\Psi = (\mathcal{I}, A, S, \{\mathcal{E}_i\}_{i\in\mathcal{I}}, f)$. We now present our main result.

THEOREM 1. *Suppose the evidence structure is given by $\mathcal{E}_i(\cdot)$. Then an SCF f is Nash-implementable in a direct revelation mechanism if and only if it is measurable with respect to $\mathcal{E}_i(\cdot)$.*

The necessity of measurability follows from the fact that if two states are associated with the same set of evidence, and the preferences are constant among states, then any mechanism must have the same set of equilibria in both states. In the following subsections, we prove the sufficiency part of Theorem 1 by constructing a direct revelation mechanism, which implements f.

3.1 *Message space*

Every agent has a typical message $m_i = (s_i, E_i) \in M_i = S \times \mathcal{E}_i$. We interpret this as a claim of state and an article of evidence. The typical message m_i therefore is of the form (s_i, E_i). In the following, we denote the full message profile (of all agents) by $m \in \Pi_{i\in\mathcal{I}} M_i$.

3.2 *Outcome*

We define the outcome function of the mechanism as

$$g(m) = f(s_1).$$

That is, we implement the social outcome according to the state claim made by the first agent. However, we will show that in any equilibrium all agents report the true state. Hence, any Nash equilibrium achieves the desirable social outcome.

3.3 *Transfers*

There are four different types of transfers in the mechanism, which we will introduce one by one. The first transfer applies when an agent refutes a state claim of another

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

agent. In this case, the agent whose claim is refuted has to pay a penalty to the agent who refutes the claim. That is,

$$\tau_{ij}^1(m) = \begin{cases} 2I+1, & \text{if } s_i \in E_j \text{ and } s_j \notin E_i; \\ -2I-1, & \text{if } s_i \notin E_j \text{ and } s_j \in E_i; \\ 0, & \text{otherwise} \end{cases}$$

where I is the number of agents.

Under the second transfer, an agent incurs a penalty if his state claim is not supported by himself or other agents. Formally,

$$\tau_i^2(m) = \begin{cases} -I, & \text{if } \exists j \in \mathcal{I} \text{ such that } E_j \not\subseteq E_j^*(s_i); \\ 0, & \text{otherwise.} \end{cases}$$

The third transfer penalizes an agent (say, agent i) if he disagrees with another agent (say, agent j) along the evidence dimension for agent i. This is expressed as follows:

$$\tau_{ij}^3(m) = \begin{cases} -1, & \text{if } E_i^*(s_i) \neq E_i^*(s_j); \\ 0, & \text{otherwise.} \end{cases}$$

The fourth transfer is a penalty proportional to the cardinality of states that are not refuted by the evidence presented by agent i. This is active when an agent has made a state claim, which one or more agents have not supported. Formally,

$$\tau_i^4(m) = \begin{cases} -\dfrac{|E_i|}{|S|}, & \text{if } E_j \not\subseteq E_j^*(s_{j'}) \text{ for some } j, j' \in \mathcal{I}; \\ 0, & \text{otherwise.} \end{cases}$$

We stress that this applies to one's own state claims as well, i.e., being unable to provide the tightest evidence for one's own state claim also incurs a penalty from this transfer.

With $\tau_i^1 = \sum_{j \neq i} \tau_{ij}^1$ and $\tau_i^3 = \sum_{j \neq i} \tau_{ij}^3$, we define the overall transfer to agent i as

$$\tau^i = \tau_i^1 + \tau_i^2 + \tau_i^3 + \tau_i^4.$$

3.4 *Proof sketch*

The mechanism deals with each type of lie in sequence. We begin with other-refutable lies, for instance the lie s^l for O in state s^m in the illustrating example. This involves the transfer τ^1. Recall that by definition, an other-refutable lie for some player (say i) is refutable by a different agent, (say j). We note that τ^1 provides agent j the incentive to refute agent i's lie irrespective of the probability with which i presents it (normality ensures that he does not have to sacrifice rewards from refuting another agent's lies). It also assures agent i of a large penalty from refutation, which can be avoided by deviating to the truth (which is irrefutable).

Before eliminating the remaining lies, we comment on the role of τ^4. The transfer τ^4 is a cardinality transfer which (when it is active) incentivizes the presentation of the

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

tightest evidence by all agents. We construct the mechanism in a way that presenting additional evidence is never harmful for an agent. Indeed, when τ^4 is active, it strictly benefits the agent, since presenting an additional article of evidence reduces $|E_i|$, and thus reduces the magnitude of the fine. This allows the mechanism to elicit the true profile of tightest evidence, which is useful for the elimination of both nonrefutable and self-refutable lies. This crucially depends upon normality, since incentivizing agents to present their tightest evidence would not achieve its goal unless such an article is *available* and *can be* presented.

Now, we eliminate nonrefutable lies, for instance the lie s^h in state s^m in the illustrating example. This step involves τ^2 and τ^4. Recall that if s' is a nonrefutable lie at s^*, then some agent must have the evidence to refute s^* at s'; for instance, $\{s^h\}$ refutes s^m at s^h. Since by assumption it is not possible to eliminate s^* if it is the truth, there is at least one agent who cannot support the claim s' if the true state is s^*. Therefore, if an agent i presents a nonrefutable lie, then he knows that it cannot be supported. In the mechanism, this has two effects. First, he gets a large fine from τ^2. Second, τ^4 is active, so that all agents have a strict incentive to present all their evidence. This has the consequence of allowing agent i to evade the fine from τ^2 by switching to the truth (if everyone presents all their evidence, the truth is supported by everyone).

If we prioritize the first two steps, then agents are restricted to presenting either the truth or self-refutable lies, each of which (from Observation 1) induces for any other agent an evidence set at least as large as the truth. In this case, we claim that all agents present their tightest evidence. Indeed, if any agent withholds evidence, they fail to support the state claims of all other agents, so that τ^4 is active, and submitting additional evidence is a profitable deviation. This step, which involves τ^2, τ^3, and τ^4, uses this fact to deal with self-refutable lies.

Finally, suppose that an agent i presents a self-refutable lie, for instance F claims s^l in state s^m. Since all evidence is tightest, the transfer τ^2 incentivizes other agents to present state claims that are consistent with agent i's tightest evidence. Following this, the cross-check in τ^3 yields a penalty for agent i because a self-refutable lie for agent i is inconsistent with agent i's own tightest evidence. This penalty can be avoided by switching to the truth, since all agents are presenting their tightest evidence. In summary, in any equilibrium everyone presents the truth with the tightest evidence. Hence, the mechanism implements the social choice function.

3.5 *Proof of implementation*

We present below a formal proof of implementation using the direct revelation mechanism in Section 3. In what follows, we denote the true state by s^*. For simplicity, we will write NRL and SRL_i to denote $NRL(s^*)$ and $SRL_i(s^*)$, respectively. Fix an arbitrary mixed-strategy Nash equilibrium σ.

3.5.1 *Preliminary results* We begin by proving a lemma, which finds use at multiple points in the proof.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

LEMMA 1. *Given the strategies of other agents, an agent never incurs a loss from presenting more evidence (while holding the state claim fixed) in a deviation. Moreover, if τ_i^4 is active with nonzero probability, then under any optimal strategy, agent i presents the tightest evidence.*

PROOF. Fix agent i. It is clear that τ_i^1 causes no loss from presenting additional evidence. Indeed if agent i refutes another agent's state claim, he may make a profit from tightening the evidence. τ_i^2 only requires that the evidence presented be as tight as some bound, so that tightening the evidence causes no loss from τ_i^2 either. τ_i^3 is a statement on state claims, so evidence is not related to this transfer. τ_i^4 clearly causes no loss from tightening, rather, it causes strict gains if it is active. Therefore, given a strategy profile for other agents, if an agent expects τ^4 to be active with positive probability, then it is optimal for him to present the tightest evidence, as otherwise he could tighten his evidence to improve his payoff. □

3.5.2 *Eliminating other-refutable lies*

CLAIM 1. *If agent i reports with positive probability a message $m_i = (s_i, E_i)$ such that agent $j \neq i$ has an article of evidence which refutes s_i, then E_j must refute s_i for every message $m_j = (s_j, E_j)$, which agent j reports with positive probability.*

PROOF. Suppose not. Then agent j's evidence does not refute s_i, i.e., $s_i \in E_j$. Consider an alternate message $\tilde{m}_j$, that only replaces E_j with $E_j^*(s^*)$, which refutes s_i. The following table documents agent j's payoff change by deviating from m_j to $\tilde{m}_j$:

g	τ_i^1	τ_i^2	τ_i^3	τ_i^4	In total
0	>0	0	≥ 0	≥ 0	>0

That is, agent j gains from τ^1, on account of the fact that he has now refuted agent i's lie. By Lemma 1, there is no loss from any of the other transfers. □

CLAIM 2. *No agent reports an other-refutable lie with positive probability.*

PROOF. From Claim 1, if agent i reports a lie s_i, which agent j can refute, then with probability one agent j must present E_j to refute s_i. Consider an alternative message $\tilde{m}_i$, which replaces s_i with s^* in m_i. The following table summarizes the payoff change by deviating from m_i to $\tilde{m}_i$:

g	τ_i^1	τ_i^2	τ_i^3	τ_i^4	In total
>-1	$2I+1$	$\geq -I$	$\geq -I+1$	≥ -1	>0

Agent i gains a minimum of $2I+1$ from τ^1 due to the fact that agent j was refuting s_i. In addition, in the worst case, agent i loses at most 1 from outcome g, at most I from τ^2, at most $I-1$ from τ^3, and less than $\frac{|E_i|}{|S|}$ (which is less than 1) from τ^4. Hence, $\tilde{m}_i$ is a profitable deviation from m_i. □

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.5.3 *Eliminating nonrefutable lies* The next two claims eliminate the possibility that nonrefutable lies are reported in equilibrium.

CLAIM 3. *If an agent reports with positive probability a message $m_i = (s_i, E_i)$ where s_i is a nonrefutable lie at s^*, then E_i is the tightest, and every agent $j \neq i$ must provide the tightest evidence available, i.e., $E_j = E_j^*(s^*)$.*

PROOF. Since $s_i \in NRL$, it follows from Observation 2 that there is an agent $j \in \mathcal{I}$ who can refute s^* at s_i. However, since the evidence structure $\mathcal{E}_i(\cdot)$ satisfies condition (e1) of Definition 1, agent j cannot present $E_j^*(s_i)$ (which must refute s^*) at the state s^*. Thus, whenever there is an agent i who reports with positive probability a message m_i with a nonrefutable lie s_i, then for every $j \neq i$, τ_j^4 must be triggered with positive probability. It then follows from Lemma 1 that each agent $j \neq i$ must present $E_j = E_j^*(s^*)$ under any optimal strategy. Further, in presenting s_i, agent i knows that he (or another agent) will be unable to support s_i so that τ_i^4 is active with probability 1. Therefore, Lemma 1 yields $E_i = E_i^*(s^*)$ since agent i plays an optimal strategy. □

CLAIM 4. *No agent reports a nonrefutable lie with positive probability.*

PROOF. Suppose not. By Claim 3, if there is an agent who reports a nonrefutable lie s_i in message $m_i = (s_i, E_i)$, then $E_i = E_i^*(s^*)$ and every agent $j \neq i$ will present $E_j = E_j^*(s^*)$. Now, consider an alternative message $\tilde{m}_i$, which replaces s_i with s^*. The following table summarizes the payoff change by deviating from m_i to $\tilde{m}_i$:

g	τ_i^1	τ_i^2	τ_i^3	τ_i^4	In total
> -1	0	$\geq I$	$\geq -I+1$	≥ 0	>0

First, fixing m_i where s_i is nonrefutable, we know that $\tau_i^2(m_i, m_{-i}) = -I$ for every m since there exists some agent j (j may be i) such that $E_j^*(s_i) \subset E_j^*(s^*)$. Hence, agent i gains at least I from τ_i^2 from the deviation. Second, he loses less than 1 from changing the outcome, incurs no loss from τ_i^1 (since the truth is not refutable), and at most $I-1$ from τ_i^3. Third, agent i incurs no losses from τ_i^4 either as he was facing a scenario of no support with s_i before the deviation and the size of his evidence set has not changed. Overall, this is therefore a profitable deviation. □

3.5.4 *Eliminating self-refutable lies* The next three claims eliminate self-refutable lies. Up to this point, we have established that agents can report only the truth or self-refutable lies in equilibrium. Observation 1 implies that in this situation, agents would need to present their tightest evidence to support other agent's state claims. This is the basic idea, which we will exploit in establishing the following claims.

CLAIM 5. *All agents present the tightest evidence with probability one. That is, for any agent i, $E_i = E_i^*(s^*)$ for any message $m_i = (s_i, E_i)$ on the support of σ_i.*

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

PROOF. Suppose to the contrary that for some agent i, $m_i = (s_i, E_i)$ is on the support of σ_i, and $E_i \neq E_i^*(s^*)$. From Claims 2 and 4, other agents report either a self-refutable lie or the truth in equilibrium. Therefore, from Observation 1, if $E_i \neq E_i^*(s^*)$, then agent i expects τ^4 to be active with positive probability as he is not supporting any other agent's claims. Deviating to $\tilde{m}_i = (s_i, E_i^*(s^*))$ from m_i is a profitable deviation in this case. □

CLAIM 6. *For every agent i, and for any message $m_i = (s_i, E_i)$ on the support of σ_i, we must have $E_j^*(s_i) = E_j^*(s^*)$ for every agent $j \neq i$.*

PROOF. Suppose to the contrary that agent i reports $m_i = (s_i, E_i)$ with $E_j^*(s_i) \neq E_j^*(s^*)$ for some agent $j \neq i$. It follows from Claims 2 and 4 that s_i is a self-refutable lie for agent i. Since $E_j^*(s_i) \neq E_j^*(s^*)$, Observation 1 implies that $E_j^*(s_i) \subset E_j^*(s^*)$. From Claim 5, all agents present the tightest evidence with probability one, i.e., $E_j = E_j^*(s^*)$ with σ_j-probability one for every agent $j \in \mathcal{I}$.

Consider a deviation for agent i to the truth, i.e., consider a message $\tilde{m}_i$, which only replaces s_i with s^* in m_i. The following table summarizes the payoff change by deviating from m_i to $\tilde{m}_i$:

g	τ_i^1	τ_i^2	τ_i^3	τ_i^4	In total
> -1	≥ 0	$\geq I$	$\geq -I+1$	≥ 0	>0

The agent gains I from τ^2. This is because, by assumption, $E_j^*(s_i) \subset E_j$ for some $j \neq i$ (i.e., j cannot support s_i at s^*) and $E_j^*(s^*) = E_j$ for every agent j. Moreover, the agent loses less than -1 due to the outcome, and at most $I-1$ to τ^3, and incurs no loss from τ^1 (the truth is not refutable) or τ^4 (all claims are supported now). In conclusion, this is a profitable deviation. □

CLAIM 7. *No agent reports a self-refutable lie with positive probability.*

PROOF. Suppose to the contrary that agent i reports $m_i = (s_i, E_i)$ where s_i is a self-refutable lie.

Consider a deviation to the truth for agent i, i.e., consider a message $\tilde{m}_i$, which only replaces s_i with s^* in m_i. The following table summarizes the payoff change by deviating from m_i to $\tilde{m}_i$:

g	τ_i^1	τ_i^2	τ_i^3	τ_i^4	In total
> -1	≥ 0	≥ 0	$\geq I-1$	≥ 0	>0

In words, since s_i is a self-refutable lie and $E_i^*(s_j) = E_i^*(s^*) \neq E_i^*(s_i)$, agent i gains at least $I-1$ from τ_i^3 (wherein the first equality is from Claim 6); moreover, the agent incurs a loss of at most 1 from the outcome and no loss or gains from other transfers. Hence, this is a profitable deviation. □

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.5.5 *Implementation* To sum up, it follows from Claims 2, 4, and 7 that with probability one each agent reports the true state. Hence, to prove implementation, we need only establish the following claim.

CLAIM 8. *In equilibrium, no transfer is incurred.*

PROOF. Since all state claims are truthful, it suffices to argue that all agents present the tightest evidence in equilibrium. Indeed, if any agent is not presenting the tightest evidence, then τ^4 is active with positive probability. It then follows from Lemma 1 that deviating to the tightest evidence is a profitable deviation. □

3.6 *Discussion*

We add a few remarks here. First, notice that from Lemma 1, presenting additional evidence is never harmful, and is beneficial under some cases. Therefore, it is a weakly dominant strategy to always present all the evidence. It is clear that the above mechanism implements under iterated elimination of weakly dominated strategies as well. This yields us double implementation, in both mixed Nash equilibrium, and iterated elimination of weakly dominated strategies.[10]

Second, in considering the above mechanism, it is evident that even though the designer can impose transfers off the equilibrium, it is not sufficient to simply penalize any profiles the designer finds undesirable. Rather, the main challenge is to allow the agents a profitable deviation as well. The idea can be seen from how the mechanism eliminates both nonrefutable and self-refutable lies. For the elimination of nonrefutable lies, it is critical that when an agent is being penalized for an unsupported nonrefutable lie, the other agents must be presenting their tightest evidence, so that the truth is actually supported, enabling the agent to avoid the penalty by deviating to the truth. If this were not the case, then in deviating to the truth, the agent's report will still be unsupported. Coming to the elimination of self-refutable lies, it is critical that we extract the article refuting a self-refutable lie from the agent himself, as otherwise the cross-checks starts from the wrong profile of evidence, and we would not be able to realign the profile toward the truth.

This leads to a somewhat counterintuitive design choice, which we make during the elimination of self-refutable lies. Notice that when an agent presents a self-refutable lie, *and* presents his tightest evidence, he actually *refutes* his own state claim. This does not in fact lead to a penalty for the agent, although penalizing an agent for being internally inconsistent could be logical in certain circumstances.[11] Our mechanism, however, is based on a series of cross-checks (in Claims 6 and 7), which only work when we begin

[10]If all agents present their tightest evidence, presenting an other-refutable lie is dominated by presenting the truth owing to τ^1. Presenting nonrefutable lies is also dominated by presenting the truth due to τ^2. Any self-refutable lie, which induces a evidence set tighter than that under the truth for other agents, is dominated by the truth due to τ^2, and then self-refutable lies are dominated by the truth due to τ^3.

[11]For instance, in a partial implementation exercise of Ben-Porath, Dekel, and Lipman (2019) (p. 545), they begin with the understanding that agents must present the maximal/tightest evidence associated with his state claim at the peril of large punishments.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

from the *correct* profile of evidence. In particular, the designer can incentivize the presentation of tightest evidence but is not able to directly incentivize the presentation of the true state.[12] In its simplest form, our mechanism takes the following position: "when there is something wrong with the message profile, prioritize obtaining the tightest profile of evidence over all else."This allows the cross-checks, which realign the profile toward the truth.

3.7 *The role of normality*

We now turn to the role of normality in our setup. When agents cannot submit arbitrary amounts of evidence, the necessary condition is obtained by Kartik and Tercieux (2012a) in their Propositions 2 and 3. First, for any state s, define by $T^f(s) = \{s' : f(s') \neq f(s)\}$ the set of states in which the desirable outcomes differ from $f(s)$. Recall that we wish to implement without relying on preference variation. In particular, under constant preferences, Proposition 3 in Kartik and Tercieux (2012a) requires that s and $T^f(s)$ be distinguishable, which means that either some agent can refute s at any state in $T^f(s)$ or refute every state in $T^f(s)$ at s using a single article of evidence. In Theorem 7 of Banerjee, Chen, and Sun (2023), we show that this condition is also sufficient for mixed-strategy implementation with two or more agents in a finite albeit indirect mechanism, which requires submission of *two* articles of evidence.

3.8 *Budget balance*

We will now establish that the above mechanism can be modified to achieve budget balance. The major challenge in achieving this goal is the redistribution of penalties to other agents without affecting the incentives of the recipient agents. In general, this cannot be achieved with two agents, as we will show in Section 5. In the following discussion, we consider a setting with at least three agents.

One way to transform a two agent unbalanced mechanism into a three agent balanced mechanism is to choose two agents at random to play the unbalanced mechanism (with the transfers for an agent appropriately scaled up to reflect the probability of being chosen) and redistribute the transfers to the third agent. While this approach is quite general, it requires a stochastic mechanism. In what follows, we provide modifications to the above mechanism, which achieves budget balance without randomization.

First, it is clear that τ^1, the transfer for the elimination of other-refutable lies is already budget balanced. The incentive to present all the evidence stemming from τ^4 is key to the removal of both self-refutable and nonrefutable lies. This incentive can be arbitrarily small, and is only active when some agent's claim is not supported by the other agents. Redistributing this small incentive to the agent whose state claim is unsupported does not affect his incentives since the penalty from his state claim not being supported is much larger. The second transfer τ^2, which penalizes an agent for his state claim being unsupported, is redistributed among the other agents, with a minor modification—τ_i^2 is redistributed evenly to only those other agents $j \neq i$ who have supported i's state

[12]This differs from the classical implementation problem with preference variation, where dictator lotteries ala Abreu and Matsushima (1994) can be used to elicit an agent's preference.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

claim s_i, and redistributed evenly to all agents if no other agents have supported s_i. The third transfer τ^3 is directly redistributed evenly to all agents. The resulting mechanism is therefore budget balanced. We refer the reader to Appendix A.1 for a discussion of how implementation is obtained under this modified mechanism.

3.9 *Implementation with small transfers*

While the transfers involved in the mechanism above have been imposed only off the equilibrium, the transfers are "large" since they need to dominate the agents' utility differences from outcomes. In reality, agents might not be willing or able to pay these fines. Here, we present a result for implementation with arbitrarily small off-the-equilibrium transfers, as long as the designer can randomize and there are at least three agents. To this end, we construct an indirect mechanism building on similar ideas from Abreu and Matsushima (1994). We begin by defining the appropriate notion of implementation prevalent in the literature for this case.

DEFINITION 4. An SCF is Nash implementable with arbitrarily small transfers if for any $\varepsilon > 0$, there is a mechanism $\mathcal{M} = (M, g, (\tau_i)_{i\in\mathcal{I}})$ such that for any profile of utility functions $v = (v_i)_{i\in\mathcal{I}}$, any state s, and any mixed-strategy equilibrium σ of the game $G(\mathcal{M}, v, s)$, we have $g(m) = f(s)$ and $\tau_i(m) = 0$ for each message profile $m \in \operatorname{supp}\sigma(s)$ and the total (off-the-equilibrium) transfer to any agent can be limited to being no greater than ε.

We now state the formal result as follows.

THEOREM 2. *Suppose the evidence structure is given by $\mathcal{E}_i(\cdot)$. If there are at least three agents, then an SCF f is Nash-implementable with arbitrarily small transfers if and only if it is measurable with respect to $\mathcal{E}_i(\cdot)$.*

Intuitively, this result is based on the following ideas. A lottery is used to *divide* the incentive for manipulating the outcome into K parts (called rounds), where K can be chosen as large as necessary to meet the transfer bound required. The first round uses the implementing mechanism with its transfers scaled down, since only a small part of the outcome is controlled by it. This allows for the true state to be revealed.[13] In each of the following rounds, agents are incentivized to agree with the unanimous first round report of the truth, failing which the first deviant is penalized an amount that is small enough that it meets the transfer bound, yet large enough that it dominates the incentive for manipulating the outcome of the round. This penalty only applies to the first round with disagreement, as repeating it will lead to a large transfer. It is sufficient for implementation, however, because no agent wants to be the first to deviate in any round. For further details, we refer the reader to the formal proof of Theorem 2 in Appendix A.2.

[13]Note that this is not dependent on the mechanism used. Any fully implementing mechanism can be adapted into this framework, a fact we will leverage later when studying other setups.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4. Costly evidence

So far, we have studied hard evidence, which corresponds to the notion that evidence, that is available to the agent is costless to present. In this section, we relax the assumption that presenting evidence is costless. Evidentiary costs are mentioned as an important area of further research in both (Bull and Watson (2007, footnote 10) and (Ben-Porath and Lipman (2012, p. 1714)).

With the above motivation, we derive an extension of our implementation result to a setting with costly evidence. More formally, the environment is the same as that in Section 2.1, except for the addition of a cost function $c_i : \mathcal{E}_i \times S \to \mathbb{R}_+$, which is bounded by a (possibly) large positive cost C. Here, we note that we allow the evidentiary cost to depend on the state.

There are two possible stances on the designer's knowledge of the cost structure. Either the designer does not know $c_i(\cdot)$, and only knows C (so that he is unable to exploit the variation of costs between states), or he knows $c_i(\cdot)$ as well (whereupon he can exploit the variation of cost among states). We treat these two cases separately in the following sections.

4.1 *Implementation regardless of cost variation*

In this section, we assume that $c_i(\cdot)$ is common knowledge among the agents but the designer only knows C. Then the designer cannot exploit cost variation. The definition of normality remains the same as in Section 2.1. Further, the notion of implementation remains the same as that in Section 2, so that the designer is indifferent to the cost of evidence submission. We obtain the following result.

Theorem 3. *An SCF f is Nash implementable in a direct revelation mechanism for every costly evidence structure $\mathcal{E}_i(\cdot)$ if and only if it is measurable.*

We provide the proof of Theorem 3 along with a detailed sketch in Appendix A.3. In the proof, we adopt the same direct revelation mechanism constructed in Section 3 but suitably adjust the relative scales of the four transfer rules. The proof requires a substantially different approach from that of Theorem 1. Due to the evidentiary cost, even after raising the transfers, lies can be eliminated only with high probability rather than probability one. This turns out to be sufficient to ensure that the agents present their tightest evidence to support the truthful state claim, which is presented with high probability. However, once the tightest evidence is presented, the agents can no longer lie in their state claims even with small probability, so that implementation is achieved.

4.2 *Implementation under cost variation*

In this section, we assume that $c_i(\cdot)$ is common knowledge among the designer and the agents. A mechanism therefore can depend on the cost structure to eliminate incorrect claims of state. This setup corresponds to that in Kartik and Tercieux (2012a). We emphasize that in this setup articles of evidence do not necessarily associate with subsets

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of the state space because of the element of cost variation. Following Kartik and Tercieux (2012a), we define the set of cheapest evidence in any state as $\mathcal{E}_i^l(s) = \arg\min_{E_i} c_i(E_i, s)$. We also normalize the costs so that the difference in costs between any two articles of evidence between any two states is less than 1 dollar. That is, $c(\cdot)$ is normalized such that $|c_i(E_i, s) - c_i(E_i, s')| < 1$ for any i, E_i, s, and s'. We present the notion of implementation we work with below.

DEFINITION 5. An SCF f is directly Nash-implementable in pure (resp., mixed) strategies if there is a mechanism $\mathcal{M} = (S \times \mathcal{E}, g, (\tau_i)_{i\in\mathcal{I}})$ such that for any profile of bounded utility functions $v = (v_i)_{i\in\mathcal{I}}$, any state s, any pure (resp., mixed) strategy Nash equilibrium σ of the game $G(\mathcal{M}, v, s)$, and any message profile (s, E) in the support of $\sigma(s)$:

(i) $g(m) = f(s)$ and $\forall i$, $\tau_i(m) = 0$;

(ii) $E \in \mathcal{E}^l(s)$

As in the hard evidence setting, implementation requires that for any profile of bounded utilities, and in each equilibrium of the implementing mechanism, the outcome be f-optimal and the transfer to each agent be zero. In addition, it also requires that only an article from the set of cheapest evidence be submitted in equilibrium. In this section, we focus on pure strategy equilibria, and discuss a treatment of mixed equilibria in Section 4.3.

4.2.1 *Evidence monotonicity* Kartik and Tercieux (2012a) establish that a condition called *evidence monotonicity* is necessary for implementation in the above setup using a mechanism, which only admits the submission of cheapest evidence in equilibrium.[14] Since we wish to implement while maintaining robustness to agents' utility functions, constant preferences is a possible scenario under which a mechanism must still implement. If we allow for transfers, then we obtain the following characterization of evidence monotonicity under our setting.

DEFINITION 6. An SCF f is evidence-monotonic under constant preferences if there exists $E^* : S \to \mathcal{E}$ such that:

(i) for all s, $E^*(s) \in \mathcal{E}^l(s, f(s))$

(ii) for all s and s', if $\forall i, t \in \mathbb{R}$, $E_i' : [-c_i(E_i^*(s), s) \geq t - c_i(E_i', s) \implies -c_i(E_i^*(s), s') \geq t - c_i(E_i', s')]$, then $f(s) = f(s')$.

Alternatively, if $f(s) \neq f(s')$, then $\exists i$, t, E_i' such that $c_i(E_i^*(s), s) \leq c_i(E_i', s) - t$ but $c_i(E_i^*(s), s') > c_i(E_i', s') - t$. This yields $c_i(E_i', s') - c_i(E_i^*(s), s') < c_i(E_i', s) - c_i(E_i^*(s), s)$. The implementing mechanism, which we present later will be direct. In context of a direct mechanism, we interpret i's action of submitting (E_i', s') instead of $(E_i^*(s), s)$ as a challenge to the state claim of s at s', and denote t as the (possibly negative) challenge

[14]Notice that Theorem 3 does not involve this restriction. We discuss the implications in Section 4.3.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

reward and E_i' as the challenge evidence. In essence, agent i credibly informs the designer that the state is not s by asking for an amount of money t for presenting an article of evidence E_i' instead of $E_i^*(s)$. This is profitable if the state is s' and not profitable if the state is indeed s. This implies that if i is to challenge s at s', then there is an article E_i', which has become cheaper relative to $E_i^*(s)$ in going from s to s'. We define $\mathcal{E}_i^\gamma(s, s')$ as the set of *challenge evidence* for agent i when challenging s at state s', and make an arbitrary selection $E_i^\gamma(s, s')$ from it, which is used in the implementing mechanism.

4.2.2 *A classification of lies* We begin with a classification of lies (for some true state s^*) and then use a direct mechanism to eliminate them in sequence. To do so, we first formalize the notion of refutability in this setup.

Definition 7. An agent i can challenge a state claim s when $\exists(s_i, E_i)$ such that $c_i(E_i, s_i) - c_i(E_i^*(s), s_i) < c_i(E_i, s) - c_i(E_i^*(s), s)$.

Note that there is a difference between the interpretation of challenge between the usual hard evidence setting and this setup. In the hard evidence setting, when an agent presents an article of evidence E, which does not contain a state s, he definitively refutes the state s. In this setting, however, articles of evidence cannot be associated with a subset of the state space in the same way, since the "reversal," which credibly signals to the designer that the state is not s requires the commitment of a certain sum of money in exchange for the challenge evidence. The inequality in the definition above assures us of the existence of a sum of money that enables this reversal. In essence, whereas refutability is a property of the setup under hard evidence, a mechanism is required for the same in this setup.

4.2.3 *Implementation* We now present our main result for this setting.

Theorem 4. *Suppose that $\mathcal{E}_i(\cdot)$ is a costly evidence structure. Then an SCF f is directly Nash-implementable in pure strategies in a direct revelation mechanism if and only if it is evidence-monotonic under constant preferences.*

We draw attention to a few interesting points regarding the above result here. First, we note that while evidence monotonicity is necessary irrespective of the mechanism used, the mechanism we present is direct. Second, Chen, Kunimoto, Sun, and Xiong (2021) contains an example, which establishes that it is not possible to obtain direct implementation of some Maskin-monotonic social choice functions where there are only two agents.[15] The impossibility essentially derives from the difficulty of figuring out which agent is challenging the other when two agents disagree in their state claims. The presence of evidence allows us to bypass this difficulty.[16] Third, while the result states that evidence monotonicity under constant preferences is necessary, we allow for variation of preferences in the proof of Theorem 4. The necessity of this condition arises out of our desire to achieve implementation regardless of preference variation. If the designer cannot exploit preference variation, implementation must obtain from variations in the cost of evidence.

[15]The example discusses rationalizable implementation, but it also applies to Nash implementation.

[16]See Lemma 2 in Appendix A.4 for more details.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4.2.4 *Implementation in mixed strategies* For implementation in mixed-strategy Nash equilibrium, which accounts for general cost variation (Definition 5), we also present an alternative treatment in Banerjee, Chen, and Sun (2023). In this treatment, we impose a stronger condition than evidence monotonicity under constant preferences. Specifically, we require that for any two states s and s' with distinct social outcomes that there be at least one agent for whom an article of evidence that was not cheapest under s is now cheapest under s'. We term this condition evidence monotonicity* and prove mixed-strategy implementation (also in a direct mechanism) under evidence monotonicity*; see Theorem 6 of Banerjee, Chen, and Sun (2023).[17] The result complements Theorem 3 in requiring that only one article of cheapest evidence be submitted in equilibrium, but not normality.[18]

4.3 *Evidence monotonicity versus measurability*

Theorems 3 and 4 approach costly evidence in different ways. It is natural to ask how these results compare. In the costly evidence setting, some articles, which are unavailable, are considered to have an infinite cost. In such a setting, measurability equates to the requirement that the set of evidence with finite cost must change between states with different social outcomes. This requirement is strictly stronger than evidence monotonicity. To see this, consider Example 2 of Kartik and Tercieux (2012a) (henceforth KT), which derives from agents having a preference for honesty. The example can be viewed as a case of costly evidence in which all states have the same evidence sets, but the cheapest evidence is distinct in each state. KT show that this makes any social choice function evidence monotonic. However, since there is no variation in the set of evidence with finite cost, only constant social choice functions are measurable.

KT show in their Corollary 4 that evidence monotonicity coincides with measurability in a hard evidence structure (where evidence is costless when it is available), which satisfies normality.[19] In contrast, our Theorem 3 allows for hard evidence, which is available and yet has a positive cost. In this setting, even when the evidence structure satisfies normality, there can still be social choice functions, which satisfy measurability but not evidence monotonicity (under constant preferences). See Appendix A.4.3 for a detailed description. Such SCFs can therefore be implemented according to Theorem 3 but not Theorem 4.

The discrepancy arises because the implementation notion in Theorem 4 requires that only one article of cheapest evidence be submitted in equilibrium. Apparently,

[17] Banerjee, Chen, and Sun (2023) also show that if we require the implementing mechanisms to use only penalties or arbitrarily small rewards to the agents, then evidence monotonicity* becomes necessary for implementation; moreover, our implementing mechanism in Theorem 6 of Banerjee, Chen, and Sun (2023) indeed satisfies the requirement.

[18] We also conjecture that Theorem 3 can be established without requiring normality, by making use of an indirect mechanism akin to the implementing mechanism in the proof of Theorem 7 in Banerjee, Chen, and Sun (2023).

[19] Note that this only holds among pairs of states, which do not satisfy Maskin monotonicity, for instance with state independent preferences since otherwise, evidence monotonicity may be satisfied via preference variation.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the requirement must be associated with a fixed costly evidence structure and thereby muted when we demand implementation regardless of evidentiary cost variation, as we do in Theorem 3. In this regard, Theorem 3 is a step toward answering a question which KT pose as to which social choice functions could be implemented if the designer allowed for the presentation of costly articles of evidence to elicit information from the agents.[20]

5. Incomplete contracts and renegotiation-proofness

In this section, we apply our results under hard evidence to revisit a classical issue in contract theory—that of bilateral contracting with observable but unverifiable information. This problem arises when two parties wish to condition a contract on certain state variables, which are commonly observable, but not verifiable by a third party, such as a court. Attempting to directly condition a contract on these variables runs into difficulties, because whoever is responsible for enforcing the contract may not be able to ascertain which state occurred, and thus may also be unable to resolve any disputes between the agents.

Implementation theory deals with issues of contract design in such situations by exploiting the idea that contracts can be made contingent on the messages reported by the parties to make the observable state verifiable. In particular, by designing suitable revelation mechanisms for the contracting parties, it is often possible to achieve the same outcomes as those arrived at with fully contingent contracts. However, most papers in this literature use mechanisms of the Moore–Repullo type (see, e.g., Maskin and Moore (1999), Maskin and Tirole (1999), and Maskin (2002)), which are vulnerable to renegotiation among the agents, because they involve penalizing both agents in certain off-equilibrium paths. In general, the possibility of renegotiation significantly limits the set of implementable outcomes. For instance, Maskin and Moore (1999) (henceforth MM) considers an example in which a seller (she) owns a product which a buyer (he) is interested in obtaining. The seller has an option to make an investment in the product at cost c to raise its value from ϕ to θ for the buyer. The investment is efficient (i.e., $c < \theta - \phi$), observable by the two parties, but unverifiable. In this setting, when the cost of the investment is more than half of the value it adds to the good, MM establish that the only renegotiation-proof contract is a null contract. This is because the buyer refuses to accept the good and renegotiates the price outside of the contract. That is, it is not possible to incentivize efficient investment in this scenario.

In the context of the example, it becomes clear why nonverifiability poses an issue. If the investment were verifiable, a mechanism could fine the buyer for refusing to trade after the investment had been made, and transfer the fine to the seller, thus preventing the buyer from lying, and indeed also preventing renegotiation. Even without verifiability, if the seller could *prove* that she has indeed made the investment or refute the claim that she has not made the investment, a court could fine the buyer (if he refuses to trade) and reward the seller based on this proof. This leading example motivates our

[20] (Kartik and Tercieux (2012a, p. 349)) cite screening as an example for why it might be interesting to allow for costly evidence provision in equilibrium.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

choice of using the hard evidence results in this context, as it is more natural that evidence in contracting situations will take the form of subsets of the state space, refuting alternative states. Arguably, it is not possible to prove an investment, which one has not made, at whatever cost. This immediately leads to the question—what conditions must the hard evidence structure satisfy so that renegotiation-proof contracts can be made?

To address this question, we begin by formalizing the notion of renegotiation. Following MM, we assume A is a finite set, and define T as the set of transfers with budget surplus.[21] With this, we define the renegotiation process via a renegotiation function $h : A \times T \times S \to A \times T$, where T defines the space of transfers to the I agents. This function can be thought of as a transformation on the outcome and profile of transfers of a mechanism $\mathcal{M} = (M, g, (\tau_i)_{i\in\mathcal{I}})$, which occurs before the agents evaluate the outcome using their utility functions u_i. That is, given a mechanism $\mathcal{M}$, agents submit their messages to the mechanism, which yields an outcome and a set of (budget surplus) transfers; agents then renegotiate this combination of outcomes and transfers to another combination, and then evaluate their utilities. We think of this in terms of the game $G(\mathcal{M}, v, h, s)$, which differs from the game $G(\mathcal{M}, v, s)$ in that instead of the outcome being defined as $g(m)$ and the transfer profile as $\tau(m)$, the outcome is defined as $h^a(g(m), \tau(m), s)$ and the transfer profile as $h^t(g(m), \tau(m), s)$. Now, we define the notion of efficiency for an allocation.

Definition 8. An allocation $(a, (t_i)_{i\in\mathcal{I}})$ is efficient with respect to a profile of utilities $v = (v_i)_{i\in\mathcal{I}}$ at state s if there does not exist $(\hat{a}, \hat{t}) \in A \times T$ such that for every agent i, $u_i(\hat{a}, s, \hat{t}_i) \geq u_i(a, s, t_i)$ with strict inequality for some i.

Following MM, we make the following three assumptions about the renegotiation function h. First, we assume that the renegotiation function h is *predictable*, which essentially amounts to saying that h is common knowledge among agents and deterministic. Second, we assume that h is *individually rational*, i.e., if at every state, all agents weakly prefer the renegotiated outcome to the original one. This is a natural restriction as no agent can be forced into renegotiation. Finally, we assume that renegotiation is *efficient*, which means that $h(\cdot, \cdot, s)$ results in efficient allocations at every s.

In what follows, we will constrain the scope of the discussion to combinations of f, v, and h such that $f(s)$ is efficient with respect to v, and h satisfies the properties described above.

Definition 9. A social choice function f is implementable with renegotiation in Nash equilibrium if there is a mechanism $\mathcal{M} = (M, g, (\tau_i)_{i\in\mathcal{I}})$ such that for any state s, any profile of utilities v, and any mixed-strategy Nash equilibrium $\sigma = (\sigma_i)_{i\in\mathcal{I}}$ of the game $G(\mathcal{M}, v, h, s)$, we have $h(g(m), \tau(m), s)(\hat{a}) = f(s)$ and $h(g(m), \tau(m), s)(\hat{t}) = 0$ for every message profile m on the support of σ.

[21] That is, $T = \{t \in \mathbb{R}^I \text{ such that } \sum_i t_i \leq 0\}$. Since this is a contract between agents, it is not possible to find money outside the contract to finance a budget deficit.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In the spirit of this paper, we require that implementation obtain regardless of the utility functions v (subject to the constraints mentioned above). We now turn to characterizing the necessary and sufficient conditions for renegotiation-proof implementation.

THEOREM 5. *Assume that $I = 2$ and $\mathcal{E}$ is the evidence structure. An SCF f is implementable with renegotiation in Nash equilibrium if and only if for any pair of states s and s' in S such that $f(s) \neq f(s')$, one of the following is true:*

(a) There is one agent who can refute s' at s and s at s'; or

(b) Both agents can refute s' at s and neither of them can refute s at s'.

For the formal proof, we refer the reader to Appendix A.5.1.

To illustrate the above conditions, we consider an alternate evidence structure in the example from MM above—what if the buyer is the only agent who can prove that the investment was made (or not)? In practice, this may often be the case, for instance if the buyer has some sort of (private) suitability test, which can check if the investment has been made. Theorem 5 tells us that a renegotiation-proof contract exists in this case. This is because the burden of proving that the investment was not made also falls to the buyer, and this is impossible when the investment has been made. This, in turn, means that any set of prices can be implemented with this evidence structure (note that any price is Pareto efficient), including, of course, the set of prices, which recoup the cost of the investment for the seller, thereby incentivizing her to make the efficient investment. In this context, we refer the reader to Appendix A.5.2 for an example of an evidence structure, which satisfies measurability but not the conditions of Theorem 5. While Conditions (a) and (b) are fairly demanding in this example, requiring a fair bit of provability by the buyer, they are nevertheless *necessary*, suggesting that renegotiation-proof implementation is significantly more challenging than the notion in Definition 3.

There are a few points of interest we wish to clarify here. First, we note that the characterization in Theorem 5 is stronger than measurability. Indeed (e1) of Definition 1 and either one of Conditions (a) or (b) yields that the agent who can refute s at s' has an article of evidence at s, which he does not have at s'. Then f trivially satisfies measurability. Second, the mechanism proposed in the proof of Theorem 5 needs only a direct revelation mechanism to work, and at least one of the Conditions (a) or (b) are necessary irrespective of the mechanism used.

We discuss the intuitions behind the conditions presented in Theorem 5 now. We prove the necessity of the conditions by constructing a pair v, h under which (a) at least one of the agents has incentive to lie, and (b) the agent in question *can* lie, regardless of the implementing mechanism.

In proving sufficiency, we construct a direct revelation mechanism with two transfers: a penalty of 1 dollar for agent i if he is unable to support his state claim and a penalty of 2 dollars if his state claim is refuted by the evidence which agent $j \neq i$ presents. Each fine is paid to the other agent, so that the mechanism has a balanced budget.[22]

[22]Recall that the maximum bound of utility is less than 1 dollar, so that these transfers dominate the utility from the outcome.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Assume that the true state is s^*. First, no agent reports other-refutable lies with positive probability, because the other agent presents the refuting evidence with probability 1; this loses the first agent 2 dollars, which is more than the utility difference from the outcome and the other transfer. Second, if an agent reports a self-refutable (resp., nonrefutable) lie s with positive probability, then Condition (a) (resp., Condition (b)) of Theorem 5 implies that this agent can also refute s^* at s. Therefore, he cannot support his state claim and will be fined 1 dollar. Since this is more than the utility difference from the outcomes, deviating to $(s^*, E_i^*(s^*))$ is profitable. Therefore, the only equilibrium is truth-telling.

To summarize, we have established the necessary and sufficient condition for renegotiation-proof Nash implementation in settings with hard evidence in the language of refutability. This allows us to answer an important question—it is well known that ex post verifiability is sufficient for a renegotiation-proof contract, but "how much provability" is really required to ensure that such contracts exist? Theorem 5 offers an answer.

6. Related literature

This paper contributes to the literature on implementation with evidence. For early work in this area, we refer the reader to Green and Laffont (1986), Bull and Watson (2007), and Deneckere and Severinov (2008). See also Deneckere and Severinov (2007) for a study of partial implementation with costly evidence and Kartik and Tercieux (2012a) for a survey of other early works on implementation with evidence.

Our implementation exercise is most closely related to Ben-Porath and Lipman (2012) and Kartik and Tercieux (2012a). Ben-Porath and Lipman (2012) present two main results. First, they achieve subgame-perfect Nash implementation using a perfect-information mechanism with large off-the-equilibrium transfers (which achieve budget balance when there are three or more agents). This result does not require normality or integer games. Then Ben-Porath and Lipman (2012) achieve Nash implementation with three or more agents by using small off-the-equilibrium transfers. This latter result does require normality and integer games. In contrast, our Theorem 1 requires normality, works with two agents, and employs a direct revelation mechanism to achieve (mixed-strategy) Nash implementation without integer games but with large off-the-equilibrium transfers (which also achieve budget balance when there are three or more agents). Neither Theorems 1 or 2 of Ben-Porath and Lipman (2012) account for evidentiary cost which our Theorem 3 deals with.[23]

Kartik and Tercieux (2012a) studies the costly evidence setting. They use a canonical, Maskin-style mechanism which relies on integer games and does not use transfers. With three or more agents, their achieve pure-strategy Nash implementation for every evidence-monotonic SCF.[24] Their evidence monotonicity notion allows for preference

[23]In their Theorems 4 and 5, Ben-Porath and Lipman (2012) show that their results are robust to a variety of preference and belief specifications. Our results are consistent with these, in the sense that the designer does not need to know anything about the preferences of the agents except for the uniform bound on utilities, although we maintain the complete-information assumption throughout.

[24]Their Theorem 2 can be modified to account for mixed Nash equilibria as well, using methods from Kartik and Tercieux (2012b) but this depends on integer games as well.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

variation and reduces to Maskin-monotonicity when all messages are cheap talk. In contrast, Theorem 4 makes use of a direct revelation mechanism without integer games to achieve pure-strategy implementation with two or more agents. Further, to focus on evidentiary cost variation, we adopt the notion of evidence monotonicity under constant preferences under which our implementation result disregards (and therefore is also robust to) preference variation.

Another related strand of literature is that of implementation with preferences for honesty, wherein it is assumed that agents prefer to tell the truth if they do not gain from lying. Preferences for honesty can be viewed as a case of costly evidence in which all states have the same evidence sets, but the least-cost evidence is distinct in each state. Kartik, Tercieux, and Holden (2014) establish that with two or more agents, preferences for honesty and a condition called separable punishments (which generalize off-the-equilibrium transfers), any SCF can be implemented in a finite but indirect mechanism without integer or modulo games. Dutta and Sen (2012) also establish that with preferences for honesty all social choice correspondences, which satisfy no veto power can be implemented by a mechanism that uses integer games. Our Theorems 3 and 4 apply to a general costly evidence structure beyond the specific setting of preference for honesty and both invoke only direct revelation mechanisms.[25]

Chen, Kunimoto, Sun, and Xiong (2022) also provides an implementation result using transfers that allows for implementation of Maskin-monotonic social choice functions without using integer games. A main focus of our exercise though is to handle implementation with state-independent or consistent preferences by making use of evidence. Our emphasis on evidence also allows for the novel classification of lies, which we propose and in turn allows for the new approach to implementation that we present here. This is a feature unavailable for implementation exercises, which solely rely on preference variation. Moreover, our treatment of the two-agent case extends to dealing with renegotiation (wherein we provide a necessary and sufficient condition for renegotiation-proof implementation), a pertinent issue when two-agent implementation is used in contracting.

7. Conclusion

In this paper, we present full implementation results in settings with hard and costly evidence. In the hard evidence setting, using a novel classification of lies according to their refutability, we construct a direct revelation mechanism, which implements any measurable SCF with respect to the evidence structure. Our mechanism invokes neither integer nor modulo games, requires only two agents, accommodates evidentiary costs, and can also be modified to account for limited solvency of the agents. Based on our classification of lies, we also derive a necessary and sufficient condition for the existence of bilateral renegotiation-proof contracts (Maskin and Moore (1999)). In the costly evidence setting, we provide a mechanism that yields pure strategy implementation in a two-agent setting without using integer or modulo games. The classification of lies approach is used here as well.

[25] As Kartik and Tercieux (2012a) point out, as long as in any state at least one agent has a preference for honesty, any SCF is evidence monotonic.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Our exercise leaves a number of open questions for future research. In particular, Banerjee and Chen (2022) pursue an extension of our exercise to an incomplete information setting with a commonly known state and uncertainty among agents along the evidence dimension. Other directions include direct implementation without transfers or structural assumptions, which serve the same role.

Appendix A

In this Appendix, we provide the details and proofs, which are omitted from the main body of the paper.

A.1 *Budget balance*

Since τ^1 is budget balanced at the outset, Claims 1 and 2 are not affected by the modifications. Lemma 1 continues to hold, because the only possible scenario under which providing additional evidence may cause a loss to an agent i (owing to the redistribution of transfers) is when providing additional evidence supports an agent j's state claim s_j and, therefore, causes losses from the redistribution of τ_j^2. However, for this to be the case, agent k must then be supporting s_j. In this case, agent i does not actually receive any part of the redistributed revenues from τ_j^2.

Claims 3 and 4 remain unaffected. If an agent i presents a nonrefutable lie, we establish that all other agents still present the tightest evidence. First, since τ^4 is only redistributed to i, it does not affect other agent's incentives. Further, an agent is only unwilling to support another agent (to avoid redistributive losses from τ^2) if he was the only agent who did not support s_i (so that in supporting i, he would switch off τ_i^2). But, due to the way that τ_i^2 is redistributed, and the existence of a third agent, he does not share in the redistribution of τ_i^2 in this case. Then agent i still faces a large fine (the redistributed revenue from τ^4 is small) and chooses to deviate to $(s^*, E_i^*(s^*))$.

Claims 5 through 7 continue to hold as well. Since agents are limited to presenting either the truth or self-refutable lies, from the above argument, all evidence must still be tightest. If an agent i plays a self-refutable lie, which implies a smaller evidence set for j than the truth, then it is not supported, and agent i prefers to deviate to the truth as above. Therefore, all claims are consistent with other agents' tightest evidence. The third transfer, τ^3 is a penalty for an agent i for implying a different evidence set for himself than that implied by others for him. His incentive for avoiding this penalty is not affected by its redistribution to other agents. Therefore, with three or more agents, this mechanism implements with budget balance.

A.2 *Proof of Theorem 2*

A.2.1 *Message space* The message space is augmented with $K+1$ additional claims of state, where we call K the number of rounds and it is chosen according to the upper bound of allowable transfers. More precisely, the message space is as follows:

$$m_i = \left(s_i^0, E_i, s_i^1, s_i^2, \ldots, s_i^{K+1}\right) \in M_i = S \times \mathcal{E}_i \times S \times \cdots \times S \quad (\text{K} + 1 \text{ times}).$$

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

A.2.2 *Outcome* The outcome is specified as follows. Define $\rho^k(m)$ (for $k = 2, \dots K+1$) as follows:

$$\rho^k(m) = \begin{cases} f(s) & \text{if } \exists s \text{ such that } \left|\left\{j : s_j^k = s\right\}\right| \geq I - 1 \\ b & \text{otherwise} \end{cases}$$

where b is an arbitrary lottery over A. Then the outcome of message profile m, denoted by $\bar{g}(m)$ is defined as follows:

$$\bar{g}(m) = \varepsilon \times g\left(s^0, E\right) + (1-\varepsilon) \times \frac{1}{K} \sum_{k=2}^{K+1} \rho^k(m)$$

where $\varepsilon > 0$ is chosen to be small, and $g(\cdot)$ is the outcome function of the base (unaugmented) mechanism.

That is, the outcome is a lottery combining the outcome of the base mechanism using the zeroth and evidence reports, and an outcome function defined for each round, which is the outcome corresponding to a state on which at least $I-1$ agents have agreed or (if there is no such agreement), some random lottery b.

A.2.3 *Transfers* The mechanism multiplies all the transfers of the baseline mechanism by ε and adds the following transfers:

$$\tau_i^5(m) = \begin{cases} -\alpha & \text{if } s_i^1 \nsim s_{i+1}^0; \\ 0 & \text{otherwise.} \end{cases}$$

That is, agent i receives a fine of α if their first report is not identical to agent $i+1$'s zeroth report.

$$\tau_i^6(m) = \begin{cases} -\beta & \text{if } \exists s \text{ such that } s = s_i^1 \ \forall i \text{ and } s_i^k \neq s \text{ and } s_n^m = s \ \forall m, n < k; \\ 0 & \text{otherwise.} \end{cases}$$

That is, agent i receives a fine of β if their k^{th} report is the first deviation from a unanimous first report.

$$\tau_i^{7,k}(m) = \begin{cases} -\gamma & \text{if } \exists s \text{ such that } s_i^k \neq s \text{ and } s_j^k = s \ \forall j \neq i; \\ 0 & \text{otherwise.} \end{cases}$$

That is, agent i receives a fine of γ if their k^{th} report is the only deviation in round k from an otherwise unanimous set of reports. This fine is applied to each round.

A.2.4 *Proof of implementation* In the following proof, we do not directly prove implementation for any values of α, β, γ, and K. Rather, we show that there exist values of these parameters such that the mechanism implements and the overall transfer to any agent can be bounded below an arbitrarily small number, which is greater than zero.

Claim 9. *In any equilibrium, for every agent i, s_i^0 is the truth.*

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

PROOF. It is clear that any agent's reports with index 0 and the evidence messages only affect their payoffs through the outcome. While all transfers in the base mechanism are scaled by ε, the maximum utility value of manipulating the outcome is also scaled by ε (owing to the randomization in the final outcome), so that in this mechanism, the reports with index 0 are indeed the truth. □

CLAIM 10. *There exist values of α, β, and γ such that in any equilibrium, for every agent i, s_i^1 is the truth.*

PROOF. Suppose not. That is, $\exists i$ such that $s_i^1 \neq s^*$. From Claim 9, we know that $s_i^0 = s^*$ $\forall i$. Then consider a deviation $s_i^1 \leftarrow s^*$. Notice that s_i^1 does not affect the outcome. Then the agent gains at least α from τ^5, and could lose up to β to τ^6. There is no effect from τ^7. If $\alpha > \beta$, then there is a profitable deviation. □

CLAIM 11. *There exists values of α, β, γ, and K such that in any equilibrium, for every agent i, s_i^k, $k = 2, \ldots K+1$ is the truth.*

PROOF. The proof proceeds by induction. First, we prove that $s_i^2 = s^*$ $\forall i$. Suppose not. Then there is an agent i for whom $s_i^2 \neq s^*$ and $s_j^2 = s^*$ $\forall j < i$. That is, agent i is the first deviant from the unanimous report of s^* in s^1. Now, there are two cases.

Case 1: $\exists j \neq i$ such that $s_j^2 \neq s^*$. Then agent i is the first deviant, but there are other deviants in s^2. Consider a deviation to $s_i^2 \leftarrow s^*$. This deviation could cause him a loss of up to $\frac{1}{K}$ from the outcome, but yields a profit of β due to τ^6. In case the agents had unanimously agreed on a state in s^2, this could also lead to a loss of at most γ from τ^7. Thus, as long as $\beta > \frac{1}{K} + \gamma$, this is a profitable deviation.

Case 2: $\nexists j \neq i$ such that $s_j^2 \neq s^*$. That is, agent i is the first and only deviant in s^2. Consider again the deviation $s_i^2 \leftarrow s^*$. There is no change to the outcome (as agent i was the only agent who was disagreeing with the otherwise unanimous true report), and even though agent i could be the first deviant in further rounds, so that this deviation does not necessarily yield any advantage from τ^6, it does yield a profit of γ owing to τ^7. Thus, we need $\gamma > 0$ to make this a profitable deviation.

Thus, $s_i^2 = s^* \forall i$. Clearly, then this argument can be used inductively for the following rounds to establish that $s_i^k = s^* \forall i, k$. □

CLAIM 12. *For any $\bar{\delta} > 0$, the overall transfer to an agent can be bounded below $\bar{\delta}$.*

PROOF. Consider any $0 < \delta < \bar{\delta}$. From the previous claims, the inequalities required to be satisfied are $\gamma > 0$, $\beta > \frac{1}{K} + \gamma$, and $\alpha > \beta$, while the largest possible transfer is $\alpha + \beta + K\gamma$. Consider the following choices: $\gamma = \frac{\delta}{3K}$, $\beta = \frac{1}{K} + \frac{\delta}{3K}$, and $\alpha = \frac{\delta}{3}$. It is clear that K can be chosen to satisfy the required inequalities, and $\alpha + \beta + K\gamma < \bar{\delta}$. □

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

A.3 *Proof of Theorem 3*

The proof is by construction of an implementing mechanism. Below, we present the mechanism and a formal proof of implementation. But first, we formally state the following property of the cost function:

$$\max_{s\in S}\max_{i\in\mathcal{I}}\max_{E_i\in\mathcal{E}_i} c_i(E_i,s) < C$$

which states that across all states, all agents and all articles of evidence, the cost of evidence is bounded by C, a positive number.

A.3.1 *Message space and outcome* The message space and outcome function remain unaltered from the original implementing mechanism.

A.3.2 *Transfers* Structurally, the transfers are similar (with the addition of a transfer penalizing disagreement), however, the amounts are not fixed a priori, rather we show the existence of transfers so that the mechanism implements. We have then the following transfers:

$$\tau_{ij}^1(m) = \begin{cases} -T_1, & \text{if } s_i \in E_j \text{ and } s_j \notin E_i; \\ T_1, & \text{if } s_i \notin E_j \text{ and } s_j \in E_i; \\ 0, & \text{otherwise.} \end{cases}$$

$$\tau_i^2(m) = \begin{cases} -T_2, & \text{if } \exists j \in \mathcal{I} \text{ such that } E_j \nsubseteq E_j^*(s_i); \\ 0, & \text{otherwise.} \end{cases}$$

$$\tau_{ij}^3(m) = \begin{cases} -T_3, & \text{if } E_i^*(s_i) \neq E_i^*(s_j); \\ 0, & \text{otherwise.} \end{cases}$$

$$\tau_i^4(m) = \begin{cases} -\dfrac{T_4}{|S|}|E_i|, & \text{if } E_j \nsubseteq E_j^*(s_{j'}) \text{ for some } j, j' \in \mathcal{I}; \\ 0, & \text{otherwise.} \end{cases}$$

The overall transfer to agent i is given (as before) by

$$\tau^i = \sum_{j\neq i} \tau_{ij}^1 + \tau_i^2 + \sum_{j\neq i} \tau_{ij}^3 + \tau_i^4$$

A.3.3 *Proof of implementation* In what follows, we assume that the true state is s^*. We pick T_1, T_2, T_3, and T_4 so that the following inequalities are satisfied:

$$T_1 > \frac{C|S|}{\varepsilon} \tag{1}$$

$$T_1 \geq 1 + T_2 + (I-1)T_3 + T_4 \tag{2}$$

$$T_2 \geq 1 + (I-1)T_3 \tag{3}$$

$$T_3 \geq 1 \tag{4}$$

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$T_4 > \frac{C|S|}{1-\varepsilon} \tag{5}$$

It is immediate that this system of inequalities has a feasible solution. For instance, first choose $\varepsilon \in (0, 1)$, followed by setting $T_4 = \frac{C|S|}{1-\varepsilon} + \varepsilon$. Clearly, inequality (5) is satisfied. Second, set T_3, T_2, and T_1 in order so that inequalities (4), (3), and (2) are satisfied. Before we proceed to present the proof, we first provide a sketch to outline the augments.

First, we can prevent the agents from reporting other-refutable lies with a probability ε or more, as the reward for refutation, T_1, can be made large enough that doing so with probability ε (or more) guarantees refutation by other agents. This guaranteed by inequality (2) and established in Claims 13–14. With this, we are able to establish that all agents present their tightest evidence (Claim 16). This is achieved by choosing T_4 high enough so that the necessity of supporting the other agent's state claims overrides the evidence cost. Notice that owing to Observation 1, every state report except an other-refutable lie requires the presentation of tightest evidence to support. Second, agents do not present nonrefutable lies, because of the penalty from τ^2, which is easily avoided by deviating to the truth (since all evidence is the tightest). This is outlined in Claims 17 and 18. Third, we establish that any self-refutable lies, which are presented must be consistent with the tightest evidence of all other agents (Claim 19) and thereby eliminate the possibility that any self-refutable lies are presented in equilibrium (Claim 20). From the preceding logic, it is clear that if agents present their tightest evidence, then no agent presents other-refutable lies, and thus implementation obtains.

We now present below a formal proof of implementation using the mechanism stated above. In what follows, we denote the true state by s^*. Fix an arbitrary mixed-strategy Nash equilibrium σ.

Bounding the probability of other-refutable lies

Claim 13. *If agent i reports with probability at least $\frac{\varepsilon}{|S|}$ a state claim s_i such that agent $j \neq i$ has an article of evidence which refutes s_i, then agent j must refute s_i with probability* 1, *i.e., E_j refutes s_i for every $m_j = (s_j, E_j)$ on the support of σ_j.*

Proof. Presenting this evidence nets agent j a reward of at least $\frac{\varepsilon}{|S|}T_1$, and costs at most C. Thus, if $T_1 > \frac{C|S|}{\varepsilon}$, then this is a profitable deviation. This is a consequence of inequality (1). □

Claim 14. *Each agent reports other-refutable lies with a total probability less than ε.*

Proof. Suppose not. Then there is an agent i who reports a message m_i with a probability $\frac{\varepsilon}{|S|}$ or more such that agent $j \neq i$ has an article of evidence E_j in $\mathcal{E}_j(s^*)$, which refutes s_i. From Claim 13, agent j presents E_j with probability 1. Then the following table shows agent i's payoff changes from switching to the truth:

g	τ_i^1	τ_i^2	τ_i^3	τ_i^4
> -1	$\geq T_1$	$\geq -T_2$	$\geq -(I-1)T_3$	$\geq -T_4$

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Since $\varepsilon < 1$, it follows from inequality (2) that $T_1 \geq 1 + T_2 + (I-1)T_3 + T_4$. Hence, this constitutes a profitable deviation. □

CLAIM 15. *If all other agents present the tightest evidence with probability one, then no agent reports an other-refutable lie.*

PROOF. Suppose not. Then there is an agent i who reports a message m_i such that agent $j \neq i$ has an article of evidence E_j in $\mathcal{E}_j(s^*)$, which refutes s_i, while agent j presents this article of evidence with probability one. Then the following table shows agent i's payoff changes from switching to the truth:

g	τ_i^1	τ_i^2	τ_i^3	τ_i^4
> -1	$\geq T_1$	$\geq -T_2$	$\geq -(I-1)T_3$	$\geq -T_4$

Inequality (2) implies that this is a profitable deviation. □

Bounding the probability of nonrefutable lies

CLAIM 16. *Every agent presents their tightest evidence with probability one.*

PROOF. Consider an arbitrary agent i. From Claim 14, other agents are presenting state claims, which agent i cannot refute with a probability $1-\varepsilon$ or more. From Observation 1, these claims require agent i to present his tightest evidence so that they can be supported. Therefore, on any message where agent i does not present his tightest evidence, he expects τ^4 to be active with probability $1-\varepsilon$ or more. Since $\frac{(1-\varepsilon)T_4}{|S|} > C$ (inequality (5)), it is a profitable deviation to present the tightest evidence on such messages. □

CLAIM 17. *If all agents present their tightest evidence, then no agent reports nonrefutable lies.*

PROOF. Suppose not. Then there is an agent i who reports a message m_i which contains a nonrefutable lie while all agents present their tightest evidence. In any message $m_i = (s_i, E_i^*(s^*))$ where s_i is a nonrefutable lie, consider a deviation to the truth. Then the following table shows agent i's payoff changes from switching to the truth:

g	τ_i^1	τ_i^2	τ_i^3	τ_i^4
> -1	0	T_2	$\geq -(I-1)T_3$	≥ 0

First, the agent loses less than 1 from changing the outcome, incurs no loss from τ^1 (as the truth is irrefutable) and at most $(I-1)T_3$ from τ^3. The agent incurs no losses from τ^4 either as s_i was unsupported but the truth is supported and the size of his evidence set has not changed. Since all the presented evidence is the tightest, it follows that the agent gains at least T_2 from τ^2 from the deviation (as the truth is supported by all agents). It follows from inequality (3) that this is a profitable deviation. □

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

CLAIM 18. *No agent reports nonrefutable lies.*

PROOF. This follows immediately from Claims 16 and 17. □

Bounding the probability of self-refutable lies

CLAIM 19. *For any agent i, a self-refutable lie s_i is reported with positive probability only if $E_j^*(s_i) = E_j^*(s^*)$ for every agent $j \neq i$.*

PROOF. Suppose to the contrary that agent i reports $s_i \in SRL_i$ such that $E_j^*(s_i) \neq E_j^*(s^*)$ for some j. Then, by Observation 1, $E_j^*(s_i) \subset E_j^*(s^*)$. Consider then an alternative message, which replaces s_i with the truth. The following table shows agent i's payoff changes:

g	τ_i^1	τ_i^2	τ_i^3	τ_i^4
> -1	0	$\geq T_2$	$\geq -(I-1)T_3$	≥ 0

First, the agent loses less than 1 to the outcome. Second, the truth is not refutable so that the agent incurs no losses to τ^1. Third, we establish that the agent gains T_2 on account of τ^2. From Claim 16, all agents present the tightest evidence with probability one. Then the truth incurs no penalty while s_i incurs a penalty since $E_j^*(s_i) \subset E_j^*(s^*)$ and, therefore, s_i was not supported. Fourth, the agent loses at most $(I-1)T_3$ to τ^3, as other agents may not be reporting state reports consistent with the truth in i's evidence, i.e., it is possible that $E_i^*(s_j) \neq E_i^*(s^*)$). Fifth, since all other agents report the tightest evidence, the truth is supported, so that there is no loss from τ^4. Inequality (3) then implies that this is a profitable deviation. □

CLAIM 20. *Agents do not report self-refutable lies.*

PROOF. Suppose not. That is, suppose that there is an agent i who reports a self-refutable lie s_i. Consider a deviation to the truth for agent i (from Claim 16, evidence presentation was already the tightest). Then the following table shows agent i's payoff changes from this deviation:

g	τ_i^1	τ_i^2	τ_i^3	τ_i^4
> -1	0	≥ 0	$\geq (I-1)T_3$	≥ 0

First, the agent loses less than 1 to the outcome. Second, the truth is not refutable so that the agent incurs no losses to τ^1. Third, the deviation causes no loss from τ^2 since every agent presents their tightest evidence and the truth is supported. Fourth, since every agent presents their tightest evidence, no other-refutable lies and nonrefutable lies are presented (Claims 15 and 17). Then a self-refutable lie disagrees with all the reports (the truth and self-refutable lies) of other agents in agent i's evidence (Observation 1), and hence incurs a penalty from τ^3. The truth avoids this fine against every message, and thus yields a profit of $(I-1)T_3$. Fifth, the truth is supported, so that τ^4 cannot lead to any losses either. Inequality (4) yields that this is a profitable deviation. □

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Thus, the only equilibria are such that all agents report the truth with probability $1-\varepsilon$ or more, where $0<\varepsilon<1$ as chosen earlier. We conclude the proof in the following claim.

CLAIM 21. *All agents report the true state with probability 1 and there is no transfer on the equilibrium.*

PROOF. By Claim 16, all agents present the tightest evidence. Hence, it follows from Claim 15 that other-refutable lies are not reported, and from Claim 17 that nonrefutable lies are not presented, and Claim 20 that self-refutable lies are not presented. Therefore, all agents report the true state with probability 1. Since the tightest evidence also supports their true state claim, there is no transfer in equilibrium. □

A.4 *Proof of Theorem 4*

A.4.1 *Implementing mechanism* The message space of the mechanism is given by $M=\Pi_i M_i$, where $M_i=S\times\mathcal{E}_i$. That is, this is a direct mechanism. A typical message for agent i is represented as $m_i=(s_i,E_i)$.

An agent i challenges a state claim s when he presents a message (s_i,E_i^γ) if $\exists t$ such that $c_i(E_i^*(s),s)\le c_i(E_i^\gamma,s)-t$ but $c_i(E_i^*(s),s_i)>c_i(E_i^\gamma,s_i)-t$. When agent i challenges s at s' using $E_i^\gamma(s,s')$, the profit he makes (under constant preferences) is given by $t-c_i(E_i^\gamma(s,s'),s')+c_i(E_i^*(s),s')$. We pre-select a value $t_i(s,s')$ for any pair of states s and s' such that agent i can challenge s at s' so that $c_i(E_i^*(s),s)\le c_i(E_i^\gamma(s,s'),s)-t_i(s,s')$ but $c_i(E_i^*(s),s')>c_i(E_i^\gamma(s,s'),s')-t_i(s,s')$. Observe that $t_i(s,s')\in(-1,1)$ since $|c_i(E_i,s)-c_i(E_i,s')|<1$ for any i, E_i, s, and s'. In this case, we write $\Gamma_i(s,s_i)=1$ to denote that agent i has challenged s at the state s_i.

The outcome is determined as follows:

$$g(m)=\{\begin{cases} f(s_1), & \text{if } \Gamma_i(s_1,s_i)=1 \text{ for } i\neq 1;\\ f(s), & \text{if } \forall i\neq 1,\ s_i=s,\ \Gamma_i(s_1,s)=0, \text{ and } \Gamma_1(s,s_1)=1;\\ f(s_1), & \text{otherwise.}\end{cases}$$

The mechanism employs the following transfers. The first transfer penalizes agent 1 an amount of 1 dollar for every agent who makes a valid challenge to s_1. That is,

$$\tau_{1,i}^1(m)=-1 \quad \text{if } \Gamma_i(s_1,s_i)=1.$$

The second transfer, which applies only to agent 1, incentivizes him to agree (in the state dimension) with a report of $(s,E_i^*(s))_{i\neq 1}$ if such a report is made unless he issues a challenge. Formally,

$$\tau_1^2(m)=\begin{cases}-1, & \text{if } s_i=s\ \forall i\neq 1 \text{ and } m_1\neq\big(s,E_1^*(s)\big) \text{ and } \Gamma_1(s,s_1)=0;\\ 0, & \text{otherwise.}\end{cases}$$

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The third transfer applies only to agents other than 1. It incentivizes them to either (i) agree with the agent with the least index challenging agent 1 along the state dimension or (ii) agree with agent 1 if no agent is challenging agent 1.

$$\tau_i^3(m) = \begin{cases} -1, & \text{if } s_i \neq s_j \text{ where } j = \min\{k : \Gamma_k(s_1, s_k) = 1\}; \\ -1 & \text{if } \{k : \Gamma_k(s_1, s_k) = 1\} = \emptyset \text{ and } m_i \neq (s_1, E_i^*(s_1)) \\ 0, & \text{otherwise.} \end{cases}$$

The fourth transfer is related to the challenge payouts. For agent 1, we have

$$\tau_1^4(m) = \begin{cases} t_1(s, s_1), & \text{if } \forall i \neq 1,\ s_i = s, \Gamma_i(s_1, s) = 0, \Gamma_1(s, s_1) = 1; \\ 0, & \text{otherwise.} \end{cases}$$

For agent $i \neq 1$, we have

$$\tau_i^4(m) = \begin{cases} t_i(s_1, s_i), & \text{if } \Gamma_i(s_1, s_i) = 1; \\ 0, & \text{otherwise.} \end{cases}$$

Intuitively, the mechanism works as follows. First, we prevent agent 1 from presenting any lies that other agents can challenge by encouraging others to challenge when possible (τ^4 provides this incentive) and penalizing him for each challenge against his claim (τ^1 yields this penalty). If the other agents are telling the truth, τ^2 provides agent 1 the incentive to agree with them. If agent 1 is challenged, τ^3 ensures that every challenge is mounted with the same state claim. Then agent 1 can deviate to match this common state avoiding all the penalties from τ^1. This forms a profitable deviation. This leaves us with the truth, and lies that only agent 1 can challenge. If agent 1 tells a lie that only he can challenge, we get the other agents to agree with him using τ^3 (this helps the designer figure out what state is being challenged) and allow agent 1 to challenge this agreement. Note that other agents are not penalized when agent 1 challenges.

A.4.2 *Proof of implementation* The following lemma, which is a property of the evidence structure, finds use later.

LEMMA 2. *If s' is a lie that only agent i can challenge at s^*, then agent j ($\neq i$) cannot challenge s^* at s' with evidence $E_j^*(s')$.*

PROOF. s' is a lie that only i can challenge at s^*. Then j cannot challenge s' at s^*. Therefore, $\forall E \in \mathcal{E}_j$, $c_j(E, s^*) - c_j(E_j^*(s'), s^*) \geq c_j(E, s') - c_j(E_j^*(s'), s')$. With $E = E_j^*(s^*)$, this yields $c_j(E_j^*(s^*), s^*) - c_j(E_j^*(s'), s^*) \geq c_j(E_j^*(s^*), s') - c_j(E_j^*(s'), s')$. If j can challenge s^* at s' with challenge evidence $E_j^*(s')$, then we must have $c_j(E_j^*(s'), s') - c_j(E_j^*(s^*), s') < c_j(E_j^*(s'), s^*) - c_j(E_j^*(s^*), s^*)$, which is a contradiction. □

Essentially, Lemma 2 allows the designer to deduce that in a profile of the form $((s_1, E_1), (s, E_2^*(s)), (s, E_3^*(s)), \ldots, (s, E_I^*(s)))$, if agent 1 cannot challenge s at s_1, then it is actually agent 1 challenging s rather than agents $i \neq 1$ challenging s_1. In a canonical mechanism, the need for three agents is often to identify the agent who is deviating

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

from the majority. Lemma 2 allows us to dispense with this requirement in the costly evidence setting (under constant preferences).

Suppose the true state is s^* and consider any pure strategy equilibrium $((s_i, E_i))_{i\in\mathcal{I}}$.

CLAIM 22. *If there is an agent $i \neq 1$ who can challenge s_1, then s_1 is challenged.*

PROOF. Suppose not. Then no agent challenges s_1. Consider a deviation for i to (s_i, E_i^γ), which challenges s_1. The outcome remains $f(s_1)$ and agent i does not incur any penalties from τ^3 as he is the first agent who challenges s_1. He gains a reward from τ^4, so that this is a profitable deviation. □

CLAIM 23. *If there is an agent $i \neq 1$ who can challenge s_1, then $\exists s \in S$ such that $s_i = s$, $\forall i \neq 1$.*

PROOF. If there are only two agents, then this claim is trivially satisfied. So, suppose there are three or more agents. From Claim 22, s_1 is challenged by some agent. Suppose i is the first agent to challenge s_1. For any agent $j \neq i$, $j \neq 1$, if $s_j \neq s_i$, then he gets a penalty of 1 dollar from τ^3. Consider a deviation to match s_i. The outcome remains $f(s_1)$, but the agent avoids the penalty from τ^3. Since $t_i(s, s') \in (-1, 1)$, any possible reward from challenging using s_j (from τ^4) is less than 1 dollar. Hence, this is a profitable deviation. □

CLAIM 24. *Agent 1 does not present lies, which others can challenge.*

PROOF. From Claim 22, if agent 1 presents such a lie, then some other agent challenges it. Further, from Claim 23, $\exists s \in S$ such that $s_i = s$, $\forall i \neq 1$. This yields agent 1 a penalty of at least 1 dollar from τ^1, which he can avoid by deviating to present $(s, E_1^*(s))$. This may change the outcome, but the utility loss from that is less than 1 dollar, so that this forms a profitable deviation. □

CLAIM 25. *If agent 1 presents a lie that only he can challenge at s^*, then every other agent agrees with him in the state and evidence dimensions.*

PROOF. We note that if s_1 is a lie that agent $i \neq 1$ cannot challenge at s^*, then it is among his best responses to submit $E_i^*(s_1)$ and obtain the outcome $f(s_1)$. If he does not present $(s_1, E_i^*(s_1))$, then agent i is either mounting a challenge, which yields him a loss (as otherwise s_1 would be a lie agent i could challenge), or disagreeing without challenging and incurring a loss of 1 dollar. In either case, it is best for agent i to deviate to match agent 1's claim in both dimensions. □

CLAIM 26. *Agent 1 does not present a lie that only he can challenge at s^*.*

PROOF. Suppose not. From Claim 25, if s_1 is a lie that only agent 1 can challenge at s^*, then every agents $i \neq 1$ presents $(s_1, E_i^*(s_1))$ and the outcome is $f(s_1)$. Disagreeing without challenging is suboptimal owing to τ^2. Therefore, consider a deviation for agent 1

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

to $(s^*, E_1^\gamma(s_1, s^*))$. From Lemma 2, when agents $i \neq 1$ present $(s_1, E_i^*(s_1))$, $\Gamma_i(s^*, s_1) = 0$. However, $\Gamma_1(s_1, s^*) = 1$. This yields him a profit from challenging since the outcome remains $f(s_1)$. □

Claim 27. *The mechanism implements.*

Proof. Owing to Claims 24 and 26, agent 1 can only present claims, which no one can challenge. In such case, it is optimal owing to τ_1^2 and τ^3 for all agents to also present the designated cheapest evidence along with such claims. This leads to no transfers, the correct f-optimal outcome, and the submission of only the cheapest evidence in equilibrium. This satisfies the definition of implementation. □

A.4.3 *Measurability does not imply evidence monotonicity* Consider the following evidence structure:

State/Agent	1	2
s_1	$\{\{s_1, s_2, s_3, s_4\}\}$	$\{\{s_1, s_2, s_3, s_4\}\}$
s_2	$\{\{s_1, s_2, s_3, s_4\}, \{s_2, s_4\}\}$	$\{\{s_1, s_2, s_3, s_4\}\}$
s_3	$\{\{s_1, s_2, s_3, s_4\}, \{s_3, s_4\}\}$	$\{\{s_1, s_2, s_3, s_4\}\}$
s_4	$\{\{s_1, s_2, s_3, s_4\}, \{s_2, s_4\}, \{s_3, s_4\}, \{s_4\}\}$	$\{\{s_1, s_2, s_3, s_4\}\}$

The cost of the article $\{s_4\}$ is positive but finite.

Clearly, any f is measurable with respect to the evidence structure since between any pair of states, agent 1 has a different endowment. To see that f is not evidence monotonic, consider the states s_1 and s_4. Note that of necessity, $E_1^*(s_1) = \{s_1, s_2, s_3, s_4\}$. We now deal with three cases. First, if $E_1^*(s_4) = \{s_1, s_2, s_3, s_4\}$, then no agent can challenge s_4 at s_1 since relative to $E_1^*(s_4)$, the cost of all articles of evidence are strictly higher at s_1 than at s_4 for agent 1, and agent 2 has no cost variation. Recall that for f to be evidence monotonic under constant preference, it is necessary that the cost of some article of evidence reduces relative to $E_1^*(s_4)$ in going from s_4 to s_1. Instead the articles $\{s_2, s_4\}$ and $\{s_3, s_4\}$ rise in cost from being costless to becoming unavailable (costing ∞).

Second, if $E_1^*(s_4) = \{s_2, s_4\}$, then no agent can challenge s_4 at s_2 because as in the above case, the articles $\{s_1, s_2, s_3, s_4\}$ & $\{s_2, s_4\}$ continue to be costless, while the article $\{s_3, s_4\}$ has now become unavailable. Third, if $E_1^*(s_4) = \{s_3, s_4\}$, then no agent can challenge s_4 at s_3 since the articles $\{s_1, s_2, s_3, s_4\}$ and $\{s_3, s_4\}$ continue to be costless, while the article $\{s_2, s_4\}$ has now become unavailable. Further, $E_1^*(s_4) \neq \{s_4\}$ since $\{s_4\}$ is not among the least costly evidence. Therefore, f is not evidence monotonic under constant preference.

A.5 *Renegotiation-proof contracting*

A.5.1 *Proof of Theorem 5* We begin by proving that it is necessary for the existence of renegotiation-proof contracts that either Conditions *(a)* or *(b)* be satisfied. For a contradiction, consider a social choice function f and suppose that both Condition *(a)* and Condition *(b)* are violated for f. Then there exist a pair of states s and s' such that $f(s) \neq f(s')$ and one of the following is true:

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(c) One of the agents can refute s' at s and the other agent can refute s at s'; or

(d) s' is nonrefutable at s, but only one agent can refute s at s'.

Further, suppose that there is a mechanism $\mathcal{M} = (M, g, (\tau_i)_{i\in\mathcal{I}})$, which implements for every pair of v and h in Nash equilibrium with renegotiation. We note that any such mechanism must be such that $\sum_i \tau_i = 0$. Indeed, if $\mathcal{M}$ results in a budget surplus ($\sum_i \tau_i < 0$), then such an allocation is not efficient with respect to v and will get renegotiated under h. Consider any equilibrium σ of this mechanism and further consider any pair of messages (m_i^s, m_j^s) on the support of σ at s and another pair of messages $(m_i^{s'}, m_j^{s'})$ on the support of σ at s'. Further, assume $v_i(f(s')) = 1$, $v_i(f(s)) = 0$, $v_j(f(s')) = 0$, $v_j(f(s)) = 1$, and for all other members a of the set A, $v_i(a) + v_j(a) = 1$, $v_i(a) < 1$, $v_j(a) < 1$. Here, v is state independent. Further, choose any h such that v and h satisfy the constraints noted earlier.[26] To be concise, we will use $\bar{u}_i(m)$ to denote the utility agent i derives from the outcome and transfers chosen by $\mathcal{M}$ under the message profile m.

If Condition *(c)* is satisfied, then without loss of generality, the only possible evidence structure is that agent i can refute s' at s and agent $j(\neq i)$ can refute s at s'. Since $\mathcal{M}$ has $\sum_i \tau_i = 0$, $\bar{u}_i(m_i^{s'}, m_j^s) + \bar{u}_i(m_i^{s'}, m_j^s) = 1$. We consider the following two cases.

Case 1: If the mechanism $\mathcal{M}$ is such that $\bar{u}_i(m_i^{s'}, m_j^s) > \bar{u}_i(m_i^s, m_j^s) = v_i(f(s))$, then agent i deviates to $m_i^{s'}$ in state s and this yields an outcome different from $f(s)$, so that it is not possible that $\mathcal{M}$ implements f in Nash equilibrium with renegotiation. Note that this is feasible for agent i.

Case 2: If $\mathcal{M}$ is such that $\bar{u}_i(m_i^{s'}, m_j^s) \leq v_i(f(s))$, then $\bar{u}_j(m_i^{s'}, m_j^s) \geq v_j(f(s)) > v_j(f(s'))$ and agent j deviates to m_j^s in state s'.

In either case, $\mathcal{M}$ cannot implement f in Nash equilibrium with renegotiation.

If Condition (d) is satisfied, suppose that s' is nonrefutable at s and without loss of generality, that only agent j can refute s at s'. Since $\mathcal{M}$ has $\sum_i \tau_i = 0$, $\bar{u}_i(m_i^{s'}, m_j^s) + \bar{u}_i(m_i^{s'}, m_j^s) = 1$. We consider the following two cases.

Case 1: If the mechanism $\mathcal{M}$ is such that $\bar{u}_j(m_i^{s'}, m_j^s) > v_j(f(s'))$, then agent j deviates to m_j^s in state s' and this yields an outcome different from $f(s')$.

Case 2: If $\mathcal{M}$ is such that $\bar{u}_j(m_i^{s'}, m_j^s) \leq v_j(f(s'))$, then $\bar{u}_i(m_i^{s'}, m_j^s) \geq v_i(f(s')) > v_i(f(s))$ and agent i deviates to $m_i^{s'}$ in state s.

In either case, $\mathcal{M}$ cannot implement f in Nash equilibrium with renegotiation.

To prove that either of Conditions *(a)* and *(b)* are sufficient, we present the following mechanism. The message space is the same as our main mechanism: $M_i = S \times \mathcal{E}_i$. The outcome is given by $f(s_1)$, where f is the SCF we desire to implement. There are two transfers: A fine of 1 dollar for agent i if he is unable to support his state claim and a fine of 2 dollars if s_i is refuted by E_j. Each fine is paid to the other agent, so that the mechanism is budget balanced. Also, recall that the maximum bound of utility is less than 1 dollar, so that these transfers dominate the utility from the outcome.

[26] Since $\mathcal{M}$ must implement for every acceptable combination of v, h, and f, we take the liberty of choosing v. Moreover, our argument works for any h, which satisfies the constraints noted above; e.g., h may be set as the identity mapping, which corresponds to "no-renegotiation.".

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Assume that the true state is s^*. We now prove that this mechanism implements f with renegotiation.

First, no agent presents an other-refutable lie with positive probability; otherwise, since $\mathcal{E}$ is normal, the other agent presents the refuting evidence with probability 1 and this loses the first agent 2 dollars of money, which is more than the value of influencing the outcome. Second, if an agent presents a self-refutable lie s with positive probability, then Condition *(a)* yields that this agent can also refute s^* at s. Therefore, the agent cannot support his state claim, and is fined 1 dollar. As this is more than the value of the outcome, deviating to $(s^*, E_i^*(s^*))$ is profitable. Third, if an agent presents a nonrefutable lie s, then Condition *(b)* yields that i can refute s^* at s. Thus, he cannot support his state claim, and again, deviating to $(s^*, E_i^*(s^*))$ is profitable. Therefore, the only equilibrium is truth-telling, which results in the outcome $f(s)$ without transfers, which is efficient.

We note here that all the off-equilibrium outcomes above involved $f(s_1)$ with a budget balanced transfer of value greater than 1 dollar, so that the resulting outcomes are Pareto efficient. Therefore, the mechanism is invulnerable to renegotiation.

A.5.2 *Evidence structure which does not allow for renegotiation proof contracting*

Agent/State	ϕ	θ
Buyer	$\{\{\theta, \phi\}\}$	$\{\{\theta\}, \{\theta, \phi\}\}$
Seller	$\{\{\theta, \phi\}\}$	$\{\{\theta, \phi\}\}$

This evidence structure does not allow for renegotiation proof contracting since at state ϕ, state θ is a nonrefutable lie, but at state θ, only the buyer can refute the state ϕ. In contrast, if in the state with low investment, ϕ, the buyer could refute θ, then renegotiation-proof contracting would once again be possible.

References

Abreu, Dilip and Hitoshi Matsushima (1994), "Exact implementation." *Journal of Economic Theory*, 64, 1–19. [798, 799]

Aghion, Philippe et al. (2012), "Subgame-perfect implementation under information perturbations." *The Quarterly Journal of Economics*, 127, 1843–1881. [785]

Banerjee, Soumen and Yi-Chun Chen (2022), "Implementation with uncertain evidence." Unpublished Paper. [809]

Banerjee, Soumen, Yi-Chun Chen, and Yifei Sun (2023), "Direct implementation with evidence." arXiv preprint arXiv:2105.12298. [786, 798, 803]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2019), "Mechanisms with evidence: Commitment and robustness." *Econometrica*, 87, 529–566. [797]

Ben-Porath, Elchanan and Barton L. Lipman (2012), "Implementation with partial provability." *Journal of Economic Theory*, 147, 1689–1724. [783, 784, 785, 787, 800, 807]

Bull, Jesse and Joel Watson (2007), "Hard evidence and mechanism design." *Games and Economic Behavior*, 58, 75–93. [785, 789, 800, 807]

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Chen, Yi-Chun, Takashi Kunimoto, Yifei Sun, and Siyang Xiong (2021), "Rationalizable implementation in finite mechanisms." *Games and Economic Behavior*, 129, 181–197. https://www.sciencedirect.com/science/article/pii/S0899825621000762. [802]

Chen, Yi-Chun, Takashi Kunimoto, Yifei Sun, and Siyang Xiong (2022), "Maskin meets Abreu and Matsushima." *Theoretical Economics*, 17, 1683–1717. [808]

Chung, Kim-Sau and Jeffrey C. Ely (2003), "Implementation with near-complete information." *Econometrica*, 71, 857–871. [785]

Deneckere, Raymond and Sergei Severinov (2007), "Optimal screening with costly misrepresentation." Unpublished Paper. [807]

Deneckere, Raymond and Sergei Severinov (2008), "Mechanism design with partial state verifiability." *Games and Economic Behavior*, 64, 487–513. [807]

Dutta, Bhaskar and Arunava Sen (2012), "Nash implementation with partially honest individuals." *Games and Economic Behavior*, 74, 154–169. [808]

Green, Jerry R. and Jean-Jacques Laffont (1986), "Partially verifiable information and mechanism design." *The Review of Economic Studies*, 53, 447–456. [807]

Jackson, Matthew O. (1992), "Implementation in undominated strategies: A look at bounded mechanisms." *The Review of Economic Studies*, 59, 757–775. [785]

Kartik, Navin and Olivier Tercieux (2012a), "Implementation with evidence." *Theoretical Economics*, 7, 323–355. [783, 784, 785, 789, 798, 800, 801, 803, 804, 807, 808]

Kartik, Navin and Olivier Tercieux (2012b), "A note on mixed-Nash implementation." Unpublished Paper. [807]

Kartik, Navin, Olivier Tercieux, and Richard Holden (2014), "Simple mechanisms and preferences for honesty." *Games and Economic Behavior*, 83, 284–290. [808]

Maskin, Eric (1999), "Nash equilibrium and welfare optimality." *The Review of Economic Studies*, 66, 23–38. [784]

Maskin, Eric (2002), "On indescribable contingencies and incomplete contracts." *European Economic Review*, 46, 725–733. [804]

Maskin, Eric and John Moore (1999), "Implementation and renegotiation." *The Review of Economic Studies*, 39–56. [786, 804, 808]

Maskin, Eric and Jean Tirole (1999), "Unforeseen contingencies and incomplete contracts." *The Review of Economic Studies*, 66, 83–114. [786, 804]

Moore, John and Rafael Repullo (1988), "Subgame perfect implementation." *Econometrica*, 1191–1220. [784]

Co-editor Marina Halac handled this manuscript.

Manuscript received 11 September, 2021; final version accepted 13 May, 2023; available online 16 May, 2023.

15557561, 2024, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/TE5015 by ZBW Kiel - Hamburg (German National Library of Economics), Wiley Online Library on [04/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License