ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Strausz, Roland

Working Paper

Mechanism Design with Partially Verifiable Information

Discussion Paper, No. 45

Provided in Cooperation with:
University of Munich (LMU) and Humboldt University Berlin, Collaborative Research Center Transregio 190: Rationality and Competition

Suggested Citation: Strausz, Roland (2017) : Mechanism Design with Partially Verifiable Information, Discussion Paper, No. 45, Ludwig-Maximilians-Universität München und Humboldt-Universität zu Berlin, Collaborative Research Center Transregio 190 - Rationality and Competition, München und Berlin

This Version is available at:
https://hdl.handle.net/10419/185715

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.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Mechanism Design with Partially Verifiable Information

Roland Strausz (Humboldt University Berlin)

Discussion Paper No. 45

August 3, 2017

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

Mechanism Design with Partially Verifiable Information

Roland Strausz*

August 3, 2017

Abstract

In mechanism design with (partially) verifiable information, the revelation principle holds if allocations are modelled as the Cartesian product of outcomes and verifiable information, giving rise to evidence-contingent mechanisms. Consequently, incentive constraints characterize the implementable set. The revelation principle does not hold when an allocation is modelled as only an outcome so that mechanisms are non-contingent. Yet, any outcome implementable by an evidence-contingent mechanism is implementable by a non-contingent mechanism, provided it can both extend and restrict reporting information. A type-independent bad outcome implies the latter property.

JEL Classification Numbers: D82
Keywords: Revelation principle, Mechanism Design, Verifiable Information

*Contact details: Humboldt-Universität zu Berlin, strauszr@hu-berlin.de. This paper was started during my visit at the Cowles Foundation at Yale University in the spring of 2016. I thank Dirk Bergemann, Jesse Bull, Eduardo Faingold, Francoise Forges, Tibor Heumann, Johannes Hörner, Navin Kartik, Frederic Koessler, Daniel Krähmer, Barton Lipman, Mallesh Pai, Vasiliki Skreta, Juuso Toikka, Juuso Välimäki, Joel Watson, and Alex Wolitzky for extremely helpful discussions and comments on earlier drafts. Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 is gratefully acknowledged.

1 Introduction

Focusing exclusively on the role of asymmetric information, mechanism design studies the extent to which the distribution of information restricts economic allocations. Ideally, the theory places no limitations on the ability of economic agents to interact and communicate, in principle allowing any type of game or mechanism to govern their communication and interactions.

The revelation principle plays a crucial role in enabling mechanism design to achieve its goal of analyzing unrestricted mechanisms. The principle is well established under non-verifiability, where economic agents can only send non-credible messages about their private information. For environments in which agents have (partially) verifiable information, the applicability of the revelation principle seems less well understood. Following observations in Green and Laffont (1986) and subsequent work, the principle that *any* implementable allocation is implementable by an incentive compatible direct mechanism holds only under specific conditions on the underlying verifiability structure. From a conceptual perspective, this is puzzling and suggests that mechanism design with verifiable information fundamentally differs from mechanism design without verifiability.

To the contrary, I argue that with an appropriate (extended) notion of an economic allocation, the classical revelation principle fully extends to settings with verifiable information. In particular, the principle obtains if the set of economic allocations is modelled as the Cartesian product of the set of outcomes and the set of verifiable information, and, following Harsanyi (1967), the agents' payoff functions over these economic allocations are modelled to reflect the verifiable information structure.[1] Defining direct mechanisms as mappings from reports about an agent's type to the set of (extended) economic allocations yields the revelation principle in its usual sense: any implementable allocation is implementable by an incentive compatible direct mechanism. This conceptual insight then also has the practical implication that the usual tools of mechanism design—direct mechanisms and incentive constraints—allow a full characterization of the set of implementable outcomes.

Because these direct mechanisms effectively condition the pay-off relevant outcome on the presentation of verifiable evidence, they can be intuitively interpreted as

[1] Contrary to other modeling approaches, this approach yields a Bayesian Game in the sense of Harsanyi (1967) (see footnote 11 for more details).

evidence-contingent. In order to investigate to what extent these evidence-contingent contract are essential for implementability, I examine the smaller class of *non-contingent mechanisms*, which are mechanisms that select only an outcome. I show that such mechanisms are able to implement all outcomes that are implementable with evidence-contingent mechanisms if two elementary operations in the design of mechanisms are available: 1) broadening communication by adding (non-credible) messages; and 2) restricting communication to a subset of available messages.[2]

The first operation is clearly a sine qua non for the construction of direct mechanisms, whereas the second operation is implicitly available in any mechanism design problem without verifiability. The reason is that a mechanism can implicitly restrict communication to exclude "unwanted" messages by assigning to them an allocation that is already available for some equilibrium (ie. wanted) message. While this assignment of unwanted messages does not enlarge the set of possible deviations in the non-verifiability framework, it may do so when information is verifiable and thereby destroy incentive compatibility. Hence, when information is verifiable, the set of implementable outcomes via evidence-contingent mechanisms is generally strictly larger than the set of implementable outcomes via non-contingent mechanisms.

As a consequence, the answer to the paper's motivating question whether there are any conceptual differences between mechanism design with and without verifiability is affirmative but subtle: With verifiable information, restricting communication is harder to achieve than without verifiability. While subtle, this difference has nevertheless practical implications for applications of mechanism design with additional constraints such as frameworks in which the disclosure of evidence is the agent's inalienable action (eg. Bull and Watson, 2007).

If however the economic environment that underlies the mechanism design problem exhibits a "bad outcome"—an outcome that, independent of his type, can serve as a unequivocal punishment on the agent—then this difference is inconsequential in terms of implementability via non-contingent mechanisms. In this case, any outcome implementable via an evidence-contingent is also implementable via a (possibly non-

[2]In line with the observations in the literature, these non-contingent mechanisms may however no longer be direct or incentive compatible. Hence, the reported failure of the revelation principle in settings with verifiability can therefore also be understood as a failure with respect to non-contingent mechanisms, whereas this paper shows that the principle holds with respect to evidence-contingent mechanisms.

direct) non-contingent mechanism. This is so, because the bad outcome provides a different, more straightforward channel by which a mechanism can implicitly restrict communication: by assigning "unwanted" messages to the type-independent bad outcome, the agent is dissuaded to send such messages in the first place.[3] Hence, with the availability of a bad outcome, the set of implementable outcomes by evidence-contingent mechanisms coincides with the set of implementable outcomes by are non-contingent mechanisms. In many applications of mechanism design, such as in settings with transfers or in evidence games (eg. Hart, Kremer, and Perry, 2017), such bad outcomes are naturally available.

2 Related literature

Stated in somewhat technical terms but boiled down to its essence, modeling an allocation as the Cartesian product of outcomes and verifiable information leads to the consideration of mechanisms that are mappings for which the evidence structure is part of their *range* rather than their *domain*. Although the validity of the revelation principal has been extensively addressed in the literature, the relevance of the mechanism's domain and range has not been noticed before. Nevertheless many of the themes in this paper have in some way or another also been raised in the literature so that a careful discussion is crucial to understand this paper's contribution.

Green and Laffont (1986) were the first to note a failure of the usual revelation principle in mechanism design problems with (partially) verifiable private information. Mechanisms in their setup are mappings whose domain directly reflects the agent's verifiable information and are, in the terms of the present paper, therefore non-contingent. The authors obtain a revelation principle for their class of mechanisms only under a so-called *nested range condition*, where the agent's verifiability exhibits a nested structure. They show by explicit examples that without this condition, the revelation principle fails. They note that this failure limits the applicability of mechanism design to study general settings with partially verifiable information, because one cannot characterize the set of implementable allocations.

[3] If the agent's bad outcome is type-dependent then this straightforward channel is no longer available, since the assignment of the unwanted message to the type-dependent bad outcome would then necessarily also need to depend on the agent's type about which the agent is however privately informed.

While the nested range condition arises naturally in many practical frameworks of verifiability, Singh and Wittman (2001) give natural examples of concrete economic environments for which it is violated. For principal-agent models that satisfy a *unanimity* condition on the agent's preferences, they derive necessary and sufficient conditions for the implementability of a social choice function regardless of the underlying verifiability structure. The authors do not discuss possible extensions of direct mechanisms such as broadening and restricting communication. Following Green and Laffont (1986), they consider non-contingent mechanisms; mechanisms are mappings whose domain coincides with the agent's verifiable information.

Also Bull and Watson (2007) address the validity of the revelation principle in mechanism design problems with partially verifiable information. An important conceptual difference is however that the authors focus on economic settings in which the presentation of verifiable information is the agent's inalienable action, leading to the additional problem of moral hazard. This moral hazard problem implies that mechanisms cannot be evidence-contingent, but are effectively non-contingent. Inalienability, moreover, implies that the operation of restricting communication is not allowed in the design of mechanisms. In line with the results in this paper, the authors show that the revelation principle in their framework does not hold generally but only under an *evidentiary normality* condition, which is related to the nested range condition of Green and Laffont (1986).[4]

In the presence of verifiable information, also Deneckere and Severinov (2008) study natural limitations on mechanisms and, in particular, limits on the amount of information which the agent can send. Similar to Green and Laffont (1986) and Bull and Watson (2007), the authors do not model the presentation of evidence as part of the economic allocation so that they also exclude the revelation of evidence from the mechanism's range. In contrast to Green and Laffont (1986) but in line with Bull and Watson (2007), they allow agents to send cheap talk messages about their types. In part of their study on the limits of communication, they further explicitly assume the existence of a type-independent bad outcome. Since the principal can use this outcome to dissuade agents from presenting certain pieces of evidence, the mechanisms which the authors study can, in the terms of the current paper, both extend and restrict communication.

[4]In their study of sequential message-sending games, Lipman and Seppi (1995) already refer to this condition as *the full reports condition*.

While Deneckere and Severinov (2008) explicitly assume the existence of a type-independent bad outcome, such an outcome is implicitly also available if the agent's utility is independent of his private information. As this is the defining feature of evidence games (eg. Glazer and Rubinstein, 2004 and 2006, Sher, 2014, Hart, Kremer, and Perry, 2017), evidence games represent frameworks in which mechanisms are able to restrict communication. In line with this paper, it therefore does not matter whether mechanisms are modelled as mappings which have the presentation of evidence as part of their range (eg. Glazer and Rubinstein, 2004 and 2006) or their domain (eg. Hart, Kremer, and Perry, 2017). Similarly, in a quasi-linear context with transfers, (eg. Bull, 2008), type-independent bad outcomes are also implicitly available because mechanisms can specify a large negative transfers when supplying some types of evidence.

Given the failures in establishing the classical revelation principle in a context with verifiable information, the literature has instead characterized classes of (indirect) mechanisms that are sufficient to achieve any implementable outcome. These studies emphasize the power of dynamic mechanisms. For instance, Bull and Watson (2007) identify three-stage dynamic mechanisms—in which agents first send cheap talk messages to the mechanism designer, who then sends messages to the agents, who, in the final third step, disclose their verifiable information—as such a sufficient class.[5] Because these dynamic mechanisms ask for the presentation of verifiable evidence in a final stage, they exhibit a strong similarity to the static type-contingent mechanisms that I study here. With respect to these studies, the insight is therefore that, with an appropriate definition of an allocation, dynamic considerations are not needed.

The literature on (unique) implementation with perfect information has also studied verifiable evidence (eg. Bull and Watson, 2004, Ben-Porath and Lipman, 2012 and Kartik and Tercieux, 2012). From the perspective of this literature, the idea of extending the outcome space as presented in this paper, is not new. In particular, Section 4 in Kartik and Tercieux (2012) considers the same kind of extended allocation space and also addresses the question whether restricting to non-contingent mechanisms reduces the set of implementable outcomes. The authors do however not discuss its implications on the main focus of this paper—the validity of the revela-

[5]Dynamic mechanisms also play a crucial role in Lipman and Seppi (1995), Glazer and Rubinstein (2004), Bull (2008), and Deneckere and Serverinov (2008).

tion principle, because this principle is not a helpful concept when demanding unique implementation.[6]

Analyzing the role of verifiable information in a game theoretical rather than a mechanism design context, Forges and Koessler (2005) study communication between players with private but partially verifiable information. Since the authors do not follow a mechanism design perspective, they do not use the notion of mechanisms as implementing some social choice function. Instead, they study the set of all feasible equilibrium outcomes given that partially verifiable information limits the agents' communication possibilities. Yet, the different versions of the revelation principle they obtain and their underlying proofs are closely linked to the one shown in this paper. Importantly, the authors also explicitly point out the importance of broadening and restricting communication for expanding the set of equilibrium outcomes in their game theoretical framework.

Finally, verifiable information arises endogenously in contexts where players can certify their private information through a certifier. Consequently, the results of this paper has also implications for this more applied literature (eg. Hagenbach, Koessler, and Perez-Richet 2014, Koessler and Skreta, 2016 or Yamashita, 2017).

3 The Green and Laffont (1986) example

This section first reiterates the example by which Green and Laffont (1986) demonstrate the failure of the revelation principle. It next shows how an extended notion of an economic allocation repairs the failure, leading to the class of evidence-contingent mechanisms that are direct and incentive compatible.

Example 1: Green and Laffont (1986)

Consider a principal and one agent, who can be of three types $\Theta_1 = \{\theta_1, \theta_2, \theta_3\}$. The set of outcomes is $X_1 = \{x_1, x_2\}$. The agent has partially verifiable information, which Green and Laffont concisely model by type-specific message sets $M(\theta_i)$ with the interpretation that type θ_i can only send messages from the set $M(\theta_i)$. In their specific example they consider the sets $M_1(\theta_1) = \{\theta_1, \theta_2\}$, $M_1(\theta_2) = \{\theta_2, \theta_3\}$, $M_1(\theta_3) = \{\theta_3\}$. The agent's utilities $u_1(x, \theta)$ are as follows:

[6] In a private communication, the authors sent notes in which they, in a mechanism design context with quasi-linearity and transfers, study counterparts of my Propositions 3 and 4.

$u_1(x,\theta)$	θ_1	θ_2	θ_3
x_1	10	5	10
x_2	15	10	15

Clearly, the direct mechanism $g_1 : \Theta \to X$ with $g_1(\theta_1) = g_1(\theta_2) = x_1$ and $g_1(\theta_3) = x_2$ induces a game that implements the social choice function $f_1(\theta_1) = x_1$, $f_1(\theta_2) = f_1(\theta_3) = x_2$. This is so, because type θ_1, who cannot send the message θ_3, optimally sends the message θ_1, which results in $x_1 = f_1(\theta_1)$. Type θ_2, who cannot send the message θ_1, optimally sends the message θ_3, which results in $x_2 = f_1(\theta_2)$. Type θ_3, who can only send the message θ_3, optimally sends the message θ_3, which results in $x_2 = f_1(\theta_3)$.

Note that while direct, the mechanism is not truthful, because it induces type θ_2 to misreport his type as θ_3. A truthful direct mechanism $\hat{g}_1$ that implements f_1, requires $\hat{g}_1(\theta_1) = x_1$, $\hat{g}_1(\theta_2) = x_2$, $\hat{g}_1(\theta_3) = x_2$. This mechanism is however not incentive compatible, because it induces type θ_1 to report θ_2. This established the failure of the revelation principle.

We can however implement the social choice function f_1 with a truthful direct mechanism if one extends the concept of an allocation as follows. In addition to an outcome $x \in X$, an allocation also describes a verifiable message $\theta \in \Theta$ which the agent is to send. Hence, let the set $Y = X \times \Theta$ represents this extended set of allocations with a typical element $y = (x, \theta) \in Y$. Define utilities as follows:[7]

$$\hat{u}(x, \theta|\theta') = \begin{cases} u(x, \theta') & \text{if } \theta \in M(\theta') \\ -\infty & \text{otherwise.} \end{cases}$$

In this extended context, a direct mechanism is a function $\tilde{y} = (\tilde{x}, \tilde{\theta}) : \Theta \to Y$ from the set of non-verifiable claims about Θ to the extended set of allocations of outcomes X and verifiable messages about Θ.[8] Its interpretation is that if the agent sends the

[7]Footnote 11 discusses in more detail the role and appropriateness of $-\infty$.

[8]Hence, claims and messages are different objects in this context and not synonyms. While I use $\hat{u}$ and the mechanism $\tilde{y}$ only as hypothetical constructs for deriving a revelation principle, they allow the following literal interpretation. Although an agent can costlessly make any cheap-talk claim about his type, he has a prohibitively high cost to back up his claim if he cannot present the verifiable information to substantiate it. For instance, a person with only \$10 dollars in his pocket, can claim he has \$20, but has a prohibitively high cost of actually retrieving \$20 from his pocket.

non-verifiable claim θ_i, the mechanism picks $\tilde{x}(\theta_i) \in X$ and the agent must present the message $\tilde{\theta}(\theta_i) \in \Theta$. The direct mechanism $y(\theta_1) = (x_1, \theta_1)$, $y(\theta_2) = (x_2, \theta_3)$, $y(\theta_3) = (x_2, \theta_3)$ is incentive compatible (truthful) and implements the allocations in X as intended by the social choice function f_1.

4 The Mechanism Design Setup

The above example suggests that by extending the concept of an implementable allocation, one can recover the revelation principle. This section argues that this insight is general. In order to understand how these extended allocations change the analysis, it is most instructive to derive this insight in the original framework of Green and Laffont (1986).

Following Green and Laffont (1986), I therefore consider a principal facing an agent with utility function $u(x, \theta)$, which depends on a characteristic $\theta \in \Theta$ and an outcome $x \in X$. For concreteness, we assume that both sets are finite: $\Theta = \{\theta_1, \ldots, \theta_K\}$ and $X = \{x_1, \ldots, x_L\}$ with $K, L \in \mathbb{N}$.[9] The agent knows θ, whereas the principal only knows that $\theta \in \Theta$. The agent has verifiable information represented by a correspondence $M : \Theta \to \Theta$ with the interpretation that type θ_i can only send messages about θ from the set $M(\theta_i)$. Hence, a type θ describes both the agent's preferences over X and an available message set. In short, one can represent the principal-agent problem with verifiable information by a structure $\Gamma = \{X, \Theta, M(\cdot), u(\cdot, \cdot)\}$, which describes all the primitives.

Fully in line with the usual goal of mechanism design, Green and Laffont (p.448) state their intention to "study the class of social choice functions f from Θ into X that can be achieved despite the asymmetry of information between the two players." For this, they define a direct mechanism as follows.

Definition 1: A mechanism $(M(\cdot), g)$ consists of a correspondence $M : \Theta \to \Theta$ such that $\theta \in M(\theta)$ for all $\theta \in \Theta$, and an outcome function $g : \Theta \to X$.

Hence, the mechanism $(M(\cdot), g)$ presents the agent with a single-person decision problem in which an agent of type θ has to pick some θ but in which his choice is

In contrast, a person with \$20 dollars in his pocket, can claim to have \$20 dollars and also produce the \$20 at zero costs.

[9] All arguments naturally extend if Θ and X are subsets of some more general Euclidean spaces.

restricted to his message set $M(\theta)$. Following Green and Laffont, one can describe the agent's optimal decision behavior as follows. Given the correspondence $M(\cdot)$, the outcome function g induces a *response rule* $\phi_g : \Theta \to \Theta$ defined by[10]

$$\phi_g(\theta) \in \arg \max_{m \in M(\theta)} u(g(m), \theta).$$

This leads to the following two notions of implementability.

Definition 2: A social choice function $f : \Theta \to X$ is $M(\cdot)$*-implementable* iff there exists an outcome function $g : \Theta \to X$ such that:

$$g(\phi_g(\theta)) = f(\theta) \text{ for any } \theta \text{ in } \Theta,$$

where $\phi_g(\cdot)$ is an induced response rule.

Definition 3: A social choice function $f : \Theta \to X$ is *truthfully* $M(\cdot)$*-implementable* iff there exists an outcome function $g^* : \Theta \to X$ such that:

$$g^*(\phi_{g^*}(\theta)) = f(\theta) \text{ for any } \theta \text{ in } \Theta$$

and

$$\phi_{g^*}(\theta) = \theta.$$

The example in the previous section proves that there exists social choice functions that are $M(\cdot)$-implementable but not truthfully $M(\cdot)$-implementable. This result establishes a failure of the revelation principle and the ensuing problem that one cannot, in general, characterize the set of implementable social choice functions for all principal-agent problems Γ.[11]

The next two examples suggest, however, that not only the notion of *truthfully* $M(\cdot)$-implementability is problematic, but that the more primitive notion of $M(\cdot)$-implementability also raises questions. In Example 2, the specified social choice

[10] Because Θ is finite, the maximum exists.

[11] Note that the agent's decision problem involves a type-dependent action set. Hence, extending this approach to multiple agents leads to the concern that the game induced by the mechanism does not, strictly speaking, correspond to a Bayesian Game. In the definitions following Harsanyi (1967), games with imperfect information require that the agent's action sets are type-independent. (See footnote 14 for more details and also Bull and Watson (p. 80, 2007) who point out that their "disclosure game [...] is a Bayesian game with type-contingent restrictions on actions".)

function is not $M(\cdot)$-implementable, whereas it is implementable if the mechanism can, in addition to the messages in $M(.)$, also condition on two non-verifiable messages. In Example 3, the specified social choice function is not $M(\cdot)$-implementable, whereas it is implementable if the mechanism can limit the messages that can be sent.

Example 2: Too few messages

Consider a third outcome x_3 by duplicating outcome x_2 in the sense that each type θ is indifferent between x_3 and x_2. Hence, the set of outcomes is $X_2 = \{x_1, x_2, x_3\}$ with the utility

$u_2(x,\theta)$	θ_1	θ_2	θ_3
x_1	10	5	10
x_2	15	10	15
x_3	15	10	15

Suppose one wants to implement the social choice function $f_2(\theta_1) = x_1$, $f_2(\theta_2) = x_2$, $f_2(\theta_3) = x_3$. Then, based on the reasoning in Example 1, it is straightforward to see that this social choice function is not $M(\cdot)$-implementable, but it is implementable by a mechanism that, in addition to reporting θ, asks for some extra cheap talk message $\hat{m} \in \hat{M} = \{a, b\}$ as follows

$$g_2(\theta, \hat{m}) = \begin{cases} x_2 & \text{if } (\theta, \hat{m}) = (\theta_3, a) \\ x_3 & \text{if } (\theta, \hat{m}) = (\theta_3, b) \\ x_1 & \text{otherwise.} \end{cases}$$

With the concept of an extended allocation as introduced in Example 1, the incentive compatible direct mechanism $y_2(\theta_1) = (x_1, \theta_1)$, $y_2(\theta_2) = (x_2, \theta_3)$, $y_2(\theta_3) = (x_3, \theta_3)$ implements the outcomes in X_2 as intended by the social choice function f_2. $\square$

Example 3: Too many messages

Consider three types $\Theta_3 = \{\theta_1, \theta_2, \theta_3\}$ with two outcomes $X_3 = \{x_1, x_2\}$, message sets $M_3(\theta_1) = \{\theta_1, \theta_2\}$, $M_3(\theta_2) = \{\theta_2, \theta_3\}$, $M_3(\theta_3) = \{\theta_3\}$, and utilities

$u_3(x,\theta)$	θ_1	θ_2	θ_3
x_1	0	1	1
x_2	1	0	0

Consider the social choice function $f_3(\theta_1) = x_1$, $f_3(\theta_2) = f_3(\theta_3) = x_2$, inducing a utility 0 for each type. This social choice function is not $M(.)$-implementable. For suppose it is $M(.)$-implementable by some function $g_3 : \Theta \to X$. There are two cases for $g_3(\theta_2)$. Case 1: $g_3(\theta_2) = x_1$, but then type θ_2 can guarantee himself 1 by sending the message θ_2, which contradicts that he is supposed to get 0 under f_3. Case 2: $g_3(\theta_2) = x_2$, but then type θ_1 can guarantee himself 1 by sending the message θ_2, contradicting that he is supposed to get 0 under f_3. Note however that by restricting the mechanism to only messages $\{\theta_1, \theta_3\} \subset \Theta$ and setting $g_3(\theta_1) = x_1$ and $g_3(\theta_3) = x_2$, the social choice function f_3 is implementable. Hence, to implement f_3 it is crucial that the agent's communication is restricted: he is not allowed to send the message θ_2. Because Green and Laffont define a mechanism as consisting of an outcome function whose domain is the entire set of types Θ, they formally do not allow such restrictions in their framework. □

5 A Revelation Principle

While the example in Section 3 demonstrated a failure of the revelation principle, the last two examples point to an even more fundamental problem that already the class of direct (but possibly non-incentive compatible) mechanisms is too restrictive. In this section I argue that also these problems disappear when extending the notion of an economic allocation.

Following Green and Laffont (1986), one may model an economic allocation solely as the outcome $x \in X$, not including in the definition of an allocation the verifiable messages themselves. On the one hand, this modeling approach may seem intuitive, since sending a verifiable message is costless and, therefore, pay-off irrelevant. On the other hand, one may, however, just as well argue that a verifiable message is extremely pay-off relevant, since the cost of a verifiable message to a type who cannot send it is effectively infinite. The latter argument suggests to model an allocation as an outcome $x \in X$ together with the presentation of some verifiable information $\theta \in \Theta$ rather than only the outcome $x \in X$.

In order to show formally that this extended notion of an economic allocation restores the revelation principle, one first has to be more concrete about the interpretation of a verifiable message. In particular, I here follow the "exhaustive" modeling

approach of Bull and Watson (2007) and assume, without loss of generality, that the agent can send at most one verifiable message.[12]

Given a principal-agent problem $\Gamma = \{X, \Theta, M, u\}$, define the *extended allocation set* $\hat{X} \equiv X \times \Theta$ and the *extended utility function* $\hat{u}(\hat{x}|\tilde{\theta}) = \hat{u}(x, \theta|\tilde{\theta})$ as

$$\hat{u}(x, \theta|\tilde{\theta}) \equiv \begin{cases} u(x, \tilde{\theta}) & \text{if } \theta \in M(\tilde{\theta}) \\ u(x, \tilde{\theta}) - C & \text{otherwise.} \end{cases} \tag{1}$$

with $C = \max_{x,\theta,x',\theta'} u(x, \theta) - u(x', \theta')$.[13,14] Hence, the expanded structure $\hat{\Gamma} = \{\hat{X}, \Theta, M, \hat{u}\}$ represents a principal-agent problem in which the principal wants to implement an *extended social choice function* $\hat{f} : \Theta \to \hat{X}$ given that the agent is privately informed about his type θ.

A social choice function $\hat{f}$ is implementable if there exists some single-person decision problem in which for any type θ there exists an optimal decision inducing the allocation $\hat{f}(\theta)$. A special class of such decision problems are incentive compatible direct mechanisms defined as follows.

Definition $\hat{1}$: An *incentive compatible direct mechanism* in $\hat{\Gamma}$ is a composite function $\hat{g} = (\hat{g}^1, \hat{g}^2)$ with $\hat{g}^1 : \Theta \to X$ and $\hat{g}^2 : \Theta \to \Theta$ such that

$$\hat{u}(\hat{g}(\theta)|\theta) \geq \hat{u}(\hat{g}(\theta')|\theta) \text{ for any } \theta, \theta' \in \Theta. \tag{2}$$

[12]Bull and Watson (2007) view the agent's verifiable message as a collection of possible pieces of evidence. They thereby offer a micro-foundation for the verifiable messages which implies that the agent can only send one verifiable message. In contrast, Deneckere and Severinov (2008) do not model this intermediate step of differentiating between the verifiable message and its underlying pieces of evidence. As the authors explain, without this distinction, it is appropriate to model the possibility that the agent can send multiple verifiable messages. Yet, using the reinterpretation of Bull and Watson (2007), one can recast such a model as one in which the agent has 2^{Θ} messages available from which he can send only one message (see also Section 7).

[13]Because Θ and X are finite, C is well-defined. If the sets Θ and X are infinite, one may take the supremum rather than the maximum, which is well-defined provided that u is bounded. If u is unbounded, all arguments still go through by picking, for a given social welfare function f, a large enough (finite) value for C.

[14]The extension follows the idea of Harsanyi (p. 168, 1967) that "the assumption that a given strategy $s_i = s_i^0$ is not *available* to player i is equivalent, from a game-theoretical point of view, to the assumption that player i will never actually *use* strategy s [emphasis in the original]." As a consequence, it renders the agent's action set type-independent and solves the issue pointed out in footnote 11.

Hence, an incentive compatible direct mechanism $\hat{g}$ represents a single-person decision problem in which it is an optimal decision for the agent to report his type truthfully. We adapt Definition 2 to $\hat{\Gamma}$ as follows.

Definition $\hat{2}$: A social choice function $\hat{f} : \Theta \to \hat{X}$ is *$\hat{g}$-implementable* iff the direct mechanism $\hat{g} = \hat{f}$ is incentive compatible.

Standard arguments yield the revelation principle for the principal-agent problem $\hat{\Gamma}$: If there exists some single-person decision problem in which for any type θ there exists an optimal decision leading to the extended allocation $\hat{f}(\theta)$, then there exists an incentive compatible direct mechanism with $\hat{g}(\theta) = \hat{f}(\theta)$. Hence, the mechanism $\hat{g}$ implements the social choice function $\hat{f}$ and the next proposition follows.

Proposition 1 (Revelation principle) *Any extended allocation $\hat{f}(\theta)$ that is the outcome of some single-agent decision problem in $\hat{\Gamma}$ is $\hat{g}$-implementable.*

While the previous proposition establishes a revelation principle for the principal-agent problem $\hat{\Gamma}$, it leaves open its relation to the underlying problem Γ. The next proposition addresses this issue.

Proposition 2 *Consider some principal-agent problem Γ and its corresponding extension $\hat{\Gamma}$. If there exists some mechanism in Γ which implements the social choice function $f : \Theta \to X$, then there exists a function $\hat{\theta} : \Theta \to \Theta$ such that the extended social choice function $\hat{f}(\cdot) = (f(\cdot), \hat{\theta}(\cdot))$ is $\hat{g}$-implementable.*

Proof of Proposition 2: Suppose some decision problem implements the social choice function f in Γ. Then for type θ, some decision(s) leading to the outcome $f(\theta)$ and some verifiable message $\hat{\theta}(\theta) \in M(\theta)$ that he sends when achieving outcome $f(\theta)$ is optimal. Consider the direct mechanism $\hat{g} : \Theta \to \hat{X}$ with $\hat{g}^1(\theta) = f(\theta) \in X$ and $\hat{g}^2(\theta) = \hat{\theta}(\theta) \in M(\theta) \subseteq \Theta$. Fix some $\theta \in \Theta$. Inequality (2) holds for any θ' s.t. $\hat{\theta}(\theta') \notin M(\theta)$, because $u(f(\theta'), \theta) - C \leq \min_{x,\tilde{\theta}} u(x, \tilde{\theta}) \leq \hat{u}(f(\theta), \hat{\theta}(\theta))$, since $\hat{\theta}(\theta) \in M(\theta)$. Moreover, the optimality of the decision(s) leading to $f(\theta)$ and message $\hat{\theta}(\theta)$ imply that inequality (2) holds for any θ' s.t. $\hat{\theta}(\theta') \in M(\theta)$. It then follows that the constructed $\hat{g}$ satisfies (2) for any $\theta, \theta' \in \Theta$ so that $\hat{g}$ is an incentive compatible direct mechanism in $\hat{\Gamma}$. Hence $\hat{f}$ is $\hat{g}$-implementable. Q.E.D.

The main insight is therefore that, despite the presence of (partially) verifiable information, there is nothing peculiar about the principal-agent problem if one specifies the concept of an implementable allocation appropriately. One can then use the

revelation principle as usual to analyze the class of implementable allocations for all possible mechanisms. In particular, the incentive constraints (2) full characterize the set of implementable social choice functions $\hat{f}$. Taking the first component of $\hat{f}$ gives us the set of implementable outcomes $x \in X$.

6 Non-contingent Mechanisms

Proposition 2 suggests the following procedure for characterizing the set of implementable social choice functions f of any principal-agent problem with verifiable information Γ. First characterize the set of implementable social choice functions $\hat{f}$ in the corresponding problem $\hat{\Gamma}$ by the incentive constraints (2). The set of all implementable social choice function f can then be obtained in a second step by taking the first component of each implementable $\hat{f}$.

The procedure characterizes the set of implementable outcome via mechanisms $\hat{g} = (\hat{g}^1, \hat{g}^2)$ that map into the extended allocation $X \times \Theta$ rather than the outcome space X. Such mechanisms are *evidence-contingent*; the agent receives the allocation $\hat{g}^1(\theta) \in X$ conditional on presenting the evidence $\hat{g}^2(\theta) \in \Theta$. While the previous section shows that evidence-contingent mechanisms allow us to characterize the set of implementable outcomes with the usual tools of mechanism design, one may object that these mechanisms may be too coercive for some practical environments, because they effectively force the agent to present his evidence.[15]

Following this concern, I next study non-contingent mechanisms as defined as follows.

Definition: A non-contingent mechanism (M, g) consists of a set M and an outcome function $g : M \to X$.

The direct mechanisms $g : \Theta \to X$, as modelled in Green and Laffont form a subclass of non-contingent mechanisms. Yet, Examples 2 and 3 show that these direct mechanisms are too restrictive. Indeed, starting with a direct mechanism and broadening it by adding cheap-talk messages (Example 2), or reducing it further by restricting communication (Example 3), yields a non-contingent mechanism that implements an additional outcome.

[15] E.g., the "right to remain silent" is a right recognized in many of the world's legal systems.

The remainder of this section shows that the two examples identify exactly those operations on the design of mechanisms that are needed. In particular, starting with a direct mechanism in the sense of Green and Laffont and using the two elementary operations of broadening it by adding cheap-talk messages and limiting it by restricting the communication of verifiable information, yields a class of non-contingent mechanisms that can implement any outcome that is implementable by some evidence-contingent mechanism. This result implies that the class of non-contingent mechanisms is not restrictive in terms of the implementable outcomes they induce. Because the proof is constructive, it shows exactly how to obtain the non-contingent mechanism that implements the same outcome as its evidence-contingent counterpart.

Given a principal-agent problem Γ with the message correspondence $M(.)$, first extend the agent's message as follows:

$$\hat{M}(\theta) \equiv M(\theta) \times \Theta.$$

As before an interpretation of this extension is that, in addition to a message from $M(\theta)$, an agent of type θ also makes some non-verifiable claim about his type Θ.[16]

Let $\hat{M} \equiv \cup_{\theta \in \Theta} \hat{M}(\theta)$ and define a mechanism as follows:

Definition $\bar{1}$: A mechanism $(\bar{M}, \bar{g})$ in Γ consists of a set $\bar{M} \subseteq \hat{M}$ such that $\bar{M} \cap \hat{M}(\theta) \neq \emptyset$ for all $\theta \in \Theta$ and an outcome function $\bar{g} : \bar{M} \to X$.

Hence, a mechanism $(\bar{M}, \bar{g})$ is non-contingent. Moreover, it is constructed by starting with the message sets $M(\theta)$, which Green and Laffont consider as primitives of the underlying principal-agent problem, extending them by adding cheap-talk messages, in the form of the set Θ, to obtain the extended messages set $\hat{M}$, and, subsequently, restricting this overall message set to $\bar{M}$, which is a (possibly strict) subset of $\hat{M}$. Hence, if, in the mechanism design problem, the principal has the ability to perform the two elementary operations of adding cheap-talk messages and restricting the agent's communication, then it is compelling that the principal can use mechanisms as defined in Definition $\bar{1}$.

As before, a mechanism $(\bar{M}, \bar{g})$ presents the agent of type θ with a single-person decision problem in which he has to pick some m from the message set $\bar{M}$ that is

[16] This additional non-verifiable message does not have to be a literal claim about the agent's type. Another interpretation is that the agent has to say some natural number between 1 and K, which, given that there are K types, effectively is like reporting some Θ.

consistent with his message set $\hat{M}(\theta)$. That is, the mechanism induces a *response rule* $\bar{\phi}_{\bar{g}} : \Theta \to \bar{M}$ defined by[17]

$$\bar{\phi}_{\bar{g}}(\theta) \in \arg \max_{m \in \hat{M}(\theta) \cap \bar{M}} u(\bar{g}(m), \theta).$$

Because the function $\bar{\phi}_{\bar{g}}$ maps Θ into the Cartesian product $\Theta \times \Theta$, it is convenient to write the composed function $\bar{\phi}_{\bar{g}}$ component-wise as $\bar{\phi}_{\bar{g}} = (\bar{\phi}^1_{\bar{g}}, \bar{\phi}^2_{\bar{g}})$ of two functions $\bar{\phi}^1_{\bar{g}} : \Theta \to \Theta$ and $\bar{\phi}^2_{\bar{g}} : \Theta \to \Theta$. The first component represents the presentation of verifiable information, while the second component represents the cheap talk message.

The adapted notions of a mechanism and a response rule lead to the following concept of implementability.

Definition $\bar{2}$: A social choice function $f : \Theta \to X$ is $\bar{M}$*-implementable* in Γ iff there exists a mechanism $(\bar{M}, \bar{g})$ with an outcome function $\bar{g}$ such that:

$$\bar{g}(\bar{\phi}_{\bar{g}}(\theta)) = f(\theta) \text{ for any } \theta \text{ in } \Theta, \tag{3}$$

where $\bar{\phi}_{\bar{g}}(\cdot)$ is a response rule with respect to the mechanism $(\bar{M}, \bar{g})$.

The next proposition makes precise the idea that any implementable outcome is implementable by a non-contingent mechanism $(\bar{M}, \bar{g})$.

Proposition 3 *Consider a principal-agent problem Γ and the corresponding problem $\hat{\Gamma}$. If $\hat{f} = (\hat{x}, \hat{\theta})$ is $\hat{g}$-implementable in $\hat{\Gamma}$ and $\hat{\theta}(\theta) \in M(\theta)$ for all $\theta \in \Theta$, then $f = \hat{x}$ is $\bar{M}$-implementable in Γ.*

Proof of Proposition 3: Fix $\hat{f} = (\hat{x}, \hat{\theta})$ and define $\bar{M}$ as

$$\bar{M} = \{(\hat{\theta}(\theta_i), \theta_i) : \theta_i \in \Theta\}.$$

Because $\hat{\theta}(\theta) \in M(\theta)$, it holds by construction of $\hat{M}(\theta)$ that $\bar{M} \subseteq \cup_\theta \hat{M}(\theta)$. Define the outcome function $\bar{g} : \bar{M} \to X$ as $\bar{g}(\hat{\theta}(\theta), \theta) = \hat{x}(\theta)$ for any $(\hat{\theta}(\theta), \theta) \in \bar{M}$. Note $\hat{M}(\theta_i) \cap \bar{M} = (\hat{\theta}(\theta_i), i) \neq \emptyset$ for any $\theta_i \in \Theta$. Hence, $(\bar{M}, \bar{g})$ is a mechanism according to Definition $\bar{1}$. Moreover, because $\hat{M}(\theta_i) \cap \bar{M} = (\hat{\theta}(\theta_i), \theta_i)$ is a singleton, $(\hat{\theta}(\theta), \theta)$ is a response rule with respect to the mechanism $(\bar{M}, \bar{g})$. Hence, $\bar{\phi}_{\bar{g}}(\theta) = (\hat{\theta}(\theta), \theta)$

[17]The provision in Definition $\bar{1}$ that $\hat{M}(\theta) \cap \bar{M} \neq \emptyset$ for all $\theta \in \Theta$ implies that the agent does not maximize over an empty set.

so that $\bar{g}(\bar{\phi}_{\bar{g}}(\theta)) = \bar{g}(\hat{\theta}(\theta), \theta) = \hat{x}(\theta) = f(\theta)$. Therefore, $f = \hat{x}$ is $\bar{M}$-implementable. Q.E.D.

By constructively deriving the incentive compatible mechanism $\hat{g}$ that implements the same outcome as some non-contingent mechanism $(\bar{M}, \bar{g})$, the next proposition makes precise the converse of the previous proposition.

Proposition 4 *Consider a principal-agent problem Γ and the corresponding problem $\hat{\Gamma}$. If f is $\bar{M}$-implementable in Γ, then the social choice function $\hat{f} = (f, \bar{\phi}^1_{\bar{g}})$ is $\hat{g}$-implementable in $\hat{\Gamma}$, where $\bar{\phi}^1_{\bar{g}}$ is the first component of the response rule $\bar{\phi}_{\bar{g}}$ corresponding to the outcome function $\bar{g}$ satisfying (3).*

Proof of Proposition 4: Given f in Γ is $\bar{M}$-implementable, there is a $\bar{g}$ and an associated response rule $\bar{\phi}_{\bar{g}} = (\bar{\phi}^1_{\bar{g}}, \bar{\phi}^2_{\bar{g}})$ satisfying (3). Fixing functions $(\bar{g}, \bar{\phi}^1_{\bar{g}}, \bar{\phi}^2_{\bar{g}})$, consider the social choice function $\hat{f} = (f, \bar{\phi}^1_{\bar{g}})$ in $\hat{\Gamma}$ and the direct mechanism $\hat{g} = \hat{f}$. The propostion follows if $\hat{g}$ is incentive compatible, ie. satisfies (2). To show this, fix a type $\theta \in \Theta$. It follows that $\hat{u}(\hat{g}(\theta), \theta) = \hat{u}(f(\theta), \bar{\phi}^1_{\bar{g}}(\theta)|\theta) = u(f(\theta), \theta)$, because $\hat{g} = \hat{f} = (f, \bar{\phi}_{\bar{g}})$ and $\bar{\phi}^1_{\bar{g}}(\theta) \in M(\theta)$. Hence, one has to show that $\hat{u}(\hat{g}(\theta')|\theta) = \hat{u}(f(\theta'), \bar{\phi}^1_{\bar{g}}(\theta')) \leq u(f(\theta), \theta)$ for any $\theta' \in \Theta$. Note first that while it holds $\bar{\phi}_{\bar{g}}(\theta') \in \bar{M}$, one can have $\bar{\phi}^1_{\bar{g}}(\theta') \notin M(\theta)$ or $\bar{\phi}^1_{\bar{g}}(\theta') \in M(\theta)$. First, suppose that $\bar{\phi}^1_{\bar{g}}(\theta') \notin M(\theta)$, it then follows $\hat{u}(\hat{g}(\theta')|\theta) = u(f(\theta'), \theta) - C \leq \max_{\tilde{x}, \tilde{\theta}} u(\tilde{x}, \tilde{\theta}) - C = \min_{\tilde{x}, \tilde{\theta}} u(\tilde{x}, \tilde{\theta}) \leq u(f(\theta), \theta)$. Next, suppose that $\bar{\phi}^1_{\bar{g}}(\theta') \in M(\theta)$, it then follows $\hat{u}(\hat{g}(\theta')|\theta) = u(f(\theta'), \theta) = u(\bar{g}(\bar{\phi}_{\bar{g}}(\theta')), \theta) \leq u(\bar{g}(\bar{\phi}_{\bar{g}}(\theta)), \theta)$, where the inequality follows because $\bar{\phi}_{\bar{g}}(\theta)$ maximizes $u(\bar{g}(m), \theta)$ over all $m \in \hat{M}(\theta) \cap \bar{M}$, which includes $\bar{\phi}_{\bar{g}}(\theta')$. Q.E.D.

Combining these two propositions with the previous two implies that Definition $\bar{1}$ gives a canonical representation of mechanisms in the sense that, in terms of implementable outcomes, there is no loss of generality in restricting attention to these mechanisms; any implementable outcome is implementable by some mechanism corresponding to Definition $\bar{1}$.

Example 1 revisited:
As an illustration to see how one can check the implementability of any social choice function in any principal-agent problem Γ and find the non-contingent mechanism which implements it, reconsider the principal-agent problem $\Gamma_1 = (X_1, \Theta_1, M_1, u_1)$ of example 1 and the social choice function f_1. First construct the hypothetical

principal-agent problem $\hat{\Gamma}_1 = (X_1 \times \Theta_1, \Theta_1, M_1, \hat{u}_1)$ where the hypothetical utility function $\hat{u}_1$ follows from its definition in (1):

$\hat{u}_1(x,\theta\|\theta_1)$	θ_1	θ_2	θ_3
x_1	10	10	0
x_2	15	15	5

$\hat{u}_1(x,\theta\|\theta_2)$	θ_1	θ_2	θ_3
x_1	-5	5	5
x_2	0	10	10

$\hat{u}_1(x,\theta\|\theta_3)$	θ_1	θ_2	θ_3
x_1	0	0	10
x_2	5	5	15

Next check whether there exists a social choice function $\hat{f}_1 = (f_1, \hat{\theta}_1)$ that is $\hat{g}$-implementable in $\hat{\Gamma}_1$. Given that the revelation principle holds in $\hat{\Gamma}$, this can be done as usual: find an incentive compatible direct mechanism $\hat{g}_1 = (\hat{g}_1^1, \hat{g}_1^2) : \Theta \to \hat{X}$ with $\hat{g}_1^1 = f_1$ and $\hat{g}_1^2 = \hat{\theta}_1$ which satisfies the familiar incentive compatible conditions (2). Using these incentive constraints one can verify that $\hat{g}_1(\theta_1) = (x_1, \theta_1)$, $\hat{g}_1(\theta_2) = \hat{g}_1(\theta_3) = (x_2, \theta_3)$ is such an incentive compatible direct mechanism. Hence, the conclusion follows that f_1 is indeed $\bar{M}$-implementable in Γ_1.

While this procedure confirms that f_1 is $\bar{M}$-implementable by the familiar means of checking incentive constraints of direct mechanisms, it does not yield the mechanism $(\bar{M}_1, \bar{g}_1)$ which actually implements f_1 in the principal-agent problem Γ_1. The constructive proof of Proposition 3 shows how to recover this mechanism from $\hat{g}_1$. Using that $\hat{g}_1 = (\hat{g}_1^1, \hat{g}_1^2) = (f_1, \hat{\theta}_1) = \hat{f}_1$, it follows

$$\bar{M}_1 = \{(\hat{\theta}_1(\theta_i), \theta_i) : \theta_i \in \Theta\} = \{(\hat{g}_1^2(\theta_i), \theta_i) : \theta_i \in \Theta\} = \{(\theta_1, \theta_1), (\theta_3, \theta_2), (\theta_3, \theta_3)\}.$$

This set yields the required $\bar{g}_1$ after linking it to the social choice function f_1 by setting $\bar{g}_1(\hat{\theta}(\theta), \theta) = \hat{f}_1^1(\theta) = f_1(\theta)$ for each $(\hat{\theta}(\theta), \theta) \in \bar{M}_1$. For Example 1, this yields $\bar{g}_1(\theta_1, \theta_1) = x_1$, $\bar{g}_1(\theta_3, \theta_2) = \bar{g}_1(\theta_3, \theta_3) = x_2$. □

7 Evidence Sets

I showed my results in the original context of Green and Laffont (1986), which models verifiable information in a reduced form by directly hard wiring it to the agent's type. To illustrate how these results translate to a setup in which verifiable information is derived from actual pieces of evidence, let $\hat{E}$ represent the (finite) set of possible evidence and the set $\mathcal{E} \subseteq 2^{\hat{E}}$ the possible combinations of evidence which the agent may possess and present. Let $u(x, \theta)$ once again represent the utility of (payoff) type $\theta \in \Theta$ from an outcome $x \in X$. The agent's (meta) type $\hat{\theta} = (\theta, E)$ is a combination

of his payoff type $\theta \in \Theta$ together with his evidence type $E \subseteq \hat{E}$, representing the evidence he can present. Hence, a tuple $\Gamma = \{X, \hat{\Theta}, u\}$ describes the primitives of a principal-agent problem with evidence sets, where $\hat{\Theta} \subseteq \Theta \times \mathcal{E}$ represents the possible combinations of payoff and evidence types.

Denoting by $e_\emptyset$ the presentation of no evidence, we can, as before, define the set of (extended) allocations as $\hat{X} \equiv X \times (\mathcal{E} \cup \{e_\emptyset\})$ and extend the utility of a type $\hat{\theta} = (\theta, E) \in \hat{\Theta}$ over the allocations $\hat{X}$ by defining

$$\hat{u}(\hat{x}|\hat{\theta}) \equiv \begin{cases} u(x, \theta) & \text{if } E' \subseteq E \cup \{e_\emptyset\} \\ -C & \text{otherwise,} \end{cases}$$

with $C \in \mathbb{R}$ large. Hence, given the primitives $\Gamma = \{X, \hat{\Theta}, u\}$, the associated collection $\hat{\Gamma} = (\hat{X}, \hat{\Theta}, \hat{u})$ describes its associated extension.

Similar to Section 5, the revelation principle obtains in the extended representation $\hat{\Gamma} = (\hat{X}, \hat{\Theta}, \hat{u})$ but not in the original $\Gamma = \{X, \hat{\Theta}, u\}$. In particular, a social choice function $\hat{f} : \hat{\Theta} \to \hat{X}$ is implementable if and only if the direct mechanism $\hat{g} : \hat{\Theta} \to \hat{X}$ with $\hat{g}(\hat{\theta}) = \hat{f}(\hat{\theta})$ for all $\hat{\theta} \in \hat{\Theta}$ is incentive compatible. This shows how Proposition 1 translates to a setup where verifiable information is modelled as originating from evidence sets. In contrast, there exist frameworks $\Gamma = \{X, \hat{\Theta}, u\}$ and social choice functions $f : \hat{\Theta} \to X$ that are implementable but not necessarily by an incentive compatible direct mechanism $g : \hat{\Theta} \to X$.

Similarly, Proposition 2 translates as follows: If some mechanism $g : \hat{\Theta} \to X$ implements the social choice function $f : \hat{\Theta} \to X$ in $\Gamma = \{X, \hat{\Theta}, u\}$, then there is a social choice function $\hat{f} : \hat{\Theta} \to \hat{X}$ in its extension $\hat{\Gamma}$ that is implementable.

The next example illustrates the importance of restricting communication with evidence sets. It is similar in spirit to example 3, but has a simpler structure, since it requires only two types.

Example 4: Restricting communication with evidence sets
Consider two types $\Theta_4 = \{\theta_1, \theta_2\}$ with two outcomes $X_4 = \{x_1, x_2\}$ and utilities

$u_4(x, \theta)$	θ_1	θ_2
x_1	0	1
x_2	1	0

The evidence set is $E = \{e_a, e_b, e_c\}$. Type θ_1 has evidence set $E_1 = \{e_a, e_c\}$ and type θ_2 has evidence set $E_2 = \{e_b, e_c\}$.[18] Hence, there are two meta types $\hat{\theta}_1$ and $\hat{\theta}_2$ so that $\hat{\Theta} = \{(\theta_1, 2^{\{e_a,e_c\}}), (\theta_2, 2^{\{e_b,e_c\}})\}$.

Consider the social choice function $f : \Theta \to X$ with $f_4(\theta_1) = x_1$, $f_4(\theta_2) = x_2$, picking the least favorite allocation for each type. The outcome associated with this social choice function is not implementable by a mechanism, $g_4 : M \times E \to X_4$, which has the evidence set E as (part of) its domain and where M is some message set. This is so because the mechanism g_4 has to specify an allocation x_1 or x_2 if the agent supplies the verifiable evidence e_c. As either type can produce this evidence, any mechanism with the evidence set in its domain allows at least one of the types his more favorable option. In contrast, the outcome is implementable when considering the extended allocation $\hat{X} \equiv X \times (\mathcal{E} \cup \{e_\emptyset\})$ and associated mechanisms, $\hat{g}_4 : \hat{\Theta} \to \hat{X}$, which have the evidence set E as part of its range. Take $\hat{g}_4(\theta_1, \cdot) = (x_1, e_a)$ and $\hat{g}_4(\theta_2, \cdot) = (x_2, e_b)$. □

8 Conclusion

In mechanism design with (partially) verifiable information, the usual revelation principle holds if allocations are modelled as the Cartesian product of outcomes and verifiable information so that direct mechanisms are mapping from (cheap-talk) messages about types to these allocations, giving rise to evidence-contingent mechanisms. As a result, mechanism design with verifiable information does not fundamentally differ from mechanism design without verifiability. Its usual tools of direct mechanisms and incentive constraints still enable to fully characterize the set of implementable allocations.

Moreover, any outcome associated with some evidence-contingent mechanism is also implementable by a non-contingent mechanism that does not condition outcomes on the presentation of evidence, provided that such an non-contingent mechanism can use two properties: 1) use cheap-talk messages in excess of reports about the agent's private information; and 2) limit communication by restricting agents to send messages about their private information.

[18] Note that the evidence structure satisfies the full report condition in Lipman and Seppi (1995) and is even strongly normal in the sense of Bull and Watson (2007).

Since the second property is inherently possible in a setting without any verifiable information, the main conceptual difference between mechanism design with and without verifiable information is the weaker ability to restrict communication when information is (partially) verifiable.

References

Ben-Porath, E. and B. Lipman (2012), "Implementation with partial provability," *Journal of Economic Theory*, 147, 1689-1724.

Ben-Porath, E., E. Dekel, and B. Lipman (2014), "Optimal Allocation with Costly Verification", *American Economic Review* 104, 3779-3813.

Ben-Porath, E., E. Dekel, and B. Lipman (2017), "Mechanisms with Evidence: Commitment and Robustness", mimeo Boston University.

Bull, J. (2008), "Mechanism Design with Moderate Evidence Cost," *The B.E. Journal of Theoretical Economics* 8: Article 15.

Bull, J. and J. Watson (2004), "Evidence disclosure and verifiability," *Journal of Economic Theory*, 118, 1-31.

Bull, J. and J. Watson (2007), "Hard evidence and mechanism design," *Games and Economic Behavior*, 58, 75-93.

Deneckere, R. and S. Severinov (2008), "Mechanism design with partial state veriability," *Games and Economic Behavior*, 64, 487-513.

Forges, F. and F. Koessler (2005), "Communication equilibria with partially veriable types," *Journal of Mathematical Economics*, 41, 793-811.

Glazer, J. and A. Rubinstein (2004). "On Optimal Rules of Persuasion". *Econometrica* 72: 17151736.

Glazer, J. and A. Rubinstein (2006). "A study in the pragmatics of persuasion: a game theoretical approach.", *Theoretical Economics 1*, 395-410.

Hart, S., I. Kremer, and M. Perry* (2017) "Evidence Games: Truth and Commitment", *American Economic Review* 107: 690713.

Green, J., Laffont, J.-J. (1986), "Partially veriable information and mechanism design," *Review of Economic Studies*, 53, 447-456.

Hagenbach, J., F. Koessler, and E. Perez-Richet (2014), "Certifiable Pre-Play Communication: Full Disclosure," *Econometrica*, 82, 1093-1131.

Harsanyi, J. (1967), "Games with Incomplete Information Played by "Bayesian" Players, I-III. Part I. The Basic Model" *Management Science*, 14, 159-182.

Hart, S., I. Kremer, and M. Perry, (2017) "Evidence Games: Truth and Commitment," *American Economic Review*, 107, 690-713.

Kamenica, E. and M. Gentzkow (2011), "Bayesian Persuasion," *American Economic Review*, 101, 2590-2615.

Kartik, N. and O. Tercieux (2012), "Implementation with evidence," *Theoretical Economics*, 7, 323-355.

Koessler, F. and V. Skreta (2016), "Selling with Evidence," mimeo UCL, London.

Lipman, B. and D. Seppi (1995), "Robust inference in communication games with partial provability," *Journal of Economic Theory* , 66 (2), 370-405.

Postlewaite, A. and D. Schmeidler (1986), "Implementation in differential information economies," *Journal of Economic Theory*, 39, 14-33.

Rayo, L. and I. Segal (2010), "Optimal Information Disclosure." *Journal of Political Economy*, 118, 94987.

Sher, I. (2014), "Persuasion and dynamic communication." *Theoretical Economics*, 9: 99136.

Singh, N. and D. Wittman (2001), "Implementation with partial verification," *Review of Economic Design*, 6, 63-84.

Yamashita, T. (2017), "Optimal Public Information Disclosure by Mechanism Designer," mimeo Toulouse University.