CONTINUOUS IMPLEMENTATION WITH DIRECT REVELATION MECHANISMS

YI-CHUN CHEN[A], MANUEL MUELLER-FRANK[B], AND MALLESH M. PAI[C]

ABSTRACT: We investigate how a principal's knowledge of agents' higher-order beliefs impacts their ability to robustly implement a given social choice function. We adapt a formulation of Oury and Tercieux (2012): a social choice function is continuously implementable if it is partially implementable for types in an initial model and "nearby" types. We characterize when a social choice function is *truthfully* continuously implementable, i.e., using game forms corresponding to direct revelation mechanisms for the initial model. Our characterization hinges on how our formalization of the notion of nearby preserves agents' higher order beliefs. If nearby types have similar higher order beliefs, truthful continuous implementation is roughly equivalent to requiring that the social choice function is implementable in strict equilibrium in the initial model, a very permissive solution concept. If they do not, then our notion is equivalent to requiring that the social choice function is implementable in unique rationalizable strategies in the initial model. Truthful continuous implementation is thus very demanding without non-trivial knowledge of agents' higher order beliefs.

KEYWORDS: continuous implementation, robust implementation, contagion, higher-order beliefs.

JEL CLASSIFICATION: D82, D83.

[A]DEPARTMENT OF ECONOMICS, NATIONAL UNIVERSITY OF SINGAPORE, ECSYCC@NUS.EDU.SG
[B]IESE BUSINESS SCHOOL, UNIVERSITY OF NAVARRA, MMUELLERFRANK@IESE.EDU
[C]DEPARTMENT OF ECONOMICS, RICE UNIVERSITY, MALLESH.PAI@RICE.EDU
JUNE 21, 2021
The authors thank Rahul Deb, Matt Jackson, Maher Said, Satoru Takahashi, Olivier Tercieux, Siyang Xiong and Muhamet Yildiz for helpful comments. Pai was partially supported by NSF CCF-1763349.

1. INTRODUCTION

The literature on Robust Mechanism Design, starting with the seminal work of Bergemann and Morris (2005) studies settings where the designer does not perfectly understand the information structure among agents. It investigates the design of mechanisms that perform robustly well across various information structures among agents that the principal considers possible. In this paper, our aim is to isolate how a desire for robustness impacts a principal who is solely unsure about agents' higher-order beliefs, i.e. beliefs of agents about each other's beliefs etc. Distinguished contributions in the game theory literature inform us that predictions in a given strategic situation can be very sensitive to agents' higher-order beliefs (e.g. Rubinstein (1989) or Weinstein and Yildiz (2007)). Our question thus concerns how these higher-order beliefs play a role when the principal can design the game among the agents.

We start from a standard Bayesian implementation setting: there are finite sets of agents, states and alternatives, and there is a commonly known information structure that describes the information of the agents. The planner would like to (partially) implement a given social choice function, i.e. a function from profiles of types to alternatives. In this case, any Bayesian incentive compatible social choice function can be partially implemented with a direct revelation mechanism. But what if the principal is unsure about the exact information structure among agents, but would nevertheless like the social choice function to be partially implemented "close to" a reference information structure? Formally, we adapt the formulation of Oury and Tercieux (2012) and revisit the question of when a social choice function is continuously implementable.[1]

Our main results characterize when a social choice function is *truthfully* continuously implementable, i.e., using game forms corresponding to direct revelation mechanisms for the initial model. One way to interpret our restriction is that it formalizes conditions under which a principal who believes a baseline information structure and therefore uses a direct revelation mechanism is nevertheless able to implement his desired social choice function when he is "slightly" wrong. Under this interpretation, our notion of truthful continuous implementation is a robustness check to the standard revelation principle—we build on this interpretation by presenting results on the set of continuously implementable social choice functions. An alternate interpretation is that by limiting the message space, we rule out "detail-free" mechanisms that simply elicit these details from the

[1]Our paper substantially builds off their work, we defer a fuller discussion of the details of their work, the closely related characterization of Oury (2015), and other related papers to Section 6, after we have formally stated our own results.

agents and then proceed akin to standard mechanism design. Such mechanisms, it may be argued, obey the letter but not the spirit of a robustness exercise.[2]

Intuitively, the characterization depends on the underlying topology with respect to which we demand continuity. We study two well understood topologies in this setting. The first, the product topology, only preserves lower order beliefs. It is the topology studied in Oury and Tercieux (2012) (also, the topology implicitly used in Rubinstein (1989) and explicitly appealed to in Weinstein and Yildiz (2007)). The second is the uniform-weak topology of Chen, Di Tillio, Faingold, and Xiong (2010), which preserves higher-order beliefs. The latter is studied for two reasons. Firstly, we would argue, this is of independent interest: being a finer topology, continuity with respect to this topology captures a weaker notion of robustness. Conceptually, one can argue that these capture two disparate ways an information structure can be close to a given information structure: the latter involves agreement at all arbitrarily higher-order beliefs, while the former topology only constrains lower-order beliefs. Second, at a more technical level, our results in the latter are a building block for our results in the former—we detail this further below in Section 1.1. In Section 5, we develop an example of a standard government natural resource auction setting to motivate these topologies.

At a high level, our findings can be summarized thus: settings like the latter, where despite not knowing the exact information structure, the principal has information about the agents' higher-order beliefs, are not much more constraining than the baseline of *exact* knowledge of the information structure. By contrast, if the agents' higher-order beliefs may be arbitrary, then the principal is severely restricted.

Further, we show that a "revelation principle" applies for the latter notion. In that setting, if a social choice function can be continuously implemented, it can be truthfully continuously implemented by a direct revelation mechanism. A revelation principle does not obtain in the more general setting. Requiring this stronger notion, therefore, may necessitate the use of more complex mechanisms to continuously implement some social choice functions (in particular, mechanisms containing messages that are not sent by any type in equilibrium in the baseline information structure considered by the principal). Further, we provide a partial characterization of continuous implementation in this setting, and thus explain the gap between continuous implementation and truthful continuous implementation.

[2]Of course, a principal may opt for a different "simple" mechanism rather than a direct revelation mechanism. To that end, note that while our results are formally stated for direct revelation mechanisms, our proof techniques apply to any mechanism where the equilibrium in the baseline is full-range, i.e. for every message available to any agent, there is some type of agent in the baseline information structure which sends that message. We expand on this observation below after presenting our formal results.

1.1. *Model and Results*

Let us now describe the setting and our results more formally. There are finite sets of agents, states and alternatives.[3] There is given a social choice function of interest. There is a baseline information structure that the principal considers. The actual information structure that obtains among agents is unknown to the principal. We wish to understand when the social choice function can be truthfully continuously implemented: i.e. in any (epistemic) model that embeds the baseline model, there is an equilibrium of the direct revelation mechanism such that the baseline types report their types truthfully (resulting in the desired social choice function), and further the strategy of closeby types converges. We term this requirement *truthful continuous implementation* (the additional modifier of "truthful" to the notion of Oury and Tercieux (2012) reflecting our restriction to the truthful equilibrium of a direct revelation mechanism).

We study continuity with respect to two topologies on types. The first, the product topology, places no restrictions on agents' higher-order beliefs. We show that under this topology, truthful continuous implementation is equivalent to requiring that the social choice function be implementable with a mechanism such that, in the baseline model, each agent has a unique rationalizable action, and the desired alternative of the social choice function obtains if each agent plays this unique rationalizable action (Theorem 1). The second, the uniform-weak topology, (see e.g. Monderer and Samet (1989) and Chen, Di Tillio, Faingold, and Xiong (2010)) is roughly a topology that preserves higher-order beliefs. We show that under this topology, a social choice function is truthfully continuously implementable if and only if it can be implemented in Strict equilibrium in the baseline model (Theorem 2).

Finally, we shed some light on the gap between continuous implementation and truthful continuous implementation. We show that a social choice function is continuously implementable with respect to the uniform-weak topology if and only if it is truthfully continuously implementable with respect to the uniform-weak topology (Theorem 3). Therefore a revelation principle holds for continuous implementation with respect to the uniform-weak topology. However, we show that one does not get a revelation principle with respect to the product topology. [4] In particular, our methods show something stronger— if a social choice function is not truthfully continuously implementable, but is continuously implementable, then the implementing mechanism must necessarily have messages that are not being sent at the baseline.

[3]Throughout, we assume a richness condition on the environment: see Section 2.3 for details.

[4]We can give a partial characterization of continuous implementation with respect to the product topology: we show that any continuously implementable social choice function must be strictly rationalizable implementable. The converse need not be true.

At a technical level, we would like to highlight our characterization results in the product topology. To get some intuition for this result, recall the work of Weinstein and Yildiz (2007). They consider a *given* game of incomplete information. They assume a form of richness: for each player, and each action of that player, there exists a "crazy type" whose preferences make that action strictly dominant. Their main result is to show that for any action a that is rationalizable for a (normal) type in the game, there exist close-by types in the product topology for whom that action is the unique rationalizable action. The possibility of aforementioned crazy types is used to start a contagion process, with the strict dominance used to break ties. In an implementation setting, this assumption of crazy types is not well grounded, since the game form is chosen by the planner and therefore not fixed a priori. Further, we are after a partial equilibrium result, i.e. there exists one equilibrium of the game with the desired properties.[5]

Instead our result in the product topology builds off of our result in the uniform-weak topology. Closeness in the uniform-weak topology implies closeness in the product topology. By our results in the former, we know that the social choice function must be implementable in Strict Bayes-Nash Equilibrium. Recall further that we are considering implementation with DRMs, i.e. for every message an agent could send there is a corresponding type: in other words, the equilibrium has full range. Strict equilibrium implies that for that type it is a strict best response for him to send the corresponding message. We use these types as a substitute for the crazy types described above—these are sufficient since we are indeed arguing the existence (or lack thereof) of a single equilibrium.

Take any rationalizable strategy s_i for a player i. We construct a sequence of types that converge to the baseline type in the product topology for which this strategy is the unique best response, in a manner similar to Weinstein and Yildiz (2007) (and also Weinstein and Yildiz (2004): see discussion after the proof of the theorem). Roughly, put most of the mass of i's beliefs on the fact the others will play the strategies that rationalize s_i, and a small probability of the type corresponding to the strategy s_i. The latter makes this a strict best response. Therefore, at *any* Bayes-Nash Equilibrium of the incomplete information game in this model, these constructed types must be playing the rationalizable strategy s_i. From the fact that the social choice function is continuously implementable, therefore, we have rationalizable implementation as desired.

The paper is organized as follows. Section 2 defines the model. Section 3 characterizes truthful continuous implementation. Section 4 studies the original continuous implementation of Oury and Tercieux (2012) in this setting and the gap between the two. Section 5 develops an application in the context of natural resource auctions and explains the implications of our results. Section 6 discusses the related literature and connections.

[5]In this sense, there is a tighter connection between our results and those of Weinstein and Yildiz (2004), we discuss the details after we introduce our formal results. See also Weinstein and Yildiz (2011).

2. MODEL

There is a state of the world $\theta \in \Theta$, unknown to the planner. There is a set of alternatives A. Unless otherwise stated, both A and Θ are finite. There is a finite set of I agents. Agent i has a utility function $u_i : A \times \Theta \to \mathbb{R}$. Sometimes, we might refer directly to the implied ordinal preferences over alternatives, with the standard notations $\succ_{i,\theta}$ for the strict part of the preference of agent i at state θ, $\sim_{i,\theta}$ for indifferences, and $\succeq_{i,\theta}$ for weak preference.

2.1. *Epistemic Preliminaries*

A model $\mathcal{T}$ is a pair (T, κ) where $T = T_1 \times T_2 \times \cdots \times T_I$ is a countable type space and $\kappa_{t_i} \in \Delta(\Theta \times T_{-i})$ denotes the associated beliefs for each $t_i \in T_i$.

Given a type t_i in a model (T, κ), we can compute the first-order belief of t_i (i.e., his belief about Θ) by setting t_i^1 equal to the marginal distribution of κ_{t_i} on Θ. We can also compute the second-order belief of t_i (i.e., his belief about (θ, t^1)) by setting

$$t_i^2[E] = \kappa_{t_i}\left[\left\{(\theta, t_{-i}) : \left(\theta, t_i^1, t_{-i}^1\right) \in E\right\}\right], \forall E \subset \Theta \times (\Delta(\Theta))^I.$$

We can compute the entire hierarchy of beliefs $\left(t_i^1, t_i^2, \ldots, t_i^k, \ldots\right)$ iteratively.

Now, write $X^0 = \Theta$ and for each $k \geq 1$: $X^k = \left[\Delta\left(X^{k-1}\right)\right]^I \times X^{k-1}$. Observe that $t_i^k \in \Delta\left(X^{k-1}\right)$ for every $k \geq 1$. Let d^0 be the discrete metric on Θ and d^1 be the Prohorov distance on 1st-order beliefs $(\Delta(\Theta))$.[6] Then, recursively, for any $k \geq 2$, endow $\Delta\left(X^{k-1}\right)$ with the Prohorov distance d^k where X^{k-1} is endowed with the sup-metric induced by d^0, d^1,...., d^{k-1}. Mertens and Zamir (1985) construct the universal type space $T_i^* \subset \times_{k=0}^{\infty} \Delta\left(X^k\right)$. The universal type space has the property that $t_i = (t_i^1, t_i^2, \ldots) \in T_i^*$ if there exists some type t_i' in some model such that t_i and t_i' have the same n-th-order belief for every n. Endowed with the product topology, T_i^* is a compact metrizable space and admits a homeomorphism $\kappa_i^* \colon T_i^* \to \Delta(\Theta \times T_{-i}^*)$.

We say that a sequence of types $\{t_{i,n}\}_{n=1}^{\infty}$ converges uniform-weakly to a type t_i if:

$$d_i^{\text{uw}}(t_{i,n}, t_i) \equiv \sup_{k \geq 1} d_i^k\left(t_{i,n}^k, t_i^k\right) \to 0.$$

[6]For a metric space (X, ρ), the Prohorov distance between any two $\mu, \mu' \in \Delta(X)$ is

$$\inf\{\gamma > 0 : \mu'(E) \leq \mu(E^\gamma) + \gamma \text{ for every Borel set } A \subseteq X\},$$

where $E^\gamma = \{x \in X : \inf_{y \in E} \rho(x, y) < \gamma\}$.

Moreover, write $d^{\mathrm{uw}}(t_n, t) \to 0$ if $d_i^{\mathrm{uw}}(t_{i,n}, t_i) \to 0$ for each i.[7] Similarly, a sequence of types $\{t_{i,n}\}_{n=1}^{\infty}$ converges in the product topology to a type t_i if

$$d_i^{\mathrm{p}}(t_{i,n}, t_i) \equiv \sum_{k=1}^{\infty} 2^{-k} d_i^k \left(t_{i,n}^k, t_i^k\right) \to 0.$$

Again, write $d^{\mathrm{p}}(t_n, t) \to 0$ if $d_i^{\mathrm{p}}(t_{i,n}, t_i) \to 0$ for each i.

Following Oury and Tercieux (2012), for two models $\mathcal{T} = (T, \kappa)$ and $\mathcal{T}' = (T', \kappa')$, we will write $\mathcal{T} \supset \mathcal{T}'$ if $T \supset T'$, and for $t_i \in T_i' : \kappa_{t_i}[E] = \kappa'_{t_i}[(\Theta \times T'_{-i}) \cap E]$ for any measurable $E \subset \Theta \times T_{-i}$.

The principal considers a baseline model which we denote by $\overline{\mathcal{T}} = (\overline{T}, \overline{\kappa})$. We assume that the baseline model is finite, i.e., $|\overline{T}| < \infty$. For instance, this includes as a special case the standard mechanism design setting with a common prior over payoff-relevant types. More precisely, we may set $\Theta = \times_{i \in I} \Theta_i$, $T_i = \Theta_i$, and each κ_{t_i} is induced from a common prior $\mu \in \Delta(\Theta)$ such that $\mathrm{marg}_{\Theta_i} \mu[\theta_i] > 0$ for each θ_i, i.e., $\kappa_{t_i}[(\theta_i, \theta_{-i}, t_{-i})] = 1_{\{\theta_i = t_i, \theta_{-i} = t_{-i}\}} \mu(\theta_{-i} | \theta_i)$.

2.2. *Mechanisms and Notion of Implementation*

A social choice function (SCF) is a mapping $f : \overline{T}_0 \to A$ where $\overline{T}_0 \subseteq \overline{T}$. In general $\overline{T}_0 = \overline{T}$, but in some examples we may have strict containment. Assume also that $\{t_i\} \times \mathrm{supp} \kappa_{t_i} \subset \overline{T}_0$ for every $t_i \in \overline{T}_i$ (the reason for this support condition is so that social choice function is well-defined for every profile that every type considers possible).

A mechanism, denoted $\mathcal{M} = (M, g)$ is a message space M_i for each player i, with $M \equiv \prod_i M_i$, and an outcome function $g : M \to A$. A countable (respectively, finite) mechanism is one where the message space M is countable (respectively, finite) in cardinality. Given a mechanism $\mathcal{M}$ and a model $\mathcal{T}$, we write $U(\mathcal{M}, T)$ for the induced incomplete information game. A Bayes-Nash Equilibrium (BNE) is a strategy profile $(\sigma_i)_{i \in I}$ with $\sigma_i : T_i \to \Delta(M_i)$ such that for $t_i \in T_i$, each message $m_i \in \mathrm{supp}\ \sigma_i(t_i)$ maximizes the expected payoff of agent i with respect to the opponents' strategy profile σ_{-i}.

A direct revelation mechanism is defined as is standard, i.e. the message space of every player equals the set of types the principal considers possible in the baseline model, and the outcome function is denoted $g : \prod_i \overline{T}_i \to A$. We can now define truthful continuous implementation in this setting:

DEFINITION 1. *We say f is* truthfully continuously implementable w.r.t. metric d *if the direct revelation mechanism is such that for any model $\mathcal{T} \supset \overline{\mathcal{T}}$, there is a (possibly mixed) BNE σ in the game $U(\mathcal{M}, \mathcal{T})$ such that for every $t \in \overline{T}_0$:*

a. $g(t) = f(t)$, and,

[7] See Chen, Di Tillio, Faingold, and Xiong (2010) for further details about this topology.

b. for any sequence $\{t_n\} \subset T$ *with* $d(t_n, t) \to 0$, $\sigma(t_n) \to \delta_t$. .

Definition 1 is directly comparable to the definition of continuous implementation of Oury and Tercieux (2012) (Definition 2 in their paper)—see Definition 2 below for their definition in our notation. Note that truthful continuous implementation is more demanding than continuous implementation in two ways. Firstly, it fixes the form of the mechanism used: the former restricts attention to direct revelation mechanisms where the latter considers general mechanisms. Secondly, it demands robustness of a specific equilibrium of this mechanism (i.e., the truth-telling equilibrium), whereas the latter focuses on outcomes.[8]

DEFINITION 2. *Given any SCF* f, *mechanism* $\mathcal{M} = (M, g)$, *and model* $\mathcal{T} = (T, \kappa)$ *with* $\mathcal{T} \supset \overline{\mathcal{T}}$, *say that a mixed-strategy Bayes-Nash Equilibrium (BNE)* σ *continuously implements (resp. strictly continuously implements)* f *in* $U(\mathcal{M}, \mathcal{T})$ *w.r.t. a metric* d *if*

a. $\sigma|_{\overline{T}}$ *is a pure-strategy Bayes-Nash Equilibrium (resp. strict Bayes-Nash equilibrium) in* $U(\mathcal{M}, \overline{\mathcal{T}})$;
b. *For any* $t \in \overline{T}_0$, $g(\sigma(t_n)) \to f(t)$ *for any sequence of type profiles* $\{t_n\} \subset T$ *with* $d(t_n, t) \to 0$.

We say that f *is* continuously implementable (resp. strictly continuously implementable) w.r.t. metric d *if there is a mechanism* $\mathcal{M} = (M, g)$ *such that for any model* $\mathcal{T} \supset \overline{\mathcal{T}}$, *there is an equilibrium which continuously implements (resp. strictly continuously implements)* f *in* $U(\mathcal{M}, \mathcal{T})$.

2.3. *Reduced Normal Forms and a Richness Assumption*

A recurring issue in our setting is breaking indifferences, since we have no transfers. To get results within a classical implementation setting we therefore need a richness assumption.[9] In order to introduce our assumption, first consider the following standard definition of strategic equivalence adapted to our setting.

DEFINITION 3. *For a DRM* g, *we say* t_i *is* strategically equivalent *to* t_i' *for an agent* i *if agent* i *is indifferent between the two reports regardless of the state and others' reports, i.e.:*

$$\forall t_{-i}, \theta : g(t_i, t_{-i}) \sim_{i,\theta} g(t_i', t_{-i}).$$

[8] We discuss this further in Section 6.

[9] Our assumption serves the same technical role as the assumption of costly messages in Oury and Tercieux (2012) and local payoff uncertainty in Oury (2015). We discuss those assumptions when we compare to the related literature in Section 6. At a high-level though, the difference is conceptual—our assumption is one that can be verified in the context of the baseline model considered by the principal. Their assumptions are richness assumptions on the elaborations of their model with respect to which continous implementation is desired.

In light of this we can define the reduced normal-form of a DRM, again, in line with standard terminology.

DEFINITION 4. *A* reduced normal-form *of a DRM g, denoted $\tilde{g}$, is a mechanism in which all the strategically equivalent messages are identified. For each t_i, let $\tilde{t}_i$ denote the message in $\tilde{g}$ corresponding to the set of messages strategically equivalent to t_i in g.*

It is possible in the original mechanism g that two messages are strategically equivalent for some agent i but deliver different outcomes at some profile of messages from other agents, i.e. the mechanism $\tilde{g}$ is not well defined. The following assumption rules this out.

ASSUMPTION 1. *We say that a DRM g* admits a reduced normal-form *if $\tilde{g}$ is well defined, i.e., for an agent i and any two messages t_i and t_i' which are strategically equivalent, $g(t_i, \cdot) = g(t_i', \cdot)$.*

This is reminiscent of the non-bossiness assumption of Satterthwaite and Sonnenschein (1981), which is often invoked in social choice/ allocation settings. Roughly, it requires that if an agent changing his report (all else equal) changes the selected alternative, then the agent cannot be indifferent between the two alternatives. However, non-bossiness is standardly defined only for private-value settings, so we do not expound further.

This assumption is novel and therefore perhaps not well understood. Observe that the following simple richness assumption implies that Assumption 1 is always satisfied: in particular this assumption is purely on the environment rather than Assumption 1 which is on the environment *and* the desired social choice function f.

ASSUMPTION 2. *For every agent i and any two alternatives $a, a' \in A$, there is some θ such that agent i is not indifferent between a and a' under θ.*

This latter assumption may not be appropriate for some settings of interest. For example, in a private-good allocation setting, agents may be always indifferent between alternatives that only differ in the allocations of other agents. Even here, however, the desired social choice function f may be such that Assumption 1 is satisfied, even though the environment does not satisfy Assumption 2.

To see this consider the following private-good, private-value allocation setting. There are three agents $1, 2, 3$, and three alternatives $1, 2, 3$, with each alternative to be thought of as the corresponding agent getting the good. Each agent i has a type $t_i \in [0, 1]$ which is their value for receiving the good, and an outside option of 0 for not receiving the good, with $\theta = (t_1, t_2, t_3)$, $\Theta = [0, 1] \times [0, 1] \times [0, 1]$. Observe first that in this setting, Assumption 2 is not satisfied—e.g. agent 1 is always indifferent between alternatives 2 and 3. However, note that the social choice function which assigns the good efficiently, $f(t_1, t_2, t_3) = \arg\max_i(t_1, t_2, t_3)$ is such that any DRM g that implements it must satisfy

Assumption 1—an agent's report will sometimes affect her own allocation. In fact, in this example, there are no strategically equivalent messages.

In what follows, we invoke the weaker Assumption 1. The reader may mentally substitute the stronger Assumption 2 if they prefer. Either way, we emphasize that either of these assumptions are directly verifiable on the primitives of the model.

3. Characterizing Truthful Continuous Implementation

Our main result in this section is a characterization of the set of truthfully continuously implementable social choice functions in the product topology. The following definition of interim correlated rationalizable messages (c.f. Dekel, Fudenberg, and Morris (2007)) will be useful:

Definition 5. *Let $R_i^\infty(t_i, \mathcal{M})$ denote the set of* interim correlated rationalizable *messages of type t_i in $\mathcal{M}$ defined as follows:*

Let $R_i^0(t_i, \mathcal{M}) = M_i$. Inductively, for each $k \geq 1$, a message $m_i \in R_i^k(t_i, \mathcal{M})$ iff there is some $\mu \in \Delta(\Theta \times T_{-i} \times M_{-i})$ such that

R1: $m_i \in \arg\max_{m_i'} \int_{\Theta \times M_{-i}} u_i(m_i', m_{-i}, \theta) \text{ marg } \mu_{\Theta \times M_{-i}}[d\theta, m_{-i}]$;

R2: $\text{marg}_{\Theta \times T_{-i}} \mu = \kappa_{t_i}$;

R3: $\mu\left(\left\{(\theta, t_{-i}, m_{-i}) : m_{-i} \in R_{-i}^{k-1}(t_{-i}, \mathcal{M})\right\}\right) = 1.$

Then, $R_i^\infty(t_i, \mathcal{M}) \equiv \cap_{k=1}^\infty R_i^k(t_i, \mathcal{M})$.

We can now define implementation in unique rationalizable action profile:

Definition 6. *Let g be a DRM that admits a reduced normal-form. We say f is implementable in the unique rationalizable action profile in the reduced normal-form $\tilde{g}$ if for every $t \in \overline{T}$, $R^\infty(t, \tilde{g}) = \{\tilde{t}\}$.*

Note that this definition is slightly stronger than rationalizable implementation: the latter only requires that every rationalizable action profile results in the desired alternative, while in addition, we require that the implementing mechanism have a unique rationalizable strategy for each type.

Theorem 1. *Suppose that Assumption 1 holds. An SCF f is truthfully continuously implementable w.r.t. d^p by a DRM g if and only if it is implementable in unique rationalizable action profile in $\tilde{g}$ in the sense of Definition 6.*

Since this proof is fairly involved, a high level overview may be useful to help orient the reader. Sufficiency is fairly straightforward—if $\tilde{g}$ implements f in unique rationalizable action, then g truthfully continuously implements f—this follows straightforwardly from the upper hemicontinuity of the rationalizable correspondence.

The nontrivial direction is therefore necessity, i.e. to show that if an SCF f is truthfully continuously implementable (in the product topology) then f must be implementable in the unique rationalizable action in the sense of Definition 6.

As a key building block we use our characterization of truthful continuous implementation in uniform-weak topology below (Theorem 2). Combined with Corollary 1 this tells us that an SCF f is truthfully continuously implementable w.r.t. the uniform-weak topology if and only if it is implementable in Strict Bayes-Nash Equilibrium in the "reduced normal form." From this fact, and the fact that the uniform-weak topology is finer than the product topology, we have that if f is truthfully continuously implementable (in the product topology) then f is implementable in Strict Bayes-Nash Equilibrium.

Recall further that we are considering implementation with DRMs, i.e. for every message an agent could send there is a corresponding type: in other words, the equilibrium has full range. Strictness implies that for the type corresponding to a particular message it is a strict best response for him to send the corresponding message. We use this fact as a substitute for the costly messages of Oury and Tercieux (2012) or the local payoff uncertainty of Oury (2015).

Take any type t_i' which is a rationalizable report for a player i of type $t_i \in \overline{T}_i$. We can construct a sequence of types t_i^n that converge to t_i in the product topology for which reporting t_i' is the unique best response, in a manner similar to Weinstein and Yildiz (2007) (and also Weinstein and Yildiz (2004)). Roughly, put most of the mass of i's beliefs on the fact the others will play the strategies that rationalize t_i, and a small probability that the type is t_i'. The latter makes reporting t_i' a strict best response. Therefore, at *any* Bayes-Nash Equilibrium of the incomplete information game in this model, these constructed types must be playing the rationalizable message t_i'. Since the social choice function is continuously implementable, therefore, we have rationalizable implementation as desired.

3.1. *Uniform-Weak Topology*

We now introduce our characterization of truthful continuous implementation in the uniform-weak topology. As we pointed out above, this is useful as a stepping stone to the characterization in the product topology. Since continuity with respect to the uniform-weak topology captures a weaker notion of robustness, these results may be of independent interest. To state and prove our characterization, we introduce two more terms. We say that DRM g *strictly rewards truth-telling* at type t_i over type t_i' for agent i if

$$\sum_{(\theta,t_{-i})\in\Theta\times\overline{T}_{-i}} \left[u_i\left(g(t_i,t_{-i}),\theta\right) - u_i\left(g(t_i',t_{-i}),\theta\right)\right]\,\overline{\kappa}_{t_i}\left[(\theta,t_{-i})\right] > 0.$$

We say that t_i always weakly dominates t_i' for agent i in DRM g if

$$\forall\,(\theta,t_{-i})\in\Theta\times\overline{T}_{-i}:\quad u_i\left(g(t_i,t_{-i}),\theta\right) - u_i\left(g(t_i',t_{-i}),\theta\right) \geq 0.$$

The following lemma is key to our characterization.

LEMMA 1. *If an SCF f is truthfully continuously implementable by a DRM g with respect to d^{uw}, then, for every agent i and any pair of agent i's types t_i and t_i', either g strictly rewards truth-telling at t_i over t_i'; or t_i always weakly dominates t_i' in g.*

Suppose there exists an agent i and a pair of types t_i and t_i' such that g neither strictly rewards truth-telling at t_i, nor does t_i always weakly dominate t_i'. This in particular means that there is some state θ' and some profile of other agents' reports t'_{-i} at which agent i strictly prefers to report t_i' over t_i. We show that there exists a sequence of perturbations which converges to t_i in the uniform-weak topology, such that each type in this sequence uniquely prefers to report t_i' in the DRM. Roughly speaking, these are types that put a small mass on the state that the type is θ' and the other agents' types are t'_{-i}, but are otherwise identical to t_i. Thus the conditions described in Lemma 1 are a necessary condition for truthful continuous implementation in this setting.

Our main characterization of truthful continuous implementation follows:

THEOREM 2. *An SCF f is truthfully continuously implementable by a DRM g with respect to d^{uw} if and only if for every agent i and any pair t_i and t_i', either g strictly rewards truth-telling at t_i over t_i'; or t_i is strategically equivalent to t_i' for agent i.*

The proof of this theorem is easy to describe. The necessity of our condition is straightforward in light of Lemma 1. If g does not strictly reward truth-telling at t_i over t_i', then by the condition of the Lemma, t_i must always weakly dominate t_i'. But then g cannot strictly reward truth-telling at t_i' over t_i either. This must imply that t_i' also always weakly dominates t_i, which implies that t_i and t_i' are strategically equivalent. We show the sufficiency of our condition constructively.

COROLLARY 1. *Suppose that Assumption 1 holds. f is truthfully continuously implementable in d^{uw} if and only if the reduced normal-form DRM $\tilde{g}$ implements f in truthful strict BNE in $U\left(\mathcal{M}, \overline{\mathcal{T}}\right)$, i.e. if truthtelling is a strict Bayes-Nash equilibrium in the game $U\left(\mathcal{M}, \overline{\mathcal{T}}\right)$.*

As an aside we should note that similar permissive results would be achieved if we considered closeness in the strategic topology of Dekel, Fudenberg, and Morris (2006). This follows from a result of Chen, Di Tillio, Faingold, and Xiong (2010) who show that the two topologies are equivalent around finite types (recall that by assumption the baseline model was finite).

4. A REVELATION PRINCIPLE FOR CONTINUOUS IMPLEMENTATION?

So far, we have only studied truthful continuous implementation. We now recall the definition of continuous implementation in this setting and consider the relation between continuous implementation and truthful continuous implementation for both topologies.

We begin with a positive result, i.e. that if requiring continuous implementation with respect to the uniform-weak topology, we have a revelation principle.

To state and prove our characterization of continuous implementation, we adapt two definitions to this environment. Fix a mechanism $\mathcal{M} = (M, g)$. For agent i's type t_i in $\overline{T}_i$ and message $m_i' \in M_i$, we say that f *strictly rewards* $\bar{\sigma}(t_i)$ over m_i' in a (pure-strategy) BNE $\overline{\sigma}$ in $U\left(\mathcal{M}, \overline{\mathcal{T}}\right)$ if

$$\sum_{(\theta,t_{-i})\in\Theta\times\overline{T}_{-i}} \left[u_i\left(g\left(\overline{\sigma}_i\left(t_i\right), \overline{\sigma}_{-i}\left(t_{-i}\right)\right)\right), \theta) - u_i\left(g\left(m_i', \overline{\sigma}_{-i}\left(t_{-i}\right)\right)\right), \theta)\right] \overline{\kappa}_{t_i}\left[(\theta, t_{-i})\right] > 0.$$

We say that t_i always weakly dominates m_i' in a (pure-strategy) BNE $\overline{\sigma}$ in $U\left(\mathcal{M}, \overline{\mathcal{T}}\right)$ if

$$\forall\left(\theta, t_{-i}\right) \in \Theta \times \overline{T}_{-i}: \quad u_i\left(g\left(\overline{\sigma}_i\left(t_i\right), \overline{\sigma}_{-i}\left(t_{-i}\right)\right)\right), \theta) - u_i\left(g\left(m_i', \overline{\sigma}_{-i}\left(t_{-i}\right)\right)\right), \theta) \geq 0.$$

The following lemma is again the key to our characterization of continuous implementation. The proof is analogous to the proof of Lemma 1.

LEMMA 2. *If $\overline{T}_0 = \overline{T}$ and f is continuously implementable w.r.t. d^{uw}, then there is a pure-strategy BNE $\overline{\sigma}$ in $U\left(\mathcal{M}, \overline{\mathcal{T}}\right)$ such that for each agent i, each type t_i in $\overline{T}_i$ and message $m_i' \in M_i$, either f strictly rewards $\overline{\sigma}(t_i)$ over m_i'; or $\overline{\sigma}(t_i)$ always weakly dominates m_i' in BNE $\overline{\sigma}$.*

Lemma 2 immediately implies the following characterization (as well as revelation principle) for continuous implementation in d^{uw}. Denote by $\tilde{f}$ the reduced normal form of the DRM f.

THEOREM 3. *Suppose that Assumption 1 holds and $\overline{T}_0 = \overline{T}$. f is continuously implementable in d^{uw} if and only if the reduced normal-form DRM $\tilde{f}$ implements f in truthful strict BNE in $U\left(\mathcal{M}, \overline{\mathcal{T}}\right)$.*

The basic idea of Theorem 3 is analogous to the proof of Theorem 2. The main difference is that we need Assumption 1 to ensure that the reduced normal-form is well-defined. We can then apply similar arguments. Comparing to Theorem 2, we therefore have that, with respect to the uniform-weak topology, a social choice function is continuously implementable iff it is truthfully continuously implementable.

PROOF. ($\Rightarrow$) Let $\mathcal{M} = (g, M)$ be a mechanism such that BNE σ continuously implements f. Consider the direct revelation mechanism $\mathcal{M}' = (g', \bar{T})$ defined as $g'(t) = g(\sigma(t))$ for all $t \in \bar{T}$. By Lemma 2 and Theorem 2 such a mechanism clearly truthfully continuously implements f. The implication now follows from Corollary 1.

($\Leftarrow$) By Corollary 1, an scf f satisfying this condition is truthfully continuously implementable and therefore trivially, also continuously implementable. ■

4.1. *Product Topology*

In this section, we first show by counterexample that a revelation principle does not apply to continuous implementation with respect to the product topology. In particular, we show an example below in which the direct revelation mechanism does not continuously implement the desired social choice function (in particular, since it is easily verified that this fails the characterization of Theorem 1). We then constructively show that there is a mechanism which contains additional messages and continuously implements the desired social choice function. The example is essentially due to Oury and Tercieux (2012) (working paper version).

There are 2 agents. Each claims an object, in state θ_i, $i = 1, 2$, agent i is the legitimate owner. The set of outcomes is $A = \{(x, p_1, p_2) : x \in \{0, 1, 2, 3\}, p_1, p_2 \in \{0, \underline{\xi}, \xi, \overline{\xi}\}\}$, . If $x = 0$, the object is not given to either player, $x = 1$ or 2 connotes that it was given to the respective player, while $x = 3$ implies that neither player gets the object and both are punished. The p_i's correspond to payments from the agents to the principal. Utility functions are quasilinear and the object has a monetary value to each player. The value is v_H if the player is the true owner, $0 < v_L < v_H$ otherwise. Finally, the punishment outcome $x = 3$ is equivalent to a fine of f_L to the agent if she is the legitimate owner, and $f_H > f_L > 0$ if not.

The baseline type-space of each agent is $\{\theta_1, \theta_2\}$ with the $\bar{\kappa}(\cdot)$ being the appropriate common knowledge function.[10] The social choice function the principal would like to continuously implement is $f(\theta_i, \theta_i) = (i, 0, 0)$, $f(\theta_i, \theta_j) = (0, \xi, \xi)$ with $\xi > f_L$.

Claim 1. *This social choice function is not truthfully continuously implementable wrt d^P.*

Proof. A direct revelation mechanism in this setting has exactly two messages for each player, one corresponding to each type. The claim follows from the characterization of Theorem 1 since both messages are rationalizable for both types. ■

Claim 2. *There exists an indirect mechanism that continuously implements f with respect to d^P.*

[10] Note that $\bar{\kappa}$ thus defined results in a "diagonal" typespace that does not fit our model's requirement that the typespace be a product space. This is for expositional simplicity. Remark 1 below shows how the example extends appropriately.

PROOF. Consider an indirect mechanism where each player has 3 possible messages, (Mine, His, Mine+). The outcome is given by the matrix below with $v_L < \overline{\xi} < v_H$, $f_L < \underline{\xi} < f_H$ and $\underline{\xi} < \xi$.

	Mine	His	Mine+
Mine	$(0, \xi, \xi)$	$(1, 0, 0)$	$(2, \xi, \overline{\xi})$
His	$(2, 0, 0)$	$(0, \xi, \xi)$	$(0, \underline{\xi}, 0)$
Mine+	$(1, \overline{\xi}, \xi)$	$(0, 0, \underline{\xi})$	$(3, 0, 0)$

At θ_1, action "His" is strictly dominated by "Mine+" for player 1. Consequently, "Mine" and *Mine+* are strictly dominated by "His" for player 2. Finally, in the third round, "Mine" is strictly better than "Mine+" for Player 1. Analogous reasoning follows for type θ_2. Hence "Mine" is the unique rationalizable action for type θ_1, and "His" for type θ_2. Playing this rationalizable action results in the desired social choice function being implemented.

Therefore, the mechanism described above continuously implements the social choice function f w.r.t. d^P because the interim correlated rationalizable correspondence is upper-hemicontinous (see proof of sufficiency of Theorem 1). ■

REMARK 1. *Observe that the example as stated is one where agents' types are common knowledge among the agents. This example is easily perturbed to one with a product typespace. To that end consider a slight perturbation of this model where type θ_1 believes the other agent is of type θ_2 with probability $(1 - \varepsilon)$, and type θ_2 with probability ε, and θ_2's beliefs are defined analogously. This will satisfy our requirements for some $\varepsilon > 0$ small enough. To see this, observe that both Claims 1 and 2 continue to hold as written. Firstly, each action remains strictly rationalizable in the direct revelation mechanism (Claim 1). Further, since the set of rationalizable actions is appropriately upper-hemicontinuous (see e.g. Theorem 2 of Dekel, Fudenberg, and Morris (2006)), Claim 2 follows for ε small enough.*

4.2. *A partial characterization for Indirect Mechanisms*

Finally, we provide some results about continuous implementation with respect to the product topology in indirect mechanisms. We assume that $\overline{\mathcal{T}}$ has full support, i.e., for each $t_i \in \overline{T}_i$, we have supp$\overline{\kappa}_{t_i} = \overline{T}_{-i}$. Some new definitions are now necessary. We say that m_i is strategically equivalent to m_i' for agent i in BNE $\overline{\sigma}$ in $U\left(\mathcal{M}, \overline{\mathcal{T}}\right)$ if

$$\forall \left(\theta, t_{-i}\right) \in \Theta \times \overline{T}_{-i} : \quad u_i \left(g \left(m_i, \overline{\sigma}_{-i} \left(t_{-i}\right)\right), \theta\right) = u_i \left(g \left(m_i, \overline{\sigma}_{-i} \left(t_{-i}\right)\right), \theta\right).$$

The following assumption is essentially Assumption 1 adapted to indirect mechanisms.

ASSUMPTION 3. *For any agent i and any two messages m_i and m_i' which are strategically equivalent for some BNE $\overline{\sigma}$ in $U\left(\mathcal{M}, \overline{\mathcal{T}}\right)$, we have $g\left(m_i, \cdot\right) = g\left(m_i', \cdot\right)$.*

THEOREM 4. *Suppose that Assumption 3 holds for mechanism $\mathcal{M}$. Then, f is continuously implementable in d^p if and only if it is strictly continuously implementable in d^p.*

It is worth connecting our results to Oury and Tercieux (2012). There, Theorem 3 shows that any social choice function that is strictly continuously implementable must satisfy a form of monotonicity (formally, strict interim rationalizable monotonicity, see Definition 8 of that paper). The present theorem effectively shows that under Assumption 3, the same implication extends to all continuously implementable social choice functions.

DEFINITION 7. *Let $\mathcal{T} = (T, \kappa)$ be a model. Denote by $W_i^\infty(t_i, \mathcal{M})$ the set of (interim correlated)* strictly rationalizable *messages of type t_i in $U(\mathcal{M}, \mathcal{T})$ defined as follows:*

Let $W_i^0(t_i, \mathcal{M}) = M_i$. Inductively, for each $k \geq 1$, a message $m_i \in W_i^k(t_i, \mathcal{M})$ iff there is some $\mu_{-i} \in \Delta(\Theta \times T_{-i} \times M_{-i})$ such that

R1: $\{m_i\} = \arg\max_{m_i'} \sum_{\theta, m_{-i}} u_i(m_i', m_{-i}, \theta) \text{ marg } \mu_{\Theta \times M_{-i}}[\theta, m_{-i}]$;

R2: $\text{marg}_{\Theta \times T_{-i}} \mu_{-i} = \kappa_{t_i}$;

R3: $\mu_{-i}\left(\left\{(\theta, t_{-i}, m_{-i}) : m_{-i} \in W_{-i}^{k-1}(t_{-i}, \mathcal{M})\right\}\right) = 1.$

Then, $W_i^\infty(t_i, \mathcal{M}) \equiv \cap_{k=1}^\infty W_i^k(t_i, \mathcal{M})$.

We can now define implementation in strictly rationalizable action profiles:

DEFINITION 8. *We say f is implementable in strictly rationalizable action profiles by mechanism $\mathcal{M}$ if for every $t \in \overline{T}$, we have $g(m) = f(t)$ for every $m \in W^\infty(t, \mathcal{M})$.*

THEOREM 5. *Suppose that Assumption 3 holds. An SCF f is continuously implementable w.r.t. d^p by a finite mechanism only if f is implementable in strictly rationalizable action profiles by a finite mechanism.*

As we pointed out earlier, our proof techniques in Theorem 1 apply to any mechanism such that at the baseline information structure there is an equilibrium which is both full-range and implements our desired social choice function. The desideratum of "equilibrium continuous implementation" would be defined with respect to this equilibrium, by analogy to definition 1. It should be clear that our characterization of Theorem 1 continues to hold in such a case. The gap between Theorem 1 and Theorem 5 is that the latter allows for mechanisms that contain messages not sent by any type in the baseline information structure (as in the construction of Claim 2). This also further clarifies the trade-off between Oury and Tercieux (2012) and our paper. The trade-off is not that they allow indirect mechanism whereas we focus on direct revelation mechanisms. Our approach has more bite in the classical literature where messages are cheap talk. This enables us to study the robustness of the revelation principle (Assumption 3 reduces to Assumption 1 when applied to direct revelation mechanisms and truthful strategies being the equilibrium). The cost is that we need these kinds of "richness" assumptions to make any progress. Conversely, their approach needs no such richness assumption, but instead appeals to a vanishing cost of messages. This allows them to provide a full characterization of continuous implementation of a social choice function. In particular they show that continuous implementation is equivalent to rationalizable implementation of the social choice function in the baseline environment.

We should note that Theorem 5 only provides necessary but not sufficient conditions: the strict rationalizable correspondence need not be upper-hemicontinuous. Therefore we cannot conclude that a social choice function that satisfies this condition will be continuously implementable with respect to the product topology. Of course, we know from Oury and Tercieux (2012) that rationalizable implementability of the social choice function is sufficient. There is, therefore, a gap between the necessary and sufficient conditions in this setting. A full characterization appears out of reach.

5. An Example: Natural Resource Auctions

It may be useful at this stage to develop an example to help readers appreciate the implications of our results in a classical applied mechanism design setting.[11] To that end consider the following variant of a natural resource auction model. For ease of exposition and description, the model we describe below has sets of types and alternatives that are (uncountably) infinite in cardinality, so our results formally do not apply. However, it should be clear that this model can be appropriately discretized so that our results directly apply (at the cost of clarity/ brevity).

There is a principal (e.g. the government) who wishes to auction a license to utilize a natural resource, e.g. a license to drill wells at a particular tract of land. The tract has an unknown quantity of oil q. For simplicity, we assume that the price of oil $p \in \mathbb{R}_+$ is commonly known. There are 2 competing buyers. Buyer $i = 1, 2$ has a privately known fixed-cost c_i to operate the drill. The net value to buyer i of winning the license to operate the tract for a license fee of l is therefore $pq - c_i - l$. The quantity of oil in the tract is not known. Instead, each buyer receives surveys, with information about the quantity of oil q underlying: in particular, a survey contains a noisy estimate $e = q + \varepsilon$ where ε is mean-0 noise.

The set of alternatives the principal considers is $A = \{1, 2\} \times \mathbb{R}_+ \times \mathbb{R}_+$, that is to say which buyer the license is allotted to and how much each buyer is charged.

[11] We thank Muhamet Yildiz for suggesting this application.

Baseline Information Structure. In the baseline considered by the principal, each buyer's cost c_i is i.i.d. drawn from a commonly known distribution $\mathcal{C}$ with support on some interval $[\underline{c}, \overline{c}] \subset \mathbb{R}_+$. Further, the true quantity q is distributed according to a commonly known distribution $\mathcal{Q}$ with support on interval $[0, \overline{q}] \subset \mathbb{R}_+$. So the private information of a buyer can be denoted by a $K+1$-tuple (c_i, e_i) where $e_i = (e_i^1, \ldots, e_i^K)$. Define $q(e_i)$ as the posterior expected quantity of a buyer who sees surveys (estimates) e_i.

In the baseline information structure, the buyers both receive the same set of surveys, i.e. for any $k \leq K$, $e_1^k = e_2^k$. Note that the baseline information structure is therefore effectively one of pure private values (there is complete information of the quantity of oil, and therefore the only heterogeneity among buyers is the private costs).

The principal is interested in continuously implementing the social choice function corresponding to the dominant strategy outcome of a second-price auction in this setting. That is to say, the scf of interest is defined thus: when i has private information (c_i, e_i), calculate $pq(e_i) - c_i$, i.e. the gross expected value of buyer i for the tract. Assign the good to the buyer with the higher expected value, and charge them the other's expected value. Observe that the realized net utility of the buyer depends on the actual amount of oil present in the resource.

Remark 2. *The direct revelation mechanism version of this is to ask each buyer to report their private information (c_i, e_i), and calculating their expected value from this and then running a second-price auction. Standard results from auction theory tell us that reporting private information truthfully is a weakly dominant strategy in this mechanism.*

For any buyer, the set of all reports that results in the same expected value are strategically equivalent. In the reduced-normal-form, therefore, the set of strategies for the buyers is simply the range of feasible expected values— and we know that bidding truthfully is a strict Bayes-Nash equilibrium in this reduced normal form. However note there exist other rationalizable strategies— for example, it is an equilibrium in this setting for one agent to report corresponding to the highest possible bid ($p\overline{q} - \underline{c}$) and the other agent to report corresponding to a bid of 0.

Now let us investigate two possible perturbations of this baseline information structure:

(1) In the first, it is common knowledge that there are K surveys, but each buyer places some small probability that the other only saw some subset of these surveys. As motivation, consider that all the surveys were commissioned by some public authority, but agents consider the possibility that the other did not get access to some of the surveys in time before submitting a bid.
(2) In the second, the total number of surveys K is not common knowledge. Each buyer places some small probability that the other saw additional surveys. For

example, there are a large number of experts who have informative estimates, and each buyer considers the possibility that the competitor consulted additional experts privately.

Intuitively, the former corresponds to perturbations that remain close to the baseline in the uniform-weak topology. The latter on the other hand is an e-mail game type structure that corresponds to perturbations in the product topology.

In light of Remark 2, our results now have the following implications:

(1) Theorem 1 tells us that the principal's desired social choice function cannot be truthfully continuously implemented with respect to perturbation (2). To see why, note that even though the baseline is one of private-values, perturbation (2) exposes each agent to (severe) adverse selection. Each agent considers the possibility that their competitor has a more accurate estimate of the quantity of oil in the well, and therefore will only win when the true amount of oil is smaller than the expectation conditional on their own signals. One can therefore construct a sequence of types converging to the baseline whose unique equilibrium strategy is to bid 0.

(2) Theorem 2 tells us that the principal's desired social choice function can be truthfully continuously implemented with respect to perturbation (1). Theorem 3 tells us that expanding the class of mechanisms does not expand the set of social choice functions that can be continuously implemented with respect to perturbations of type (1).

6. Related Literature

There is a large, influential literature on the connection between higher-order beliefs and strategic behavior, beginning with the email game paper of Rubinstein (1989) and the subsequent global games paper of Carlsson and Van Damme (1993), too large to comprehensively cite here. Indeed, within this field there are now at least two influential approaches: the ex-ante approach of e.g. Kajii and Morris (1997), and the interim approach of Weinstein and Yildiz (2004) and Weinstein and Yildiz (2007). As we stated earlier, our approach borrows ideas from the latter.

There is also a large literature considering robustness in mechanism design. It bifurcates into "global" and "local" approaches.[12] In global approaches (see e.g. the pioneering works of Bergemann and Morris (2005); Chung and Ely (2007)) the planner has no information on the information structure (model) that will prevail among agents. The planner wishes to implement the social choice function on all models she considers possible. By contrast, in the local approach (see e.g. Chung and Ely (2003), Oury and Tercieux

[12]While we will not dwell on these, intermediate notions of robustness, where the principal rules out some possible beliefs among the agents, have also been recently formulated and characterized—see e.g. Ollár and Penta (2017).

(2012), Jehiel, Meyer-ter Vehn, and Moldovanu (2012) or Aghion, Fudenberg, Holden, Kunimoto, and Tercieux (2012)) the planner has some specific model in mind but is not entirely confident about it. The requirement therefore is analogously local, i.e. that the social choice function be implemented at types close to the initial model. This paper falls in the latter camp so we focus our discussion on related works in this vein.

The formulation of a "local" approach to robustness that we use in this paper was pioneered by Oury and Tercieux (2012). Our results have some counterparts to theirs. We therefore first discuss the connection to their paper before mentioning other work.

The biggest difference in setups is that we mainly consider implementation by "direct revelation mechanisms." This assumption allows us tighter characterizations of (truthful) continuous implementation under the product topology. In the "forward" direction they consider the stronger desideratum of strict continuous implementation, and show that strict monotonicity of the social choice function is necessary for strict continuous implementation. To get a full characterization, and to study continuous implementation directly (as opposed to strict continuous implementation), they enrich the model to consider that sending various messages may involve small costs to the agents. By contrast, our assumptions allow us a full characterization without either (i.e. the strengthening of desideratum to strict continuous implementation, nor the possibility of costly messages). Another critical difference between our result and theirs is that our Theorem 1 is a characterization for the implementing DRM whereas their counterpart (Theorem 4) is a characterization of implementability (i.e., the mechanism that achieves rationalizable implementation is different from the mechanism that achieves continuous implementation in general (and also in their proof)).

They do not consider the uniform-weak topology but do hint at similar results in one direction (see, e.g., Footnote 16 of their paper). Our results on the uniform-weak topology thus both strengthen their results, and also constitute a key intermediate step to our characterization in the product topology.

Another closely related paper is that of Oury (2015), who characterizes continuous implementation as equivalent to full implementation in rationalizable strategies by introducing local payoff uncertainty of the planner. Assumption 1 in that paper embeds the set of states Θ into a larger set of states Θ^*, where these additional states allow to resolve indifferences.[13]

At a high level then, the difference between our approach and these two papers is that they consider general mechanisms, and obtain their characterization by extending the

[13] In our notation, the definition of local payoff uncertainty is as follows (Assumption 1) —there is a baseline model $\overline{\mathcal{T}}$, and the set of states of the world considered by types in the baseline is Θ. However, the principal envisages a larger set of states Θ^*, where $\Theta \subseteq \Theta^*$ and for every agent i, action a and state θ there exists a state $\theta^*(\theta, a, i)$ such that $u_i(a, \theta^*(\theta, a, i) > u_i(a, \theta)$ and $u_i(a', \theta^*(\theta, a, i)) = u_i(a', \theta)$ for any other $a' \neq a$.

model (costly messages in the case of Oury and Tercieux (2012), additional states in the case of Oury (2015)). We instead cover only direct revelation mechanisms, and look at the robustness of a specific equilibrium (truthful equilibrium). Conversely our richness conditions (e.g. Assumption 1 or 2) can be verified directly within the benchmark model $\overline{\mathcal{T}}$ and our robustness exercise requires no extra payoff-relevant perturbation beyond what's specified in the benchmark set of states Θ.

A recent closely related paper that takes a different approach is Takahashi and Tercieux (2011): they study robust equilibrium *outcomes* rather than robust equilibrium *behaviors* (recall our discussion after Definition 1). Formally, they look at sequential games where there is almost common certainty of payoffs (for our purposes, "almost" refers to being close in the uniform-weak topology). The latter means that their results do not directly apply to our setting: our Theorem 3 requires the domain of the SCF to have a product structure, while almost common certainty implies the baseline typespace is entirely the diagonal. That said, their results imply that if a social choice function is implemented via a mechanism with a unique subgame perfect equilibrium outcome, continuous implementation in the uniform-weak topology follows. Therefore, when considering continuous implementation with respect to the uniform-weak topology around common certainty of payoffs, a revelation principle does not apply.[14]

As we alluded to earlier, other papers have raised similar questions about "local" robust implementation. Chung and Ely (2003) ask about the possibility of (full) implementation in undominated Nash equilibrium while additionally requiring that Bayes-Nash equilibria of settings with arbitrarily small uncertainty also be close to the social choice function. They show that monotonicity of the social choice function is a necessary condition in their setting (while full implementation in undominated Nash equilibrium is possible for any social choice function under complete information). Aghion, Fudenberg, Holden, Kunimoto, and Tercieux (2012) consider subgame-perfect implementation under similar perturbations. Jehiel, Meyer-ter Vehn, and Moldovanu (2012) get a negative result similar in interpretation to ours, but in a different setting, where the multi-dimensionality of agents' signals drives the result. Postlewaite and Wettstein (1989) pursue the idea of a feasible, continuous function that achieves Walrasian outcomes in an exchange economy. Continuity is with respect to small perturbations of the initial endowments, as a substitute to modeling incentive constraints.

Our work is also connected to the literature on informational size beginning with McLean and Postlewaite (2002). These papers consider settings close to complete information,

[14] In fact, the failure of revelation principle occurs because of our requirement that the equilibrium strategies of close by types converges to truth-telling (part (b) of Definition 1) instead of the restriction of using a DRM. Further details are available from the authors on request. We thank Satoru Takahashi and Olivier Tercieux for discussions on this topic.

and argue what can be thought of as continuity results—when the state is approximate common knowledge, small transfers are sufficient to elicit the private information of agents. Most papers in this line consider settings with transfers, except Gerardi, McLean, and Postlewaite (2009). Our results in the uniform-weak topology can be thought of as complementing their findings—both suggest that in settings with approximate common knowledge of the information structure, a desired social choice function may be implemented. While they consider richer settings, they also assume a common prior among agents that is known to the principal.

References

AGHION, P., D. FUDENBERG, R. HOLDEN, T. KUNIMOTO, AND O. TERCIEUX (2012): "Subgame perfect implementation under information perturbations," *Quarterly Journal of Economics*, 127(4), 1843–1881.

BERGEMANN, D., AND S. MORRIS (2005): "Robust mechanism design," *Econometrica*, 73(6), 1771–1813.

CARLSSON, H., AND E. VAN DAMME (1993): "Global games and equilibrium selection," *Econometrica*, 61(5), 989–1018.

CHEN, Y.-C., A. DI TILLIO, E. FAINGOLD, AND S. XIONG (2010): "Uniform topologies on types," *Theoretical Economics*, 5(3), 445–478.

CHUNG, K.-S., AND J. C. ELY (2003): "Implementation with Near-Complete Information," *Econometrica*, 71(3), 857–871.

——— (2007): "Foundations of dominant-strategy mechanisms," *Review of Economic Studies*, 74(2), 447–476.

DEKEL, E., D. FUDENBERG, AND S. MORRIS (2006): "Topologies on types," *Theoretical Economics*, 1, 275–309.

——— (2007): "Interim correlated rationalizability," *Theoretical Economics*, 2(1), 15–40.

GERARDI, D., R. MCLEAN, AND A. POSTLEWAITE (2009): "Aggregation of expert opinions," *Games and Economic Behavior*, 65(2), 339–371.

JEHIEL, P., M. MEYER-TER VEHN, AND B. MOLDOVANU (2012): "Locally robust implementation and its limits," *Journal of Economic Theory*, 147(6), 2439–2452.

KAJII, A., AND S. MORRIS (1997): "The robustness of equilibria to incomplete information," *Econometrica*, pp. 1283–1309.

MCLEAN, R., AND A. POSTLEWAITE (2002): "Informational size and incentive compatibility," *Econometrica*, 70(6), 2421–2453.

MERTENS, J.-F., AND S. ZAMIR (1985): "Formulation of Bayesian analysis for games with incomplete information," *International Journal of Game Theory*, 14(1), 1–29.

MONDERER, D., AND D. SAMET (1989): "Approximating Common Knowledge with Common Beliefs," *Games and Economic Behavior*, 1, 170–190.

OLLÁR, M., AND A. PENTA (2017): "Full implementation and belief restrictions," *American Economic Review*, 107(8), 2243–77.

OURY, M. (2015): "Continuous implementation with local payoff uncertainty," *Journal of Economic Theory*, 159, 656–677.

OURY, M., AND O. TERCIEUX (2012): "Continuous implementation," *Econometrica*, 80(4), 1605–1637.

POSTLEWAITE, A., AND D. WETTSTEIN (1989): "Feasible and continuous implementation," *The Review of Economic Studies*, 56(4), 603–611.

RUBINSTEIN, A. (1989): "The Electronic Mail Game: Strategic Behavior Under" Almost Common Knowledge"," *American Economic Review*, 79(3), 385–391.

SATTERTHWAITE, M. A., AND H. SONNENSCHEIN (1981): "Strategy-proof allocation mechanisms at differentiable points," *The Review of Economic Studies*, 48(4), 587–597.

TAKAHASHI, S., AND O. TERCIEUX (2011): "Robust Equilibria in Sequential Games under Almost Common Certainty of Payoffs," Discussion paper, mimeo.

WEINSTEIN, J., AND M. YILDIZ (2004): "Finite-order implications of any equilibrium," .

——— (2007): "A structure theorem for rationalizability with application to robust predictions of refinements," *Econometrica*, 75(2), 365–400.

——— (2011): "Sensitivity of equilibrium behavior to higher-order beliefs in nice games," *Games and Economic Behavior*, 72(1), 288–300.

APPENDIX A. OMITTED PROOFS

PROOF OF LEMMA 1. Suppose that f is continuously implementable w.r.t. d^{uw} by mechanism $\mathcal{M} = (M, g)$. Consider a model $\mathcal{T} = (T, \kappa)$ defined as follows. Let

$$T_j = \overline{T}_j \bigsqcup \bigsqcup_{\left(\theta', t_j, t'_{-j}\right) \in \overline{T}_j \times \Theta \times \overline{T}_{-j}} \bigsqcup_{n=1}^{\infty} \left\{ t_{j,n}^{\left(t_j, \theta', t'_{-j}\right)} \right\}.$$

where we set $\kappa_{t_j} = \overline{\kappa}_{t_j}$ for every $t_j \in \overline{T}_j$; moreover, let

$$\kappa_{t_{j,n}^{\left(t_j, \theta', t'_{-j}\right)}} = \left(1 - \frac{1}{n}\right) \overline{\kappa}_{t_j} + \frac{1}{n} \delta_{\left(\theta', t'_{-j}\right)}, \forall n \in \mathbb{N}.$$

It is straightforward to verify that $d_j^{\text{uw}} \left(t_{j,n}^{\left(t_j, \theta', t'_{-j}\right)}, t_j \right) \to 0$. Suppose instead that for some agent i, and some pair of types t_i and t'_i in $\overline{T}_i$, g neither strictly rewards truth-telling at t_i over t'_i nor does t_i always weakly dominate t'_i in g, i.e.,

$$\sum_{(\theta, t_{-i}) \in \Theta \times \overline{T}_{-i}} \left[u_i \left(g(t'_i, t_{-i}), \theta \right) - u_i \left(g(t_i, t_{-i}), \theta \right) \right] \overline{\kappa}_{t_i} \left[(\theta, t_{-i}) \right] \geq 0 \tag{1}$$

and for some $t'_{-i} \in \overline{T}_{-i}$ and θ',

$$u_i\left(g(t'_i, t'_{-i}, \theta'\right) - u_i\left(g(t_i, t'_{-i}, \theta'\right) > 0. \tag{2}$$

Hence, under σ_{-i}, by reporting t'_i, agent i with type $t_{i,n}^{(t_i,\theta',t'_{-i})}$ gets interim expected payoff equal to

$$\left(1-\frac{1}{n}\right)\sum_{(\theta,t_{-i})\in\Theta\times\overline{T}_{-i}} u_i\left(g(t'_i, t_{-i}), \theta\right)\,\overline{\kappa}_{t_i}\left[(\theta, t_{-i})\right] + \frac{1}{n}u_i\left(g(t'_i, t'_{-i}), \theta'\right).$$

Then, by (1) and (2), for agent i with this type, reporting t'_i is strictly better than reporting t_i for all n large enough. But then it cannot be the case that $\sigma_i\left(t_{i,n}^{(t_i,\theta',t'_{-i})}\right) \to \delta_{t_i}$. This contradicts the supposition that $\mathcal{M}$ continuously implements f. ■

PROOF OF THEOREM 1. ($\Leftarrow$): Let $\mathcal{T}$ be a model with $\mathcal{T} \supset \overline{\mathcal{T}}$. Since T is countable and $\overline{T}$ is finite, a standard fixed-point argument implies that there is a BNE σ in the game $U(g, \mathcal{T})$. Let $\tilde{\sigma}$ be the strategy profile in $\tilde{g}$ induced from σ, i.e., for each $t \in T$, we set $\tilde{\sigma}(t)[\tilde{t}] = \sigma(t)[\tilde{t}]$ where $\tilde{t}$ is the set of messages strategically equivalent to t in the DRM g. Since σ is a BNE in g, it follows that $\tilde{\sigma}$ is also a BNE in $\tilde{g}$.

Since $R^{\infty}(t, \tilde{g}) = \{\tilde{t}\}$ for $t \in \overline{T}$, by the upper hemicontinuity of the rationalizable correspondence $R^{\infty}(\cdot, \tilde{g})$ (see, e.g., Theorem 2 of Dekel, Fudenberg, and Morris (2006)), there is some $\varepsilon > 0$ such that

$$d_i^{\mathrm{P}}\left(t'_i, t_i\right) < \varepsilon \Rightarrow R_i^{\infty}\left(t'_i, \tilde{g}\right) = \{\tilde{t}_i\}$$

Since $\tilde{\sigma}$ is a BNE in $\tilde{g}$, it follows that $\tilde{\sigma}_i\left(t'_i\right) = \delta_{\tilde{t}_i}$ for any $t'_i \in T_i$ with $d_i^{\mathrm{P}}\left(t'_i, t_i\right) < \varepsilon$. Hence, for any $t'_i \in \operatorname{supp} \sigma_i(t_i)$, we have that t'_i is equivalent to $\tilde{t}_i$. Define a strategy profile σ' in $U(g, \mathcal{T})$ as

$$\sigma'_i\left(t'_i\right) \equiv \begin{cases} \delta_{\tilde{t}_i}, & \text{if } d_i^{\mathrm{P}}\left(t'_i, t_i\right) < \varepsilon; \\ \sigma_i\left(t'_i\right), & \text{otherwise.} \end{cases}$$

Since σ is a BNE in $U(g, \mathcal{T})$, that σ' is also a BNE. Moreover, $g(\tilde{t}) = f(t)$ for every $t \in \overline{T}$ and by construction σ' also satisfies requirement (b) in Definition 1.

($\Rightarrow$): Fix a DRM g that truthfully continuously implements f w.r.t d^{P}. Since f is truthfully continuously implementable by g w.r.t. d^{P}, f is truthfully continuously implementable by g w.r.t. d^{uw}. By Theorem 2 and Corollary 1, f is implementable in strict BNE in $\tilde{g}$.

The following lemma will be useful.

LEMMA 3. *For each $k \geq 1$ and $\varepsilon \in (0,1)$, there is a countable model $\mathcal{T}_{k,\varepsilon} \supset \overline{\mathcal{T}}$ such that $T_{i,0,\varepsilon} \equiv \overline{T}_i$ and $T_{i,k,\varepsilon} \equiv \left(\bigsqcup_{t_i \in \overline{T}_i} R_i^k(t_i, \tilde{g})\right) \bigsqcup T_{i,k-1,\varepsilon}$.*

Fix any BNE $\tilde{\sigma}$ of the the game $U(\tilde{g}, \mathcal{T}_{k,\varepsilon})$ with $\tilde{\sigma}(t) = \delta_{\tilde{t}}$ for every $t \in \overline{T}$. This model has the property that for each type $t_{i,k,\varepsilon}(\tilde{t}'_i, t_i)$ (the type in $T_{i,k,\varepsilon}$ that corresponds to $(\tilde{t}'_i, t_i)$ such that $\tilde{t}'_i \in R^k_i(t_i, \tilde{g})$),

(1) $d^k_i\left(t^k_{i,k,\varepsilon}(\tilde{t}'_i, t_i), t^k_i\right) < \varepsilon$;

(2) $\tilde{\sigma}_i\left(t_{i,k,\varepsilon}(\tilde{t}'_i, t_i)\right) = \delta_{\tilde{t}'_i}$.

This lemma appears a little convoluted but is at the heart of our proof. It constructs a countable model $\mathcal{T}_{k,\varepsilon}$ with following property: Consider any Bayes Nash equilibrium $\tilde{\sigma}$ of the game of incomplete information $U(\tilde{g}, \mathcal{T}_{k,\varepsilon})$ with the property that types in $\overline{T}$ all report their type " truthfully." In other words, each type t_i sends the reduced normal form message $\tilde{t}_i$ in $\tilde{g}$ corresponding to the equivalence class which the type t_i falls in. Further, consider any message $\tilde{t}'_i \in R^k_i(t_i, \tilde{g})$, i.e. any message that survives up to k rounds of iterated deletion of never best response in $\tilde{g}$ for type t_i of player i.

The model $\mathcal{T}_{k,\varepsilon}$ is constructed such that there exists a type of player i, $t_{i,k,\varepsilon}(\tilde{t}'_i, t_i)$ that is ε-close to t_i in their k-th-order beliefs; moreover, player i of type $t_{i,k,\varepsilon}(\tilde{t}'_i, t_i)$ must play $\tilde{t}'_i$ under the BNE $\tilde{\sigma}$.

Before we present the proof of Lemma 3, let us conclude the now routine proof of Theorem 1. Consider the countable model $\mathcal{T}$ where $T_i = \bigsqcup_{k=1}^{\infty} T_{i,k,\frac{1}{k}}$ and $\mathcal{T}_{k,\frac{1}{k}}$ is given as in Lemma 3.

Since f is truthfully continuously implementable w.r.t. d^{P}, there is a BNE σ in the game $U(g, \mathcal{T})$ such that requirements (a) and (b) in Definition 1 hold. Again, σ induces a BNE $\tilde{\sigma}$ in $\tilde{g}$. Since $\sigma(t) = \delta_{\tilde{t}}$ by requirement (b) of Definition 1, we have $\tilde{\sigma}(t) = \delta_{\tilde{t}}$.

Thus, it follows from Lemma 3 that for each $\tilde{t}'_i \in R^{\infty}_i(t_i, \tilde{g})$, for each k, there is a type $t_{i,k,\frac{1}{k}}(\tilde{t}'_i, t_i) \in T_i$ such that

$$d^k_i\left(t_{i,k,\frac{1}{k}}(\tilde{t}'_i, t_i), t^k_i\right) \leq \frac{1}{k}, \tag{3}$$

and

$$\tilde{\sigma}_i\left(t_{i,k,\frac{1}{k}}(\tilde{t}'_i, t_i)\right) = \delta_{\tilde{t}'_i}.$$

It follows from (3) that $d^{\mathrm{P}}_i\left(t_{i,k,\frac{1}{k}}(\tilde{t}'_i, t_i), t_i\right) \to 0$. Since σ satisfies requirement (b) in Definition 1, we know that it must be the case that $\sigma_i\left(t_{i,k,\frac{1}{k}}(\tilde{t}'_i, t_i)\right) \to \delta_{\tilde{t}_i}$. Hence, $\tilde{t}'_i = \tilde{t}_i$.

Finally, since $\tilde{t}'_i \in R^{\infty}_i(t_i, \tilde{g})$ is arbitrary, we conclude that $\tilde{t}_i$ is the unique rationalizable message profile at t in $\tilde{g}$. ■

PROOF OF LEMMA 3. Formally, fix $\varepsilon \in (0,1)$ and we prove the claim by induction. First, the claim trivially holds for $k = 0$. Now we prove the claim for $k \geq 1$, assuming that it holds for $k-1$. Denote by $\tilde{T}_i$ the messages of agent i in the reduced-form $\tilde{g}$, namely $\tilde{T}_i \equiv \{\tilde{t}_i : t_i \in \overline{T}_i\}$. By definition, each $\tilde{t}'_i \in R^k_i(t_i, \tilde{g})$ is a best response to some belief

$\mu_{-i} \in \Delta\left(\Theta \times \overline{T}_{-i} \times \tilde{T}_{-i}\right)$ such that:

$$\begin{aligned} &\operatorname{marg}_{\Theta \times \overline{T}_{-i}} \mu_{-i} = \kappa_{t_i}, \\ \text{and } &\mu_{-i}\left(\left\{\left(\theta, t_{-i}, \tilde{t}'_{-i}\right) : \tilde{t}'_{-i} \in R^{k-1}_{-i}\left(t_{-i}, \tilde{g}\right)\right\}\right) = 1. \end{aligned}$$

By the induction hypothesis, there is a mapping $\eta_{-i,k-1,\varepsilon}$ from each $t_{-i} \in \overline{T}_{-i}$ and $\tilde{t}'_{-i} \in R^{k-1}_{-i}\left(t_{-i}, \tilde{g}\right)$ to a type $t_{-i,k-1,\varepsilon}\left(\tilde{t}'_{-i}, t_{-i}\right)$ such that (1) and (2) in Lemma 3 hold.

Since $\tilde{t}'_i$ is in the reduced form $\tilde{g}$ of the DRM g, $\tilde{t}'_i$ is the equivalent class which includes some $t'_i \in \overline{T}_i$. Then, define $\kappa_{t_{i,k,\varepsilon}(\tilde{t}'_i, t_i)} \in \Delta\left(\Theta \times T_{-i,k,\varepsilon}\right)$

$$\kappa_{t_{i,k,\varepsilon}(\tilde{t}'_i, t_i)} = (1-\varepsilon)\left(\mu_{-i} \circ \eta^{-1}_{-i,k-1,\varepsilon}\right) + \varepsilon \overline{\kappa}_{t'_i}.$$

That is, with probability $(1-\varepsilon)$, type $t_{i,k,\varepsilon}\left(\tilde{t}'_i, t_i\right)$ believes that the state and the opponents' types follow a distribution that is induced from μ_{-i} (in which each $t_{-i,k-1,\varepsilon}\left(\tilde{t}'_{-i}, t_{-i}\right)$ plays $\tilde{\sigma}_{-i}\left(t_{-i,k-1,\varepsilon}\left(\tilde{t}'_{-i}, t_{-i}\right)\right) = \delta_{\tilde{t}'_{-i}}$ by the induction hypothesis); with probability ε, type $t_{i,k,\varepsilon}\left(\tilde{t}'_i, t_i\right)$ has the same belief as type t'_i. Since $\tilde{t}'_i$ is a best response against μ_{-i} and the strict/ unique best response against $\overline{\kappa}_{t'_i}$ in $\tilde{g}$ (by Corollary 1), it follows that $\tilde{\sigma}_i\left(t_{i,k,\varepsilon}\left(\tilde{t}'_i, t_i\right)\right) = \delta_{\tilde{t}'_i}$. Moreover, since

$$d^{k-1}_{-i}\left(t_{-i,k-1,\varepsilon}\left(\tilde{t}'_{-i}, t_{-i}\right), t^{k-1}_{-i}\right) < \varepsilon,$$

we have that $d^k_i\left(t^k_{i,k,\varepsilon}\left(\tilde{t}'_i, t_i\right), t^k_i\right) < \varepsilon$. ■

PROOF OF THEOREM 2. (⇒) Observe that when a DRM g strictly rewards truthtelling at t_i over t'_i, then t'_i cannot always weakly dominate t_i. Thus, it follows from Lemma 1 that if f is truthfully continuously implementable by a DRM g, then t'_i always weakly dominates t_i if and only if t_i always weakly dominates t'_i, i.e., they are strategically equivalent in the sense of Definition 3.

(⇐)Let g be a DRM that truthfully continuously implements f in the sense of Definition 1. Hence, $g(t) = f(t)$ for every $t \in \overline{T}$. Now consider a model $\mathcal{T} \supset \overline{\mathcal{T}}$.

Note that we can pick $\varepsilon > 0$ such that for each i, and t_i and t'_i such that g strictly rewards truthtelling, we have

$$(1-\varepsilon) \sum_{(\theta, t_{-i}) \in \Theta \times \overline{T}_{-i}} \left[u_i\left(g(t_i, t_{-i}), \theta\right) - u_i\left(g(t'_i, t_{-i}), \theta\right)\right] \overline{\kappa}_{t_i}\left[(\theta, t_{-i})\right] > \varepsilon D. \tag{4}$$

Here D is defined as

$$D \equiv \max_{i,t,t',\tilde{\theta}} \left|u_i\left(g(t), \tilde{\theta}\right) - u_i\left(g(t'), \tilde{\theta}\right)\right|.$$

Moreover, we may decrease ε further so that the following two conditions are satisfied: firstly, for any agent i and any t_i and t'_i, the $\left(d^{\text{uw}}_i, \varepsilon\right)$-ball around t_i does not overlap with the $\left(d_i^{\mathrm{uw}},\varepsilon\right)$-ball around t_i', i.e. these balls are disjoint; and secondly,

$$\begin{aligned} & d_i^{\mathrm{uw}}\left(t_i'',t_i\right)<\varepsilon \\ \Longrightarrow\ & \kappa_{t_i''}\left[\left\{(\theta,t_{-i})\right\}^{\varepsilon}\right]\in\left[(1-\varepsilon)\kappa_{t_i}\left[(\theta,t_i)\right],(1+\varepsilon)\kappa_{t_i}\left[(\theta,t_i)\right]\right] \end{aligned} \tag{5}$$

where $\{(\theta,t_{-i})\}^{\varepsilon}$ denotes the $\left(d_{-i}^{\mathrm{uw}},\varepsilon\right)$-ball around (θ,t_{-i}). In words, consider any type t_i'' which is ε-close to the baseline type t_i. For every (θ,t_{-i}) baseline considers possible with probability $\overline{\kappa}_{t_i}[(\theta,t_{-i}))]$, the type t_i'' puts a close by belief on the set $(\theta,t_{-i})^{\varepsilon}$ consisting of types $(d_{-i}^{\mathrm{uw}},\varepsilon)$ close to t_{-i}.

Consider the agent normal-form of the game $U(\mathcal{M},\mathcal{T})$ with the restriction that t_i'' in the $\left(d_i^{\mathrm{uw}},\varepsilon\right)$-ball around t_i must report t_i. Denote this game with restriction by $\overline{U}(\mathcal{M},\mathcal{T})$.

Since T is countable and $\overline{T}$ is finite, a standard fixed-point argument implies that $\overline{U}(\mathcal{M},\mathcal{T})$ has a BNE σ. By construction of $\overline{U}(\mathcal{M},\mathcal{T})$, for any sequence $d^{\mathrm{uw}}(t_n,t)\to 0$, we have $\sigma(t_n)=t$ for n large enough.

Furthermore, σ is a BNE in the original game $U(\mathcal{M},\mathcal{T})$. To see this note that for any agent i in the ε ball around t_i, given that all other agents $-i$ in the ε-ball around (θ,t_{-i}) are reporting t_{-i}, the unique best response is to play t_i. This follows due to (4) and (5).

Therefore, g truthfully continuously implements f with respect to d^{uw}. ■

PROOF OF LEMMA 2. Suppose that f is continuously implementable w.r.t. d^{uw} by mechanism $\mathcal{M}=(M,g)$. Consider a model $\mathcal{T}=(T,\kappa)$ defined as follows. Let

$$T_j=\overline{T}_j\bigsqcup\bigsqcup_{\left(\theta',t_j,t_{-j}'\right)\in\overline{T}_j\times\Theta\times\overline{T}_{-j}}\bigsqcup_{n=1}^{\infty}\left\{t_{j,n}^{\left(t_j,\theta',t_{-j}'\right)}\right\}.$$

where we set $\kappa_{t_j}=\overline{\kappa}_{t_j}$ for every $t_j\in\overline{T}_j$; moreover, let

$$\kappa_{t_{j,n}^{\left(t_j,\theta',t_{-j}'\right)}}=\left(1-\frac{1}{n}\right)\overline{\kappa}_{t_j}+\frac{1}{n}\delta_{\left(\theta',t_{-j}'\right)},\forall n\in\mathbb{N}.$$

It is straightforward to verify that $d_j^{\mathrm{uw}}\left(t_{j,n}^{\left(t_j,\theta',t_{-j}'\right)},t_j\right)\to 0$. Since f is continuously implementable w.r.t. d^{uw} by $\mathcal{M}$, there is an equilibrium σ which continuously implements f in $U(\mathcal{M},\mathcal{T})$. Since σ continuously implements f in $U(\mathcal{M},\mathcal{T})$, we have (a) $\sigma|_{\overline{T}}$ is a pure-strategy BNE in $U\left(\mathcal{M},\overline{\mathcal{T}}\right)$; (b) $g(\sigma(t_n))\to f(t)$ for any sequence of type profiles $\{t_n\}\subset T$ and $t\in\overline{T}$ with $d(t_n,t)\to 0$. Since $\overline{T}_0=\overline{T}$, it follows that

$$g(\sigma(t)) = f(t),\forall t\in\overline{T}. \tag{6}$$

$$g\left(\sigma_i\left(t_{i,n}^{\left(t_i,\theta',t_{-i}'\right)}\right),\sigma_{-i}(t_{-i})\right) \to f(t_i,t_{-i}),\forall t_{-i}\in\overline{T}_{-i} \tag{7}$$

Suppose to the contrary that for some agent i, the SCF f neither strictly rewards t_i over m_i'; nor does t_i always weakly dominate m_i' in σ (or more precisely $\overline{\sigma} \equiv \sigma|_{\overline{T}}$ in $U\left(\mathcal{M}, \overline{\mathcal{T}}\right)$), i.e.,

$$\sum_{(\theta,t_{-i})\in\Theta\times\overline{T}_{-i}} \left[u_i\left(g\left(m_i', \sigma_{-i}\left(t_{-i}\right)\right)\right), \theta\right) - u_i\left(g\left(\sigma_i\left(t_i\right), \sigma_{-i}\left(t_{-i}\right)\right)\right), \theta\right)\right] \overline{\kappa}_{t_i}\left[\left(\theta, t_{-i}\right)\right] \geq 0. \quad (8)$$

and for some t_{-i}' and θ',

$$u_i\left(g\left(m_i', \sigma_{-i}\left(t_{-i}\right)\right)\right), \theta\right) - u_i\left(g\left(\sigma_i\left(t_i\right), \sigma_{-i}\left(t_{-i}\right)\right)\right), \theta\right) > 0. \quad (9)$$

First, under σ_{-i}, by reporting m_i', agent i with type $t_{i,n}^{\left(t_i,\theta',t_{-i}'\right)}$ gets the interim expected payoff equal to

$$\begin{aligned}&\left(1-\frac{1}{n}\right)\sum_{(\theta,t_{-i})\in\Theta\times\overline{T}_{-i}} u_i\left(g(m_i', \sigma_{-i}\left(t_{-i}\right)), \theta\right) \overline{\kappa}_{t_i}\left[\left(\theta, t_{-i}\right)\right] \\ &+\frac{1}{n}u_i\left(g(m_i', \sigma_{-i}\left(t_{-i}'\right)), \theta'\right)\end{aligned}$$

where the equality follows from (6). Second, by (7) for each $t_{-i} \in \overline{T}_{-i}$, there is some $M_i^{t_{-i}} \subset M_i$ such that for any sufficiently large n,

$$\begin{aligned}\sigma_i\left(t_{i,n}^{\left(t_i,\theta',t_{-i}'\right)}\right)\left[M_i^{t_{-i}}\right] &\geq 1-\frac{1}{2\left|\overline{T}_{-i}\right|}; \\ g\left(m_i, \sigma_{-i}\left(t_{-i}\right)\right) &= f\left(t_i, t_{-i}\right) = g\left(\overline{\sigma}_i\left(t_i\right), \overline{\sigma}_{-i}\left(t_{-i}\right)\right)), \forall m_i \in M_i^{t_{-i}}. \quad (10)\end{aligned}$$

Since $\overline{T}_{-i}$ is finite, it follows that for sufficiently large n, we have

$$\sigma_i\left(t_{i,n}^{\left(t_i,\theta',t_{-i}'\right)}\right)\left[\bigcap_{t_{-i}\in\overline{T}_{-i}} M_i^{t_{-i}}\right] \geq \sum_{t_{-i}\in\overline{T}_{-i}} \sigma_i\left(t_{i,n}^{\left(t_i,\theta',t_{-i}'\right)}\right)\left[M_i^{t_{-i}}\right] - \left(\left|\overline{T}_{-i}\right| - 1\right) > 0 \quad (11)$$

Finally, under σ_{-i}, by reporting $m_i \in \bigcap_{t_{-i}\in\overline{T}_{-i}} M_i^{t_{-i}}$, agent i of type $t_{i,n}^{\left(t_i,\theta',t_{-i}'\right)}$ gets the interim expected payoff equal to

$$\begin{aligned}&\left(1-\frac{1}{n}\right)\sum_{(\theta,t_{-i})\in\Theta\times\overline{T}_{-i}} u_i\left(g(m_i, \sigma_{-i}\left(t_{-i}\right)), \theta\right) \overline{\kappa}_{t_i}\left[\left(\theta, t_{-i}\right)\right] \\ &+\frac{1}{n}u_i\left(g(m_i, \sigma_{-i}\left(t_{-i}'\right)), \theta'\right) \\ =\ &\left(1-\frac{1}{n}\right)\sum_{(\theta,t_{-i})\in\Theta\times\overline{T}_{-i}} u_i\left(g\left(\overline{\sigma}_i\left(t_i\right), \overline{\sigma}_{-i}\left(t_{-i}\right)\right)\right), \theta\right) \overline{\kappa}_{t_i}\left[\left(\theta, t_{-i}\right)\right] + \frac{1}{n}u_i\left(f\left(t_i, t_{-i}'\right)\right), \theta'\right)\end{aligned}$$

where the equality follows from (10). Then, by (8), (9) and (11), agent i of type $t_{i,n}^{(t_i,\theta',t'_{-i})}$ can profitably deviate by assigning the probability on $\bigcap_{t_{-i}\in\overline{T}_{-i}} M_i^{t_{-i}}$ to m'_i instead. This is a contradiction to σ being a BNE. ■

PROOF OF THEOREM 4. Suppose that $\mathcal{M}$ continuously implements f w.r.t. d^{P}. To prove that $\tilde{\mathcal{M}}$ strictly continuously implements f w.r.t. d^{P}, consider any model $\mathcal{T}' = (T', \kappa')$. Denote by $\mathcal{T}'' = (T'', \kappa'')$ the disjoint union of $\mathcal{T}' = (T', \kappa')$ and the model $\mathcal{T} = (T, \kappa)$ constructed in Lemma 2. Then, we must have some BNE σ which continuously implements f in $U(\mathcal{M}, \mathcal{T}'')$ (and there by in $U(\mathcal{M}, \mathcal{T})$). It follows from Lemma 2 that $\sigma|_{\overline{T}}$ satisfies the property that for each agent i, each type t_i in $\overline{T}_i$ and message $m'_i \in M_i$, we have either $\sigma|_{\overline{T}}$ strictly rewards t_i over m'_i; or t_i always weakly dominates m'_i in BNE $\sigma|_{\overline{T}}$. Since $\overline{\mathcal{T}}$ has full support, if t_i always weakly dominates m'_i in $\sigma|_{\overline{T}}$ and $\sigma_i|_{\overline{T}}(t_i)$ is not strategically equivalent to m'_i, the message m'_i must yield strictly lower payoff than $\sigma_i|_{\overline{T}}(t_i)$ for type t_i. Hence, it follows from Assumption 3 that $\sigma|_{\overline{T}}$ is a strict BNE in $U\left(\tilde{\mathcal{M}}, \overline{\mathcal{T}}\right)$. It follows that σ continuously implements f in $U(\mathcal{M}, \mathcal{T}')$. Hence, $\tilde{\mathcal{M}}$ strictly continuously implements f w.r.t. d^{P}. ■

PROOF OF THEOREM 5. Since $\mathcal{M} = (M, g)$ continuously implements f w.r.t. d^{P}, by Theorem 4, we may assume without loss of generality that $\mathcal{M}$ strictly continuously implements f w.r.t. d^{P}. We start by proving the following key lemma.

LEMMA 4. *For each pure-strategy strict BNE $\overline{\sigma}$ in $U\left(\mathcal{M}, \overline{\mathcal{T}}\right)$ and $k \geq 0$, there is a model $\mathcal{T}_k^{\overline{\sigma}} \supset \overline{\mathcal{T}}$ such that $T_{i,0} \equiv \overline{T}_i$ and*

$$T_{i,k}^{\overline{\sigma}} \equiv \left(\bigsqcup_{t_i\in\overline{T}_i} W_i^k(t_i, \mathcal{M})\right) \bigsqcup T_{i,k-1}^{\overline{\sigma}}.$$

Fix any BNE σ of the the game $U\left(\mathcal{M}, \mathcal{T}_k^{\overline{\sigma}}\right)$ such that $\sigma|_{\overline{T}} = \overline{\sigma}$. This model has the property that for each type $t_{i,k}(m_i, t_i)$ (the type in $T_{i,k}$ that corresponds to $m_i \in W_i^k(t_i, \mathcal{M})$),
(1) $t_{i,k}^k(m_i, t_i) = t_i^k$;
(2) $\sigma_i(t_{i,k}(m_i, t_i)) = \delta_{m_i}$.

Consider any message $m_i \in W_i^k(t_i, \mathcal{M})$, i.e. any message that survives up to k rounds of iterated deletion of never best response in $\mathcal{M}$ for type t_i of player i. The model $\mathcal{T}_k^{\overline{\sigma}}$ is constructed such that there exists a type of player i, $t_{i,k}(m_i, t_i)$ that has the same k-th-order beliefs; moreover, player i of type $t_{i,k}(m_i, t_i)$ must play m_i under the BNE σ.

Before we present the proof of Lemma 4, let us conclude the now routine proof of Theorem 5. Consider the countable model $\mathcal{T}$ where

$$T_i = \bigsqcup_{\overline{\sigma} \text{ is a pure-strategy strict BNE in } U\left(\mathcal{M},\overline{\mathcal{T}}\right)} \left(\bigsqcup_{k=0}^{\infty} T_{i,k}^{\overline{\sigma}} \right)$$

and $T_{i,k}^{\overline{\sigma}}$ is given as in Lemma 4. Since $\mathcal{M}$ strictly continuously implements f w.r.t. d^{P}, there is some BNE σ which strictly continuously implements f in $U\left(\mathcal{M},\mathcal{T}\right)$. Hence, $\sigma|_{\overline{T}}$ is a pure-strategy strict BNE in $U\left(\mathcal{M},\overline{\mathcal{T}}\right)$. It follows from Lemma 4 that for each k and each $m_i \in W_i^k(t_i,\mathcal{M})$, there is a type $t_{i,k}^k\left(m_i,t_i\right) \in T_i$ such that

$$t_{i,k}^k\left(m_i,t_i\right) = t_i^k \tag{12}$$

and

$$\sigma_i\left(t_{i,k}\left(m_i,t_i\right)\right) = \delta_{m_i}.$$

It follows from (12) that $d_i^{\mathrm{P}}\left(t_{i,k}\left(m_i,t_i\right),t_i\right) \to 0$. Since $\mathcal{M}$ strictly continuously implements f, we know that it must be the case that $g\left(\sigma\left(t_k\left(m,t\right)\right)\right) \to f\left(t\right)$. Since $\sigma\left(t_k\left(m,t\right)\right) = m$, it follows that $g\left(m\right) = f\left(t\right)$ for every $m \in W^\infty\left(t,\mathcal{M}\right)$. ■

Proof of Lemma 4. First, since $\sigma|_{\overline{T}} = \overline{\sigma}$, the claim trivially holds for $k=0$. Now we prove the claim for $k \geq 1$, assuming that it holds for $k-1$. By definition, each $m_i \in W_i^k\left(t_i,\mathcal{M}\right)$ is a strict best response to some belief $\mu_{-i} \in \Delta\left(\Theta \times \overline{T}_{-i} \times M_{-i}\right)$ such that $\operatorname{marg}_{\Theta\times\overline{T}_{-i}} \mu_{-i} = \kappa_{t_i}$ and $\mu_{-i}\left(\left\{\left(\theta,t_{-i},m_{-i}\right) : m_{-i} \in W_{-i}^{k-1}\left(t_{-i},\mathcal{M}\right)\right\}\right) = 1$. By the induction hypothesis, there is a mapping $\eta_{-i,k-1}$ from each $t_{-i} \in \overline{T}_{-i}$ and $m_{-i} \in W_{-i}^{k-1}\left(t_{-i},\mathcal{M}\right)$ to a type $t_{-i,k-1}\left(m_{-i},t_{-i}\right)$ such that (1) and (2) in Lemma 4 holds. Define $\kappa_{t_{i,k}(m_i,t_i)} \in \Delta\left(\Theta \times T_{-i,k}^{\overline{\sigma}}\right)$ as

$$\kappa_{t_{i,k}(m_i,t_i)} = \mu_{-i} \circ \eta_{-i,k-1}^{-1}.$$

That is, type $t_{i,k}\left(m_i,t_i\right)$ believes that the state and the opponents' types follow a distribution that is induced from μ_{-i} (in which each $t_{-i,k-1}\left(m_{-i},t_{-i}\right)$ plays m_{-i} in BNE σ by the induction hypothesis). Since m_i is a best response against μ_{-i}, it follows that $\sigma_i\left(t_{i,k}\left(m_i,t_i\right)\right) = \delta_{m_i}$. Moreover, since $t_{-i,k-1}^{k-1}\left(m_{-i},t_{-i}\right) = t_{-i}^{k-1}$, we have that $t_{i,k}^k\left(m_i,t_i\right) = t_i^k$. ■