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 his 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. If only ordinal preferences are common knowledge among agents, a mild richness condition implies that the social choice function must be dictatorial. Truthful continuous implementation is thus impossible without non-trivial knowledge of agents' higher order beliefs. We further show that without such knowledge, a revelation principle does not apply: the set of social choice functions which can be continuously implemented is strictly larger.

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
JANUARY 13, 2020
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 implementation setting: there are finite sets of agents, states and alternatives. The planner would like to (partially) implement a given social choice function, i.e. a function from possible states to alternatives. The state is unknown to the principal. As a baseline, suppose the state is common knowledge among agents. In this case, it is well known that any social choice function is trivially partially implementable given three or more agents with a direct revelation mechanism. But what if the principal is not sure whether the state is *exactly* common knowledge among agents, but would nevertheless like the social choice function to be partially implemented "close to" complete information? 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 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.

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

Intuitively, the characterization depends on the underlying topology with respect to which we demand continuity. As motivation for the topologies we consider, consider the example of a standard government natural resource auction setting. In the baseline, all agents in the auction rely on, and know that other agents also rely on etc., the same set of extremely accurate geological surveys. If this were exactly the case, not only do agents know that all others agree with their estimates, but also know that other agents know that others agree and so on—the estimate is common knowledge among agents. Further, if the government understands that this is the situation, then it knows that the estimates are common knowledge among agents.

First consider the variant that with some small probability, some of the agents might not have received some of these surveys. Alternately, consider the variant that with some small probability, agents might receive an additional, accurate survey that only they privately see. For a small enough probability, both these settings may be thought of as "close" to common knowledge. While these two may appear superficially similar, it is well known that they are different in terms of the higher-order beliefs they induce in the agents.[2] Further, in either variant, the principal may not know the exact information structure among agents, and may therefore wish any mechanism she designs to be robust with respect to small changes in the information structure.

We wish to understand how the desideratum of robustness with respect to these two kinds of settings constrains the principal. We focus on these specifically for two reasons. First, conceptually, one can argue that these capture two disparate ways an information structure can be close to complete information: the former involves agreement at all arbitrarily higher-order beliefs, while the latter only constrains lower-order beliefs. Second, at a more technical level, our results in the former are a building block for our results in the latter—we detail this further below in Section 1.1.

At a high level, our findings can be summarized thus: Settings like the former , i.e. where despite not knowing the exact information structure, the principal nevertheless has some information about the agents' higher-order beliefs, are not much more constraining than the baseline of *exact* common knowledge. By contrast, if the agents' higher-order beliefs may be arbitrary, then the principal is severely restricted—to the point where in some settings the only implementable social choice functions are those that are dictatorial.

Further, we show that a "revelation principle" applies for the weaker notion. That is to say one in which the principal knows agents' higher-order beliefs i.e. that if a social

[2]More precisely suppose the true state is either H or L, and the various surveys accurately reveal the state with very high probability. The former situation we described corresponds to one where it is common knowledge that there are k surveys, and each agent sees, for each survey, either the result of the survey or with some small probability a null signal φ. The latter situation corresponds to one where the number of surveys is not common knowledge, and having seen k surveys, agents put some probability that other agents have seen $k+1$, setting up a belief structure akin to Rubinstein (1989).

choice function can be continuously implemented, it can be robustly 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. Nevertheless we are able to provide a partial characterization of continuous implementation in this setting, and thus explain the gap between continuous implementation and truthful continuous implementation.

1.1. *Model and Results*

Let us now be slightly more formal in describing the setting and our results. There are finite sets of agents, states and alternatives. There is given a social choice function of interest.[3] There is a baseline information structure that the principal considers: we start with the special case of common knowledge of the state in Section 3, and generalize in Section 4. 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 that yields the desired social choice function at all types close to the baseline types. We term this requirement *truthful continuous implementation* (the additional modifier of "truthful" to the notion of Oury and Tercieux (2012) reflecting our restriction to this limited class of mechanisms).

We formalize the examples discussed previously by considering continuity with respect to two topologies on types. The first, 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, roughly, a social choice function is truthfully continuously implementable if and only if it can be implemented in Strict equilibrium in the baseline model (Theorems 2, 6).

The second, the product topology, used in Weinstein and Yildiz (2007) places no restrictions on agents' higher-order beliefs. We show that under this topology, roughly, 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 (Theorems 3, 7).

To argue that this is a very demanding solution concept, we introduce a modeling innovation not previously considered in the robust implementation literature: we revisit the ordinal setting considered in the early implementation literature. In the baseline common

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

knowledge state we only assume agents know each others' ordinal preferences over alternatives, not the cardinal utilities.[4] We characterize truthful continuous implementation with respect to the product topology in this setting: it requires that the social choice function be uniquely ordinal rationalizable implementable in the original common knowledge model (Theorem 5).[5] The results of Börgers (1995) then imply that as long as the set of possible states includes all unanimous profiles over the alternatives, the social choice function must be dictatorial (Corollary 7).

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 8). 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.[6]

At a technical level, a key novelty that we would like to highlight is 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.[7]

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 topology, we know that the social choice function must

[4]In the context of normal-form games of complete information, there has been a view that fixing the cardinal/ von-Neumann Morgenstern utilities is a strong assumption. The works of Börgers (1993) or Weinstein (2016) relax this, and develop counterparts to standard solution concepts (the former) or study how the outcomes under solution concepts is affected by different cardinal utilities that represent the same ordinal preferences (the latter).

[5]Rationalizability here involves considering all cardinalizations that are consistent with the ordinal preferences: see Definition 8 and the discussion that follows for details.

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

[7]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 result, Theorem 3. See also Weinstein and Yildiz (2011).

be implementable in Strict Nash Equilibrium. Our contagion begins from the putative equilibrium where an agent with a type that believes a state is common knowledge sends the message corresponding to that state. Recall further that we are considering implementation with DRMs, i.e. for every message an agent could send there is a corresponding state: in other words, the equilibrium has full range. Strict NE implies that in the state where an agent believes that a state is common knowledge 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 at complete information at some state. We construct a sequence of types that converge to the complete information 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 that the state corresponding to the strategy s_i is indeed the true state and everyone is playing that. 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. We defer a fuller verbal description to after we introduce the formal result

The organization of the rest of the paper is as follows. Section 2 defines the model. Section 3 characterizes truthful continuous implementation in the special case where the baseline information structure the principal considers is that agents have common-knowledge of the state of the world. Section 3.4 introduces the ordinal model, and characterizes truthful continuous implementation in this model. Section 4 generalizes to settings where the baseline information structure may be one of incomplete information. Section 5 studies general continuous implementation (i.e., as opposed to truthful continuous implementation). Section 6 discusses the related literature and connections.

2. Model

There is a state of the world $\theta \in \Theta$, unknown to the planner. There is a set of alternatives A. The planner would like to implement a social choice function $f : \Theta \to A$. We assume that both A and Θ are finite (some of our results do not require this assumption).

There is a finite set of I agents. Agent i has a utility function $u_i : A \times \Theta \to \Re$. 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 his indifferences, and $\succeq_{i,\theta}$ for weak preference.

2.1. *Common Prior Perturbations*

To build intuition for our results, we first consider the following simple common prior setting. Agent i receives a signal $s_i \in S_i \equiv \Theta$, $S \equiv \prod_i S_i$.

Signals are drawn according to a joint prior probability distribution over signals and states of the world, $P \in \Delta(\Theta \times S)$. Viewing P as a point in the $(I+1)|\Theta|$ dimensional Euclidean space, we say a model P is ε-close to P' if $||P - P'||_1 \leq \varepsilon$.

For each $s_i \in S_i$, denote by $P(\cdot|s_i)$ the conditional probability on $\Theta \times S_{-i}$. Denote the signal of agent i by s_i^θ when agent i receives a signal corresponding to state of the world θ; moreover, we write s^θ for the signal profile $\left(s_i^\theta\right)_{i \in I}$ and s_{-i}^θ for $\left(s_j^\theta\right)_{j \neq i}$. Say P^{CI} is a complete information prior if $P^{\text{CI}}(\theta, s) = 0$ for every $s \neq s^\theta$. Fix some P^{CI} which assigns positive probability on each $\theta \in \Theta$.[8]

A mechanism, denoted (M, g) is a message space M_i for each player i, and an outcome function $g : M \to A$. A Bayes-Nash Equilibrium (BNE) is a strategy profile $(\sigma_i)_{i \in I}$ with $\sigma_i : S_i \to \Delta(M_i)$ such that for $s_i \in S_i$, each message $m_i \in \text{supp}\, \sigma_i(s_i)$ maximizes the expected payoff of agent i with respect to the opponents' strategy profile σ_{-i} and $P(\cdot|s_i)$.

As is standard, we say that a game form (M, g) (partially) implements a SCF f at model P if there is a mixed BNE of the game form $\sigma_i : S_i \to \Delta(M_i)$ such that $g(m) = f(s)$ for every message profile $m \in \textit{supp}\, \sigma(s)$ for all signal profiles $s \in S$. We will consider "direct revelation mechanisms" (DRMs), i.e. a game form where $M = S$.

The following definition is closely related to Definitions 2,3 of Oury and Tercieux (2012), restricting attention to "direct revelation mechanisms," i.e. where the message space equals the set of possible states of the world Θ; and adapted to the model of perturbations considered above.

Definition 1. *We say f is truthfully continuously implementable if there is a DRM g such that for any sequence of models $P_n \to P^{CI}$:*

(a) For each $\theta \in \Theta$, $g\left(s^\theta\right) = f(\theta)$;

(b) There exists $\underline{n}$ large enough such that truth-telling (σ^T) is a BNE of g under P_n for any $n \geq \underline{n}$.[9]

2.2. *Universal type space and topologies*

We now introduce the machinery to study continuous implementation under general perturbations. 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

[8]This is to ensure that if $P_n \to P^{\text{CI}}$, then $P_n(\cdot|s_i)$ is defined by Bayes' rule for large n. The latter property is used in Theorem 1. The choice of P^{CI} does not affect our definitions or results.

[9]The truth-telling strategy profile σ^T is simply the one where $\sigma_i^T(\theta) = \delta_{s_i^\theta}$ for all θ and i, where δ is the standard Dirac-delta function.

mechanism $\mathcal{M}$ and a model $\mathcal{T}$, we write $U(\mathcal{M}, T)$ for the induced incomplete information game. Let $\overline{\mathcal{T}} = (\overline{T}, \overline{\kappa})$ be the complete-information model, i.e., $\overline{T}_i = \{t_i^\theta : \theta \in \Theta\}$ and $\overline{\kappa}_{t_i^\theta}\left[\left(\theta, t_{-i}^\theta\right)\right] = 1$ for each $\theta \in \Theta$.

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

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

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 $(\Delta(\Theta))$.[10] 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:

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

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.[11] 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 following adapts Definition 1 to now consider continuity in these topologies defined above, and uses the definition above to define perturbations to a baseline typespace

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

$$\inf\{\gamma > 0 : \mu'(E) \leq \mu(E^\gamma) + \gamma \text{ for every Borel set } A \subseteq X\},$$
$$\text{where } E^\gamma = \{x \in X : \inf_{y \in E} \rho(x, y) < \gamma\}.$$

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

(as opposed to perturbations to an ex-ante prior over types in Definition 1)—Theorem 2 below connects the two notions. Note that we still restrict attention to DRMs as before, i.e. the message space of each player still equals $S_i \equiv \Theta$.

DEFINITION 2. *We say f is truthfully continuously implementable w.r.t. a metric d if there is a DRM g such that for any model $\mathcal{T} \supset \overline{\mathcal{T}}$, there is a (possibly mixed) BNE σ in the game $U(\mathcal{M}, \mathcal{T})$ with the property that for any sequence of type profiles $\{t_n\} \subset T$ with $d\left(t_n, t^\theta\right) \to 0$, for every $\theta \in \Theta$ we have:*

(a) $g(s^\theta) = f(\theta)$, and,

(b) $\sigma(t_n) = \delta_{s^\theta}$ for any n large enough.

Definition 2 is directly comparable to Definition 2 of Oury and Tercieux (2012). Note that truthful continuous implementation is more demanding on the surface than continuous implementation in two ways. Firstly, it fixes the form of the mechanism used. Secondly, it demands robustness of a specific equilibrium of this mechanism (i.e., the truth-telling equilibrium).[12]

Our requirement (b) superficially strengthens the requirement in their paper, with the difference arising from the fact that we only have a finite number of alternatives. They only require that, at a sequence of types t_n converging to a common-knowledge of $\theta \in \Theta$ profile t^θ, the outcome of g under the Bayes-Nash equilibrium σ considered converges to the desired outcome under the social choice function , $f(\theta)$; i.e. whenever $\{t_n\}$ converges to t (in either topology), $(g \circ \sigma)(t_n) \to f(\theta)$. By contrast, we require that f is the exact outcome for all type profiles "close enough." The difference can be explained by observing that we require these nearby types tell the truth and hence $(g \circ \sigma)(t_n) \to f(\theta)$ is equivalent to $g(s^\theta) = f(\theta)$. In contrast, truth-telling in general has no meaning with an arbitrary equilibrium in the indirect mechanisms considered in Oury and Tercieux (2012).

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. 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 s_i is* strategically equivalent *to s'_i for an agent i if agent i is indifferent between the two reports regardless of the state and others' reports, i.e.:*

$$\forall s_{-i}, \theta'' : g(s_i, s_{-i}) \sim_{i,\theta''} g(s'_i, s_{-i}).$$

[12]Looking at outcome convergence without focusing on the truth-telling equilibrium is the exercise of Takahashi and Tercieux (2011). We discuss this further in Section 6.

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 s_i, let $\tilde{s}_i$ denote the message in $\tilde{g}$ corresponding to the set of messages strategically equivalent to s_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 s_i and s'_i which are strategically equivalent, $g(s_i, \cdot) = g(s'_i, \cdot)$.*

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.

In our setting, this assumption is non-standard because it may be satisfied by some combination of restrictions on the environment, the social choice function f to be implemented, and the object of study, i.e. the mechanism g, itself. To alleviate this, observe that the following simple richness assumption purely on the environment implies that Assumption 1 is always satisfied.

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 value $v_i \in [0, 1]$ for receiving the good, and an outside option of 0 for not receiving the good, with $\theta = (v_1, v_2, v_3)$, $\Theta = [0, 1] \times [0, 1] \times [0, 1]$. 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, $f(v_1, v_2, v_3) =$

$\arg\max_i(v_1, v_2, v_3)$ is such that any DRM g that implements it must satisfy Assumption 1—an agent's report will sometimes affect her own allocation.

In what follows, we invoke the weaker/ necessary Assumption 1. The reader may mentally substitute the stronger/ sufficient Assumption 2 if they prefer.

3. Complete Information Baseline Model

In this section, we describe our results for the baseline information structure where the state of the world is common knowledge among agents. First, Section 3.1 considers robustness in the sense of Section 2.1, i.e. with respect to perturbations of a common prior. This helps build intuition for our main results: characterizations of truthful continuous implementation with respect to the uniform-weak and product topologies (Section 3.2 and 3.3 respectively). Section 3.4 introduces the ordinal model (i.e. ones where the baseline common knowledge types only have common knowledge of agents' ordinal preferences) and characterizes truthful continuous implementation in this setting.

3.1. *Common Prior Perturbations*

Before we can state our characterization, we need to introduce two more terms. We say that DRM g *strictly rewards unanimity* at θ over θ' for agent i if

$$g(s_i^\theta, s_{-i}^\theta) \succ_{i,\theta} g(s_i^{\theta'}, s_{-i}^\theta).$$

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

$$\forall s_{-i}, \theta'' : \quad g(s_i^\theta, s_{-i}) \succeq_{i,\theta''} g(s_i^{\theta'}, s_{-i}).$$

The following simple lemma is key to our characterization.

Lemma 1. *If an SCF f is truthfully continuously implementable by a DRM g in the sense of Definition 1 then, for every agent i and any pair θ and θ', either g strictly rewards unanimity at θ over θ'; or θ always weakly dominates θ' in g.*

Proof. Suppose instead that for some agent i, and some pair θ and θ', g neither strictly rewards unanimity at θ over θ' nor does θ always weakly dominate θ' in g, i.e.:

$$u_i\left(g(s_i^{\theta'}, s_{-i}^\theta), \theta\right) \geq u_i\left(g(s_i^\theta, s_{-i}^\theta), \theta\right) \tag{1}$$

and for some $\bar{s}_{-i}$ and θ'',

$$u_i\left(g(s_i^{\theta'}, \bar{s}_{-i}), \theta''\right) > u_i\left(g(s_i^\theta, \bar{s}_{-i}), \theta''\right). \tag{2}$$

Consider $P_n \in \Delta(\Theta \times S)$ such that

$$P_n\left(\tilde{\theta},\tilde{s}\right) = \begin{cases} \left(1-\frac{1}{n}\right) P^{\text{CI}}\left(\theta, s^{\theta}\right), & \text{if } \left(\tilde{\theta},\tilde{s}\right) = \left(\theta, s^{\theta}\right); \\ \frac{1}{n} P^{\text{CI}}\left(\theta, s^{\theta}\right), & \text{if } \left(\tilde{\theta},\tilde{s}\right) = \left(\theta'', s_i^{\theta}, \bar{s}_{-i}\right); \\ P^{\text{CI}}\left(\tilde{\theta},\tilde{s}\right), & \text{otherwise.} \end{cases}$$

Clearly, $P_n \to P^{\text{CI}}$. Thus, under σ^T_{-i} and P_n, by reporting s_i, agent i who has received a signal of s_i^{θ} gets the interim expected payoff equal to

$$\left(1-\frac{1}{n}\right) u_i\left(g\left(s_i, s_{-i}^{\theta}\right), \theta\right) + \frac{1}{n} u_i\left(g\left(s_i, \bar{s}_{-i}\right), \theta''\right).$$

Then, by (1) and (2), for agent i with signal s_i^{θ}, reporting $s_i^{\theta'}$ is strictly better than truth-telling, and thus truth-telling is not a BNE of g under P_n for every n, a contradiction. ■

Our main characterization follows:

Theorem 1. *An SCF f is truthfully continuously implementable by a DRM g in the sense of Definition 1 if and only if the following hold:*

(a) $g(s^{\theta}) = f(\theta)$ for each $\theta \in \Theta$,

(b) For every agent i and any pair θ and θ', either g strictly rewards unanimity at θ over θ'; or s_i^{θ} is strategically equivalent to $s_i^{\theta'}$ for agent i.

Proof. ($\Rightarrow$) If a DRM g strictly rewards unanimity at θ over θ', then θ' cannot always weakly dominate θ. It follows from Lemma 1 that if f is truthfully continuously implementable by a DRM g, then θ' always weakly dominates θ if and only if θ always weakly dominates θ', i.e., s_i^{θ} and $s_i^{\theta'}$ are strategically equivalent in the sense of Definition 3.

($\Leftarrow$): Consider a sequence of models $\{P_n\}$ with $P_n \to P^{\text{CI}}$. We want to show that σ^T is a BNE of g under P_n for any sufficiently large n. Pick $\varepsilon > 0$ such that for each θ and θ' such that whenever g strictly rewards unanimity at θ over θ', we have

$$(1-\varepsilon)\left[u_i\left(g(s_i^{\theta}, s_{-i}^{\theta}), \theta\right) - u_i\left(g(s_i^{\theta'}, s_{-i}^{\theta}), \theta\right)\right] > \varepsilon D \tag{3}$$

$$\text{where, } D \equiv \max_{i,s,s',\tilde{\theta}} \left|u_i\left(g\left(s\right), \tilde{\theta}\right) - u_i\left(g\left(s'\right), \tilde{\theta}\right)\right|. \tag{4}$$

Then, if s_i' is strategically equivalent to s_i^{θ}, agent i with signal s_i^{θ} gets the same payoff announcing either s_i^{θ} or s_i'. Moreover, for any sufficiently large n, each s_i^{θ} assigns at least probability $1-\varepsilon$ on $\left(\theta, s_{-i}^{\theta}\right)$. It follows from (3) that s_i^{θ} is strictly better than s_i' for agent i with signal s_i^{θ}, if s_i' is not strategically equivalent to s_i^{θ}. ■

How permissive are the conditions identified in Theorem 1? After all, these are conditions stated on the game form g, rather than its implications on what social choice functions are continuously implementable in the sense of Definition 1.

DEFINITION 5. *Let g be a DRM which admits a reduced normal-form. We say f is implementable in strict NE in the reduced normal-form $\tilde{g}$ if for every $\theta \in \Theta$,*
(a) $\tilde{g}(\tilde{s}^\theta) = f(\theta)$;
(b) $\tilde{s}^\theta$ is a strict NE in $\tilde{g}$ at state θ for every $\theta \in \Theta$.

We obtain the following corollary:

COROLLARY 1. *Suppose that Assumption 1 holds. An SCF f is truthfully continuously implementable in DRM g if and only if it is implementable in strict NE in $\tilde{g}$.*

PROOF. ($\Rightarrow$) By Assumption 1 and the fact that f is truthfully continuously implementable, $\tilde{g}(\tilde{s}^\theta) = g(s^\theta) = f(\theta)$. Moreover, it follows from Theorem 1 that $\tilde{s}^\theta$ is a strict NE at θ in $\tilde{g}$.

($\Leftarrow$) Since $\tilde{g}$ satisfies requirements (a) and (b) in Definition 5, g satisfies the requirements of Theorem 1, which implies that f is truthfully continuously implementable. ■

This is the main finding of this section—f must be implementable in Strict Nash Equilibrium in the original common knowledge environment. The requirement of implementation in Strict Nash Equilibrium is known to be a weak one—the following corollaries characterize it. However, we should also point out that slightly outside the model, if we allowed for the possibility of arbitrarily small transfers, and with three or more agents, implementation in Strict Nash Equilibrium is trivially possible. This can be done by defining the outcome of the game form when only one agent's report differs from the consensus as the same as at consensus, but requiring that agent to make a small payment.

Implementation in Strict Nash Equilibrium requires deviators to be strictly punished. In the absence of payments, this of course requires that there exist alternatives strictly worse for a potential deviator than what she could achieve if she truthfully reported the state. The following concept is therefore helpful: we say f is *never pessimal* for agent i if, at any state, the alternative selected $f(\theta)$ is not his worst alternative, i.e.:

$$\forall \theta : \; f(\theta) \notin \arg\min_{a \in A} u_i(a, \theta).$$

COROLLARY 2. *If a DRM g truthfully continuously implements f and Assumption 1 holds, g is invariant to the reports of any agent for whom f is not never pessimal.*

PROOF. Consider an agent i for whom f is pessimal at some state θ. By Corollary 1, θ must be strategically equivalent to any other θ' for this agent. By Assumption 1 strategically equivalent messages must lead to the same alternative being selected. Therefore, g must always ignore i's reports. ■

As long as there are three or more agents for whom the social choice function f is never pessimal, implementation in Strict Nash Equilibrium and therefore truthful continuous implementation is "for free," as summarized by the following corollary.

COROLLARY 3. *An SCF f is truthfully continuously implementable if there are three or more agents for whom f is never pessimal.*

3.2. *Uniform-Weak Topology*

We now consider truthful continuous implementation with respect to the uniform-weak topology d^{uw} on the universal type space. The following result shows that truthful continuous implementation (w.r.t. the common-prior perturbations, as in Definition 1) is equivalent to truthful continuous implementation with respect to d^{uw} (Definition 2).

In doing so, we demonstrate that the driving force behind Theorem 1 is the fact that θ is "almost common knowledge" among agents (recall that the uniform-weak topology preserves approximate common knowledge). The common prior ex-ante stage makes the proofs of Theorem 1 easier but does not play a pivotal role.

The proof is fairly straightforward. In one direction, common prior perturbations are a special case of close-by types in the uniform weak topology. The other direction makes use of Theorem 1. In this direction, the proof essentially argues that if g truthfully continuously implements f in the sense of Definition 1, then truth telling is a BNE among types close enough under d^{uw} to the common knowledge types. By Theorem 1, truth-telling is a *strict* Nash Equilibrium under g (modulo strategically equivalent messages). Thus, like Monderer and Samet (1989), regardless of what other types are playing, as long as other agents' types close to the common knowledge type are truth-telling, it follows that truth-telling is a unique best response for this agent.

THEOREM 2. *An SCF f is truthfully continuously implementable if and only if it is truthfully continuously implementable with respect to d^{uw}.*

PROOF. ($\Leftarrow$): Let P_n be a sequence of common prior models such that $P_n \to P^{CI}$. Consider a model $\mathcal{T} = (T, \kappa)$ defined as follows: Let $T_i = \bigsqcup_{n=1}^{\infty} S_{i,n}$ where each $S_{i,n} \equiv S_i$; moreover, for each $s_{i,n} \in S_{i,n}$, let $\kappa_{s_{i,n}} = P_n(\cdot|s_{i,n})$. Hence, $\kappa_{s_{i,n}}[\Theta \times S_{-i,n}] = 1$.

Since f is truthfully continuously implementable with respect to d^{uw}, let σ be the BNE that satisfies (a) and (b) in Definition 2.

To see that f is truthfully continuously implementable, consider any $\theta \in \Theta$ and $s_{i,n} = \theta$. Since P^{CI} has full support and $P_n \to P^{CI}$, it follows from Fudenberg and Tirole (1991, Theorem 14.5) that for any $p < 1$, it is common-p at any s_n that θ has realized for every sufficiently large n. Hence, $d^{\text{uw}}\left(s_n, t^{\theta}\right) \to 0$ (see Chen, Di Tillio, Faingold, and Xiong (2010)). Since S is finite, it follows from condition (b) in Definition 2 that $\sigma(s_n) = \delta_{s^{\theta}}$ for any n large enough. That is, σ^T is a BNE under P_n.

($\Rightarrow$): Let g be a DRM that truthfully continuously implements f in the sense of Definition 1. Hence, $g\left(s^{\theta}\right) = f(\theta)$ for every θ. Now consider a model $\mathcal{T} \supset \overline{\mathcal{T}}$. By Theorem 1, we can pick $\varepsilon > 0$ such that for each θ and θ' such that g strictly rewards unanimity at θ over θ', we have, for D defined as in (4),

$$(1-\varepsilon)\left[u_i\left(g(s_i^\theta, s_{-i}^\theta), \theta\right) - u_i\left(g(s_i^{\theta'}, s_{-i}^\theta), \theta\right)\right] > \varepsilon D. \tag{5}$$

Moreover, we may decrease ε further so that the following two conditions are satisfied: (1) for any agent i and any $\theta \neq \theta'$, the $\left(d_i^{\text{uw}}, \varepsilon\right)$-ball around $\left(\theta, t_i^\theta\right)$ does not overlap with the $\left(d_i^{\text{uw}}, \varepsilon\right)$-ball around $\left(\theta', t_i^{\theta'}\right)$, i.e. these balls are disjoint; (2)

$$d_i^{\text{uw}}\left(t_i, t_i^\theta\right) < \varepsilon \implies \kappa_{t_i}\left[\left\{\left(\theta, t_{-i}^\theta\right)\right\}^\varepsilon\right] > 1 - \varepsilon, \tag{6}$$

where $\left\{\left(\theta, t_{-i}^\theta\right)\right\}^\varepsilon$ denotes the $\left(d_{-i}^{\text{uw}}, \varepsilon\right)$-ball around $\left(\theta, t_{-i}^\theta\right)$, i.e., any type which is ε-close to the common knowledge type t_i^θ also believes that with probability at least $1-\varepsilon$ all other agents $-i$ have types within distance ε from t_{-i}^θ.

Consider the agent normal-form of the game $U(\mathcal{M}, \mathcal{T})$ with the restriction that t_i in the $\left(d_i^{\text{uw}}, \varepsilon\right)$-ball around t_i^θ must report s_i^θ. Denote this game with restriction by $\overline{U}(\mathcal{M}, \mathcal{T})$. Since T is countable and S 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 $d^{\text{uw}}\left(t_n, t^\theta\right) \to 0$, we have $\sigma(t_n) = s^\theta$.

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^θ, given that all other agents $-i$ in the ball (θ, t_{-i}^θ) are reporting s_{-i}^θ, the unique best response is to play s_i^θ. This follows due to (5) and (6).

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

As a result of Theorem 2 the characterization of social choice functions which are truthfully continuously implementable with respect to the uniform-weak topology is exactly the same as that under the common prior perturbations outlined in Section 3.1. The same interpretations therefore apply: continuous implementation in this topology is, as we would argue, permissive.

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 (here, the common knowledge type).

3.3. *Product topology*

Finally, we consider truthful continuous implementation in the product topology. The following definition of interim correlated rationalizable messages (c.f. Dekel, Fudenberg, and Morris (2007)) will be useful:

DEFINITION 6. *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: $m_i \in \arg\max_{m_i'} \int_{\Theta \times M_{-i}} u_i(m_i', m_{-i}, \theta)$ marg $\mu_{\Theta \times M_{-i}}[d\theta, m_{-i}]$;

R2: 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 7. *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 $\theta \in \Theta$,*
(a) $\tilde{g}(\tilde{s}^\theta) = f(\theta)$;
(b) $R^\infty(t^\theta, \tilde{g}) = \{\tilde{s}^\theta\}$.

Note that part (2) of the definition has t^θ as the first argument. In other words, our definition requires $\tilde{s}^\theta$ to be the (unique) rationalizable action profile when θ is common knowledge, for every state $\theta \in \Theta$. Since agents' vN-M utilities over alternatives are common knowledge, interim correlated rationalizability reduces to the standard (correlated) rationalizability of Bernheim (1984) or Pearce (1984).

Further, as we pointed out earlier, note that $\tilde{s}_i^\theta$ is agent i's *unique* rationalizable action. In this sense, our requirement is slightly stronger than the usual (full) implementation in rationalizability (see e.g. Bergemann, Morris, and Tercieux (2011) for a characterization of implementation in rationalizability). The latter only requires that the social choice function be implemented at every rationalizable action profile; we additionally require that the rationalizable action profile be unique.

Our main result characterizes truthful continuous implementation w.r.t. d^{p} as equivalent to implementability in rationalizability. As we stated earlier, the result is similar to the main result of Oury and Tercieux (2012)—we also use the contagion argument of Weinstein and Yildiz (2007) in a mechanism design context. However unlike the former paper, we do not need to extend the model to consider costly messages etc. to break ties.

THEOREM 3. *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 the unique rationalizable action profile in $\tilde{g}$ in the sense of Definition 7.*

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 7.

As a building block we have Theorem 2, which combined with Theorem 1 and Corollary 1 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 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 already know that f is truthfully continuously implementable (in the product topology) then f is implementable in Strict Nash Equilibrium.

Recall further that we are considering implementation with DRMs, i.e. for every message an agent could send there is a corresponding state: in other words, the equilibrium has full range. Strict NE implies that in the state where an agent believes that a state θ is common knowledge 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).

Take any rationalizable strategy s_i for a player i at complete information at some state. We can construct a sequence of types that converge to the complete information 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 that the state corresponding to the strategy s_i is indeed the true state and everyone is playing that. The latter playing this strategy makes s_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 strategy s_i. Since the social choice function is continuously implementable, therefore, we have rationalizable implementation as desired.

PROOF OF THEOREM 3. ($\Leftarrow$): Let $\mathcal{T}$ be a model with $\mathcal{T} \supset \overline{\mathcal{T}}$. Since T is countable and S 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 $t \in T$ and each $\tilde{s}$, $\tilde{\sigma}(t)[\tilde{s}] = \sigma(t)[\tilde{s}]$, where in the latter $\tilde{s}$ is identified with the set of equivalent messages in the DRM g. Since σ is a BNE in g, it follows that $\tilde{\sigma}$ is also a BNE in $\tilde{g}$.

Since $R^\infty(t^\theta, \tilde{g}) = \{\tilde{s}^\theta\}$, 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

$$d_i^{\mathrm{p}}\left(t_i, t_i^\theta\right) < \varepsilon \Rightarrow R_i^\infty(t_i, \tilde{g}) = \left\{\tilde{s}^\theta\right\}$$

Since $\tilde{\sigma}$ is a BNE in $\tilde{g}$, it follows that $\tilde{\sigma}_i(t_i) = \delta_{\tilde{s}_i^\theta}$ for any $t_i \in T_i$ with $d_i^{\mathrm{p}}\left(t_i, t_i^\theta\right) < \varepsilon$. Hence, for any $s_i \in \operatorname{supp} \sigma_i(t_i)$, we have $\tilde{s}_i = \tilde{s}_i^\theta$.

To concludes the proof, define a strategy profile $\overline{\sigma}$ in $U(g, \mathcal{T})$ as

$$\overline{\sigma}_i(t_i) \equiv \begin{cases} \delta_{s_i^\theta}, & \text{if } d_i^{\mathrm{p}}\left(t_i, t_i^\theta\right) < \varepsilon; \\ \sigma_i(t_i), & \text{otherwise.} \end{cases}$$

Since σ is a BNE in $U(g, \mathcal{T})$, it follows from the fact that $\tilde{s}_i = \tilde{s}_i^\theta$ for any $s_i \in \operatorname{supp} \sigma_i(t_i)$ that $\overline{\sigma}$ is also a BNE.[13] Moreover, $g\left(s^\theta\right) = f(\theta)$ and by construction $\overline{\sigma}$ also satisfies requirement (b) in Definition 2.

($\Rightarrow$): 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 NE in $\tilde{g}$. Now consider the following Lemma:

Lemma 2. *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 $T_{i,k,\varepsilon} \equiv \left(\bigsqcup_{\theta \in \Theta} R_i^k\left(t_i^\theta, \tilde{g}\right)\right) \bigsqcup T_{i,k-1,\varepsilon}$.*

Fix any BNE $\tilde{\sigma}$ of the the game $U(\tilde{g}, \mathcal{T}_{k,\varepsilon})$ with $\tilde{\sigma}\left(t^\theta\right) = \delta_{\tilde{s}^\theta}$ for every θ. This model has the property that for each type $t_{i,k,\varepsilon}(\tilde{s}_i, \theta)$ (the type in $T_{i,k,\varepsilon}$ that corresponds to $\tilde{s}_i \in R_i^k\left(t_i^\theta, \tilde{g}\right)$),

(1) $d_i^k\left(t_{i,k,\varepsilon}^k(\tilde{s}_i, \theta), \left(t_i^\theta\right)^k\right) < \varepsilon$;
(2) $\tilde{\sigma}_i\left(t_{i,k,\varepsilon}(\tilde{s}_i, \theta)\right) = \delta_{\tilde{s}_i}$.

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 common knowledge types all report the state " truthfully." In other words, each type t_i^θ sends the reduced normal form message $\tilde{s}_i^\theta$ in $\tilde{g}$ corresponding to the equivalence class which the state θ falls in. Further, consider any message $\tilde{s}_i \in R_i^k(t_i^\theta, \tilde{g})$, i.e. any message that survives up to k rounds of iterated deletion of never best response in $\tilde{g}$ under common knowledge of θ for any player i. The model $\mathcal{T}_{k,\varepsilon}$ is such that there exists a type of player i, $t_{i,k,\varepsilon}(\tilde{s}_i, \theta)$ that is ε-close to the common knowledge of θ type in their k-th-order beliefs, t_i^θ; moreover, player i of type $t_{i,k,\varepsilon}(\tilde{s}_i, \theta)$ must play $\tilde{s}_i$ under the BNE $\tilde{\sigma}$.

Before we present the proof of Lemma 2, let us conclude the now routine proof of Theorem 3. Consider the countable model $\mathcal{T}$ where $T_i = \bigsqcup_{k=1}^\infty T_{i,k,\frac{1}{k}}$ as in Lemma 2.

[13]We should clarify that Assumption 1 invoked in the Theorem is used in this step to ensure that messages in the original game from the same equivalent class in the reduced form game result in the same outcome. This is important to ensure that we do not affect incentives when we go from the reduced form game $\tilde{g}$ to the original game g.

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 2 hold. Again, σ induces a BNE $\tilde{\sigma}$ in $\tilde{g}$. Since $\sigma\left(t^{\theta}\right)=\delta_{s^{\theta}}$ by requirement (b) of Definition 2, we have $\tilde{\sigma}\left(t^{\theta}\right)=\delta_{\tilde{s}^{\theta}}$.

Thus, it follows from Lemma 2 that for each $\tilde{s}_i \in R_i^{\infty}\left(t_i^{\theta},\tilde{g}\right)$, for each k, there is a type $t_{i,k,\frac{1}{k}}\left(\tilde{s}_i,\theta\right)\in T_i$ such that

$$d_i^k\left(t^k_{i,k,\frac{1}{k}}\left(\tilde{s}_i,\theta\right),\left(t_i^{\theta}\right)^k\right)\leq\frac{1}{k}, \tag{7}$$

$$\text{and } \tilde{\sigma}_i\left(t_{i,k,\frac{1}{k}}\left(\tilde{s}_i,\theta\right)\right)=\delta_{\tilde{s}_i}.$$

It follows from (7) that $d_i^{\mathrm{P}}\left(t_{i,k,\frac{1}{k}}\left(\tilde{s}_i,\theta\right),t_i^{\theta}\right)\to 0$. Since σ satisfies requirement (a) in Definition 2, we know that it must be the case that $\sigma_i\left(t_{i,k,\frac{1}{k}}\left(\tilde{s}_i,\theta\right)\right)=\delta_{s_i^{\theta}}$ for any k large enough. Hence, $\tilde{s}_i=\tilde{s}_i^{\theta}$. Finally, since $\tilde{s}_i\in R_i^{\infty}\left(t_i^{\theta},\tilde{g}\right)$ is arbitrary, we conclude that $\tilde{s}^{\theta}$ is the unique rationalizable message profile at θ in $\tilde{g}$. ■

It remains to provide a proof of Lemma 2. The argument resembles the proof of Proposition 1 in Weinstein and Yildiz (2007), i.e. a contagion argument. In their construction, the contagion is from a class of "crazy" types, for whom a given action is dominant. Here, by contrast we piggyback from the fact that if a social choice function is truthfully continuously implementable w.r.t. d^p, then it must be continuously implementable w.r.t. d^{uw}. By Theorem 2, the social choice function must therefore be implementable in Strict Nash Equilibrium in the reduced normal form game $\tilde{g}$.

For any rationalizable action $\tilde{s}_i$ in $R_i^{\infty}(t_i^{\theta},\tilde{g})$, we can construct a sequence of close-by types for which this action is the unique possible BNE strategy. We break ties by adding a small probability that it is common knowledge that the state of the world is the state corresponding to $\tilde{s}_i$ (in which case playing $\tilde{s}_i$ is a strict best response given the conjectured BNE). As is routine in these contagion arguments, this small probability can be inductively moved into higher and higher orders of belief, achieving convergence to the common knowledge type. Since g truthfully continuously implements f, by condition (1) of the definition of truthful continuous implementation (Definition 2), it must be the case that this rationalizable action is indeed $\tilde{s}_i^{\theta}$.

A closer parallel may be to Weinstein and Yildiz (2004) (see also Weinstein and Yildiz (2011)) where a particular equilibrium is fixed, and the question is what equilibrium actions are consistent with types whose first k orders of belief are known (and higher orders are unrestricted). Under the assumption that the equilibrium has full range, they show that the set of possible equilibrium actions must include the set of all actions that survive k rounds of iterated elimination of never strict best replies, and is upper-bounded by the set of actions that are k level rationalizable. Our DRM game and the putative truthful equilibrium at the common knowledge types trivially has full range, while Theorem 1 and Corollary 1 imply that in the reduced normal form, truth telling is a Strict Nash Equilibrium for common-knowledge types that has full range. In this case, their upper and lower bounds are the same for types that are k levels consistent with common knowledge types. Our requirement of implementation in unique rationalizable action as necessary can thus be seen to follow as a consequence.

PROOF OF LEMMA 2. 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 $k-1$. By definition, each $\tilde{s}_i \in R_i^k\left(t_i^\theta, \tilde{g}\right)$ is a best response to some belief $\lambda_{-i} \in \Delta\left(R_{-i}^{k-1}\left(t_{-i}^\theta, \tilde{g}\right)\right)$. By the induction hypothesis, there is a mapping $\eta_{-i} : R_{-i}^{k-1}\left(t_{-i}^\theta, \tilde{g}\right) \to T_{-i,k-1,\varepsilon}$ such that

$$\eta_{-i,k-1,\varepsilon}\left(\tilde{s}_{-i}\right) = t_{-i,k-1,\varepsilon}\left(\tilde{s}_{-i}, \theta\right).$$

Then, define $\kappa_{t_{i,k,\varepsilon}(\tilde{s}_i)} \in \Delta\left(\Theta \times T_{-i,k,\varepsilon}\right)$ as

$$\kappa_{t_{i,k,\varepsilon}(\tilde{s}_i,\theta)} = (1-\varepsilon)\left(\delta_\theta \times \left(\lambda_{-i} \circ \eta^{-1}_{-i,k-1,\varepsilon}\right)\right) + \varepsilon\delta_{\left(s_i, t^{s_i}_{-i}\right)}.$$

That is, with probability $(1-\varepsilon)$, type $t_{i,k,\varepsilon}\left(\tilde{s}_i, \theta\right)$ believes that the state is θ and the opponents' types follow a distribution that is induced from λ_{-i} (in which each $t_{-i,k-1,\varepsilon}\left(\tilde{s}_{-i}, \theta\right)$ plays $\tilde{\sigma}_{-i}\left(t_{-i,k-1,\varepsilon}\left(\tilde{s}_{-i}, \theta\right)\right) = \delta_{\tilde{s}_{-i}}$ by the induction hypothesis); with probability ε, type $t_{i,k,\varepsilon}\left(\tilde{s}_i, \theta\right)$ believes that the state is some s_i from the equivalent class $\tilde{s}_i$ and that the opponents' type profile $t^{s_i}_{-i}$ has common belief about the state being s_i (and thereby plays $\tilde{\sigma}_{-i}\left(t^{s_i}_{-i}\right) = \tilde{s}^{s_i}_{-i}$). Since $\tilde{s}_i$ is a best response against λ_{-i} and the strict/ unique best response against $\left(s_i, \tilde{s}^{s_i}_{-i}\right)$ in $\tilde{g}$, it follows that $\tilde{\sigma}_i\left(t_{i,k,\varepsilon}\left(\tilde{s}_i, \theta\right)\right) = \delta_{\tilde{s}_i}$. Moreover, since

$$d_i^{k-1}\left(t^{k-1}_{-i,k-1,\varepsilon}\left(\tilde{s}_{-i}, \theta\right), \left(t^\theta_{-i}\right)^{k-1}\right) < \varepsilon,$$

it follows that $d_i^k\left(t^k_{i,k,\varepsilon}\left(\tilde{s}_i, \theta\right), \left(t^\theta_i\right)^k\right) < \varepsilon$. ∎

As in Section 3.1, our results so far characterizing truthful continuous implementation in the product topology are in terms of g or its reduced normal form $\tilde{g}$. We can now derive implications on how restrictive this notion is on the social choice function f.

Our first result has a close counterpart in Theorem 1 of Oury and Tercieux (2012). They do not assume any richness compared to Assumption 1 imposed here. However, they strengthen the requirement to strict continuous implementation, and show that any strictly continuously implementable f is strictly Maskin monotonic. A social choice function $f : \Theta \to A$ is strictly Maskin monotonic if, for every pair of states θ and θ':

(1) We have that $f(\theta) = f(\theta')$ whenever

$$\forall i, a : u_i(f(\theta), \theta) > u_i(a, \theta) \implies u_i(f(\theta), \theta') \geq u_i(a, \theta'),$$

(2) Or, equivalently, $f(\theta) \neq f(\theta')$ implies

$$\exists i, a : u_i(f(\theta), \theta) > u_i(a, \theta) \text{ and } u_i(a, \theta') > u_i(f(\theta), \theta').$$

COROLLARY 4. *Suppose that Assumption 2 holds. If an SCF f is truthfully continuously implementable w.r.t. d^p, then f is strictly Maskin monotonic.*

PROOF. This follows from Theorem 3 and Proposition 1 of Bergemann, Morris, and Tercieux (2011). ■

COROLLARY 5. *Suppose that $|A| \geq 3$. If f is onto, has full domain (i.e. the set of states is such that every ordinal preference profile over alternatives is possible), and is truthfully continuously implementable w.r.t. d^p, then f is dictatorial.*

PROOF. This follows from Corollary 4 and the Muller-Satterthwaite Theorem (Muller and Satterthwaite (1977)) which states that any monotonic social function on a full domain of preferences and is onto with at least three alternatives must be dictatorial. ■

COROLLARY 6. *Suppose that Assumption 2 holds and $|A| = 2$. If f is onto and truthfully continuously implementable w.r.t. d^p, then f is dictatorial.*

PROOF. This follows from Theorem 3 and Theorem 3 of Xiong (2017). ■

3.4. *An Ordinal Setting*

Given our results so far, it remains unclear whether implementation in the unique rationalizable action profile is in general substantially different from implementation in dominant strategies. Corollary 5 makes a connection—if the full domain assumption is satisfied, then the social choice function must be dictatorial, which indeed is dominant strategy implementable as well.

However, full domain is a strong assumption, and it is already known that just monotonicity is very taxing under rich domains (for example, Saijo (1987) shows that in a universal domain, only constant social choice functions are monotonic). To make further headway toward answering this question without appealing to the full force of rich domains, we consider an ordinal setting: our common knowledge states will now be ones where agents commonly know each others' ordinal preferences over alternatives, but not each others' von Neumann-Morgenstern utilities.

It can be argued that such a setting is of independent interest, and in the spirit of the exercise this paper is undertaking. After all it may be dissonant to consider a robustness exercise, especially in a non-transferable utility setting, where nevertheless agents are sure of each others' cardinal utilities over alternatives. Our results in this setting thus delineate the extent to which the common knowledge of cardinal utilities in the baseline model was driving the results, and indeed verifies that they do not play a major role. We also note that this is in keeping with the original implementation literature which only assumed common knowledge of ordinal preferences rather than cardinal utilities.

At each $\theta \in \Theta$, each agent i has a preference ordering over the set of alternatives A. An SCF is a mapping $f : \Theta \to A$ and hence the social goal only depends on agents' ordinal preference profiles.

We still define $S_i = \Theta$ and a DRM as a mapping $g : S \to \Delta A$. In other words, a DRM only asks for the agents' reports on their ordinal preferences.

We still need cardinal utilities when we consider information perturbations. We model these by extending the state space to a cardinal state space defined as follows. Consider the Euclidean space $[0,1]^{|A|}$, and further consider the subset

$$U \equiv \{(r_1, r_2, \dots r_{|A|}) \in [0,1]^{|A|} : r_1 > r_2 > \dots > r_{|A|}\}.$$

We define the cardinal state space

$$\Theta^* = \left\{ \left(\theta, (u_i)_{i \in I}\right) \in \Theta \times U^I \right\}.$$

Here the ordered vector u_i associated with agent i is to be thought of as the cardinal utilities associated with his top alternative(s), second best alternative(s) etc. We will purposefully overload notation and also refer to the agent i's cardinal utility as a function $u_i : A \times \theta \to [0,1]$ for ease of notation. Given a $u_i \in U$, there is a unique implied utility function $u_i : A \times \theta \to [0,1]$.[14]

Endow U^I with the Euclidean topology and Θ^* with the product topology and endow both with the Borel σ-algebra. A model is now a pair (T, κ) where $T = T_1 \times T_2 \times \cdots \times T_I$ is a countable type space and $\kappa_{t_i} \in \Delta\left(\Theta^* \times T_{-i}\right)$ denotes the associated beliefs for each $t_i \in T_i$. Say (T, κ) is an *ordinal-complete information model* if $T_i = \cup_{\theta \in \Theta} T_i^\theta$ such that $\kappa_{t_i^\theta}\left[\{\theta\} \times U^I \times T_{-i}^\theta\right] = 1$ for every $t_i^\theta \in T_i^\theta$, $\theta \in \Theta$, and $i \in I$. In other words, each $t_i^\theta \in T_i^\theta$ believes that the ordinal preference profile is given by θ and his opponents' type profile belongs to T_{-i}^θ. Again, denote by $t_i^\theta \in T_i^\theta$ a typical element in an ordinal-complete information type space. Fix a *finite* ordinal-complete information model $\overline{\mathcal{T}} = \left(\overline{T}, \overline{\kappa}\right)$.

Each type in a model still induces a hierarchy of beliefs over Θ. Thus, both $d^{\text{uw}}\left(t_n, t^\theta\right)$ and $d^{\text{p}}\left(t_n, t^\theta\right)$ remain well defined and measure only distance of two hierarchies of beliefs about ordinal preferences. Truthful continuous implementation w.r.t. d^{uw} and truthful continuous implementation w.r.t. d^{p} are defined as previously.

[14]Of course, if the ordinal preferences of agent i have indifferences at θ, then there may be multiple $u_i \in U$ that represent the same utility function.

THEOREM 4. *An SCF f is truthfully continuously implementable in DRM g with respect to d^{uw} if and only if the following hold:*
(a) $g(s^\theta) = f(\theta)$ for each $\theta \in \Theta$,
(b) For every agent i and any pair θ and θ', either g strictly rewards unanimity at θ over θ'; or s_i^θ is strategically equivalent to $s_i^{\theta'}$.

The proof of this Theorem is very similar to the proof of Theorem 2. In the forward direction, we can fix a particular cardinalization and only consider perturbations on that, thus recovering the old cardinal model. In the backward direction, the argument is, as before, carefully arguing that if the condition is satisfied, it is a BNE for types close enough to the ordinal-complete information types to report the state, as desired. All the previous corollaries of Section 3.1 then follow immediately.

The duplication is because implementation in Strict Nash Equilibrium is an ordinal notion, and the particular cardinalization or lack thereof does not affect the result. However, the characterization of Theorem 3 in the product topology involved an inherently cardinal solution concept, namely, rationalizability. To develop the analog in this ordinal model, an additional definition is required.

DEFINITION 8. *Let $R_i^\infty(\theta, \mathcal{M})$ denote* ordinal rationalizable messages *of mechanism $\mathcal{M} = (M, g)$. Let $R_i^0(\theta, \mathcal{M}) = M_i$. Inductively, for each $k \geq 1$, a message $m_i \in R_i^k(\theta, \mathcal{M})$ iff there is some $\pi \in \Delta\left(U \times R_{-i}^{k-1}(\theta, \mathcal{M})\right)$ such that*

$$m_i \in \arg\max_{m_i'} \int_{U \times R_{-i}^{k-1}(\theta, \mathcal{M})} u_i\left(g(m_i', m_{-i}), \theta\right) \pi\left[du_i, m_{-i}\right].$$

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

Note that we allow the rationalizing belief to have correlations between the messages of others and agent i's utility index. This is therefore more permissive than the iterated pure strategy dominance notion of Börgers (1993) where the two must be independent: formal definitions are in the proof of Corollary 7.

DEFINITION 9. *Suppose a DRM g admits a reduced normal-form. We say that f is implementable in the unique ordinal rationalizable message in the reduced normal-form $\tilde{g}$ of DRM g if for every $\theta \in \Theta$,*
(a) $\tilde{g}(\tilde{s}^\theta) = f(\theta)$;
(b) $R^\infty(\theta, \tilde{g}) = \{\tilde{s}^\theta\}$.

THEOREM 5. *Supose 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 the unique ordinal rationalizable message in $\tilde{g}$ in the sense of Definition 9.*

At the core of the proof of Theorem 5 is a contagion argument that is very similar to the proof of Theorem 3. There are some technical difficulties mostly in ensuring we topologize the type space appropriately to get an analog of the upper hemicontinuity of the rationalizable correspondence for the backward direction.[15] However, the intuition remains the same and hence we defer the proof to the appendix. The payoff of Theorem 5 is in the following Corollary, which leverages a result of Börgers (1995) to characterize implementation in ordinal rationalizable messages in this setting.

First, we present two definitions: we say that θ generates a unanimous preference profile if all agents have the same preference ordering on A. Further, we say that Θ includes all unanimous preference profiles if every strict preference ordering over A can be induced as a unanimous preference profile generated by some $\theta \in \Theta$. In contrast to Corollary 5, the following result requires a substantially weaker domain richness condition and requires neither $|A| \geq 3$ nor f be onto.

COROLLARY 7. *Assume that Θ includes all unanimous preference profiles. If f is truthfully continuously implementable w.r.t. d^p, then f is dictatorial.*

PROOF. Since all unanimous preference profiles are possible, in particular, Assumption 2 is satisfied (and therefore Assumption 1).

Suppose that f is truthfully continuously implementable. Then, by Theorem 5, f is implementable in ordinal rationalizable messages in the sense of Definition 9, i.e., in $R^{\infty}(\theta, \tilde{g})$. Börgers (1993) defines a notion he calls *iterated pure strategy dominance*: Let $B_i^0(\theta, \mathcal{M}) = M_i$. Inductively, for each $k \geq 1$, a message $m_i \in B_i^k(\theta, \mathcal{M})$ iff there is some $\pi \in \Delta\left(B_{-i}^{k-1}(\theta, \mathcal{M})\right)$ and $u_i \in U$ such that

$$m_i \in \arg\max_{m_i'} \sum_{m_{-i}} u_i\left(m_i', m_{-i}\right) \pi\left[m_{-i}\right].$$

Then, $B_i^{\infty}(\theta, \mathcal{M}) \equiv \cap_{k=1}^{\infty} B_i^k(\theta, \mathcal{M})$. Note that our notion of ordinal rationalizability is more permissive than Börgers (1993) in that we allow agents to rationalize a message using a (possibly correlated) belief over their own cardinal indices and and their opponents' messages. Since $R^{\infty}(\theta, \tilde{g}) \supset B^{\infty}(\theta, \tilde{g})$, implementation in $R^{\infty}(\theta, \tilde{g})$ implies implementation $B^{\infty}(\theta, \tilde{g})$. By Proposition (unnumbered in original manuscript) and footnote 6 in Börgers (1995), under the maintained assumptions on Θ, f is implementable in $B^{\infty}(\theta, \tilde{g})$ only if f is dictatorial. ■

[15] The authors thank Satoru Takahashi for parts of the argument in this direction.

4. Incomplete Information Baseline Model

In this section we extend our results to settings where the baseline model is one of incomplete information. To that end, we update some of the definitions and introduce some notation to accomodate this richer setting.

The principal considers a baseline model which we denote by $\overline{\mathcal{T}} = \left(\overline{T}, \overline{\kappa}\right)$. We assume that the baseline model is finite, i.e., $\left|\overline{T}\right| < \infty$. 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 $\text{marg}_{\Theta_i}\mu[\theta_i] > 0$ for each θ_i, i.e., $\kappa_{t_i}[(\theta_i, \theta_{-i}, t_{-i})] = 1_{\{\theta_i = t_i, \theta_{-i} = t_{-i}\}}\mu(\theta_{-i}|\theta_i)$.

A social choice function (SCF) is a mapping $f : \overline{T}_0 \to A$ where $\overline{T}_0 \subset \overline{T}$. Assume also that $\{t_i\} \times \text{supp}\kappa_{t_i} \subset \overline{T}_0$ 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). Observe that this still includes complete-information models $\mathcal{T}^{\text{CI}}$ where $f : \Theta \to A$ and $\mathcal{T}^{\text{CI}}$ is a pair $\left(\overline{T}^{\text{CI}}, \overline{\kappa}\right)$ such that $T_i^{\text{CI}} = \Theta$ and $\overline{\kappa}_{t_i^\theta}\left[\left(\theta, t_{-i}^\theta\right)\right] = 1$ for every $t_i^\theta \in T_i^{\text{CI}}$ that corresponds to the payoff-relevant state $\theta \in \Theta$. We can now define truthful continuous implementation in this setting:

Definition 10. *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,

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

4.1. *Uniform-Weak Topology*

To state and prove our characterization of truthful continuous implementation, 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 3. *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.*

The basic idea of the proof of this Lemma is similar to Lemma 1. Indeed, several of the results presented here are close analogs of results in the case where the baseline model is one of complete information. In the interest of brevity we only discuss when the import of the result is novel. Our main characterization of truthful continuous implementation follows from Lemma 3.

THEOREM 6. *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.*

COROLLARY 8. *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 $U\left(\mathcal{M},\overline{\mathcal{T}}\right)$.*

4.2. *Product topology*

We can prove an analogous result to Theorem 3 for the product topology.

DEFINITION 11. *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 $t \in \overline{T}$, $R^{\infty}\left(t,\tilde{g}\right) = \{t\}$.*

THEOREM 7. *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 11.*

The nontrivial direction as before is necessity, i.e. to show that if an SCF f is truthfully continuously implementable (in the the product topology) then f must be implementable in the unique rationalizable action in the sense of Definition 11. The proof uses the results of Theorem 6 and Corollary 8 analogous to the proof of Theorem 3.

5. A REVELATION PRINCIPLE FOR CONTINUOUS IMPLEMENTATION?

We now define continuous implementation in this richer setting and consider the relation between continuous implementation and truthful continuous implementation for both topologies.

DEFINITION 12. *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 $U\left(\mathcal{M},\mathcal{T}\right)$ w.r.t. a metric d if*

- *a. $\sigma|_{\overline{T}}$ is a pure-strategy Bayes-Nash Equilibrium (resp. strict Bayes-Nash equilibrium) in $U\left(\mathcal{M},\overline{\mathcal{T}}\right)$;*

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 $U(\mathcal{M}, \mathcal{T})$.*

5.1. *Uniform-Weak Topology*

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 $U(\mathcal{M}, \overline{\mathcal{T}})$ if

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

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

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

The following lemma is again the key to our characterization of continuous implementation. The proof is analogous to the proof of Lemma 3.

LEMMA 4. *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 $U(\mathcal{M}, \overline{\mathcal{T}})$ 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 4 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 8. *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 $U(\mathcal{M}, \overline{\mathcal{T}})$.*

The basic idea of Theorem 8 is analogous to the proof of Theorem 6. 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 6, we therefore have that, with respect to the uniform-weak topology, a social choice function is continuously implementable iff it is truthfully continuously implementable.

5.2. *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 7). 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.[16] 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 7 since both messages are rationalizable for both types. ■

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

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 $\underline{\xi} < \xi$.

[16] 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.

	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 7). ■

Remark 1. *Observe that the example as stated is one where agents' types are common knowledge among the agents. The support of the typespace is therefore "diagonal," i.e. not the product typespace we require in our model.*

However 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.

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.[17] 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 (2012), Jehiel, Meyer-ter Vehn, and Moldovanu (2012) or Aghion, Fudenberg, Holden, Kunimoto, and Tercieux (2012)) the planner has some specific model in mind (e.g. in our paper, for example that the state of the world is common knowledge among agents) 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," i.e. mechanisms where the message space of agents is exactly the set of relevant states when preferences are common knowledge. 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 discuss sufficiency, they enrich the model to consider that sending various messages may involve small costs to the agents (and get the same characterization of rationalizable implementability). 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).[18] Another critical difference between our result and theirs is that our Theorem 3 or 5 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

[17] 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).

[18] We also refer the reader to Oury (2015), who characterizes continuous implementation as equivalent to full implementation in rationalizable strategies by introducing local payoff uncertainty of the planner.

thus both strengthen their results, and also constitute a key intermediate step to our characterization in the product topology. Further, they do not consider the case where only ordinal preferences are common knowledge (i.e., Section 3.4 of the present paper).

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 10). Formally, they look at sequential games where there is almost common certainty of payoffs (for our purposes, "almost" refers to close in the uniform-weak topology). The latter means that their results do not directly apply to our setting: our Theorem 8 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.[19]

At a conceptual level, we use these tight characterizations to suggest that their results may have an alternate interpretation. They suggest that their necessary conditions build a "first bridge between partial and full implementation." By contrast, as we argued previously, we suggest that the requirement of truthful continuous implementation under the product topology is as demanding as implementation in rationalizability.

To argue that implementation in rationalizability (Theorem 5, Corollary 7) is restrictive, we use results of Börgers (1995) which argues that implementation of a social choice function in iteratively undominated strategies is equivalent to the social choice function being dictatorial whenever the set of possible preferences is a superset of the unanimous preference profiles. Here undominated refers to "pure strategy dominance," a notion initially defined in Börgers (1993). As we mentioned earlier, our result uses a notion of ordinal rationalizability that is more permissive than the notion of Börgers—see Definition 8 and the discussion that follows for details. This models a setting where only agents' ordinal preferences over outcomes, rather than their von Neumann-Morgenstern utilities, are taken as known among them.

As we alluded to earlier, other papers have raised similar questions about "local" robust implementation. Chung and Ely (2003) consider a very similar question, asking about the possibility of (full) implementation in undominated Nash equilibrium while

[19]In fact, the failure of revelation principle occurs because of our requirement that truth-telling be played with probability 1 (part (b) of Definition 10) 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.

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 also consider settings close to complete information, 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 (Theorems 1, 2) can be thought of as complementing their findings—both suggest that in settings with approximate common knowledge of the state, 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.

In light of this bifurcation in approaches to robust mechanism design, our main takeaway can also be phrased thus: for a local approach to be non-trivially different than a global approach, it must place non-trivial constraints on agents' higher order beliefs. Recall that the results on global robustness (Bergemann and Morris (2005), Chung and Ely (2007)) provide foundations for dominant strategy/ Ex-Post IC mechanisms. An application of the Gibbard-Satterthwaite theorem therefore suggests these must be dictatorial.[20] In this sense, considering only truthful continuous implementation is no more permissive than general robust implementation.

[20]More precisely, we refer to a private-value setting where an agent's (payoff) type fully pins down his preference and thus ex post implementation and dominant-strategy implementation are equivalent. We also restrict attention to the case of a social choice function and hence our environment is separable in the sense of Bergemann and Morris (2005).

Appendix A. Omitted Proofs

Proof of Corollary 3. Consider the DRM g among agents for whom f is never pessimal defined as:

$$g(s) = \begin{cases} f(\theta) & \text{if } s = s^\theta, \\ a_i(\theta) & \text{if } s = (s_i', s_{-i}^\theta), \\ a & \text{otherwise.} \end{cases}$$

where $a_i(\theta) \in A$ is selected such that $a_i(\theta) \prec_{i,\theta} f(\theta)$ for every i, θ. Since f is never pessimal for these agents, we can always select some alternative, i.e. $a_i(\cdot)$ is well defined. The g thus constructed implements f in strict NE for those agents, and therefore $\tilde{g}$ is in strict NE (agents who are not never pessimal have only a single/ trivial message in $\tilde{g}$). By Corollary 1, therefore, g truthfully continuously implements f. ■

Proof of Theorem 4. ($\Rightarrow$) For each $\theta \in \Theta$, fix some $u_i^\theta \in U$ for each $i \in I$. Then, each (ordinal) state θ corresponds to a unique cardinal state $\left(\theta, \left(u_i^\theta\right)_{i \in I}\right) \in \Theta^*$ and hence Θ can be identified with a subset $\overline{\Theta}$ of Θ^*. Truthful continuous implementation in models over $\overline{\Theta}$ is precisely the notion studied in previous sections. Theorems 1 and 2 imply our result.

($\Leftarrow$) Since $\overline{T}$ is finite,

$$\overline{\Theta} \equiv \left\{\theta^* \in \Theta^* : \text{marg}_{\Theta^*} \kappa_{t_i} [\theta^*] > 0 \text{ for some } t_i \in \overline{T}_i\right\}$$

is also a finite set. Thus, we can pick $\varepsilon > 0$ such that whenever g strictly rewards unanimity at θ over θ', we have

$$(1-\varepsilon)\left[u_i\left(g(s_i^\theta, s_{-i}^\theta), \theta^*\right) - u_i\left(g(s_i^{\theta'}, s_{-i}^\theta), \theta^*\right)\right] > \varepsilon D, \forall \theta^* \in \overline{\Theta}.$$

where

$$D \equiv \max_{\theta^* \in \overline{\Theta}, i, s, s'} \left| u_i\left(g\left(s\right), \theta^*\right) - u_i\left(g\left(s'\right), \theta^*\right) \right|.$$

The rest of the argument is similar to the proof of Theorem 2. ■

Proof of Theorem 5. Before we proceed with this proof, we recall the definition of the set of interim correlated rationalizable messages in Definition 6. We apply the definition to the ordinal setup by replacing Θ with Θ^* and still denote by $R_i^\infty(t_i, \mathcal{M})$ the set of interim correlated rationalizable messages.[21] Observe that for every ordinal complete-information type $R_i^\infty\left(t_i^\theta, \mathcal{M}\right) \subset R_i^\infty(\theta, \mathcal{M})$.

[21]An argument similar to the backward direction shows that R_i^∞ has the usual best-response property, which ensures that the iterative definition is still proper even though Θ^* is not compact.

We are now in a position to proceed with the proof.

($\Leftarrow$): As in the proof of Theorem 3, it suffices to prove that if $d^{\mathrm{p}}\left(t_n, t^\theta\right) \to 0$ and $\tilde{s}_i \in R_i^\infty\left(t_{i,n}, \tilde{g}\right)$ for all n, then $\tilde{s}_i \in R_i^\infty\left(\theta, \tilde{g}\right)$: since $R_i^\infty(\theta, \tilde{g})$ contains a unique ordinally rationalizable message by Definition 9, it would follow that there is a unique interim correlated ordinally rationalizable message for nearby types.

Since each agent has only finitely many messages in $\tilde{g}$, $R_i^\infty\left(t_i^\theta, \tilde{g}\right) = R_i^{k^*}\left(t_i^\theta, \tilde{g}\right)$ for some finite k^*. Thus, it suffices to prove that for each k, if $d^{\mathrm{p}}\left(t_n, t^\theta\right) \to 0$ and $\tilde{s}_i \in R_i^k\left(t_{i,n}, \tilde{g}\right)$ for all n, then $\tilde{s}_i \in R_i^k\left(\theta, \tilde{g}\right)$.

Observe that Θ^* is an open subset of the Polish space $\Theta \times ([0,1]^{|A|})^I$ and hence also a Polish space. Moreover, write $u_i\left(\cdot, \theta^*\right)$ for the cardinalization of agent i at $\theta^* \in \Theta^*$. Obviously, $u_i\left(\cdot, \theta^*\right)$ is continuous in θ^*.[22]

We proceed by induction. The case with $k = 0$ is trivial. Now we prove the claim for $k \geq 1$, i.e. that if $d^{\mathrm{p}}\left(t_n, t^\theta\right) \to 0$ and $\tilde{s}_i \in R_i^k\left(t_{i,n}, \tilde{g}\right)$ for all n, then $\tilde{s}_i \in R_i^k\left(\theta, \tilde{g}\right)$, assuming that it is true for $k-1$.

Suppose that $\tilde{s}_i \in R_i^k\left(t_{i,n}, \tilde{g}\right)$ for every n. Hence, there is some $\mu_n \in \Delta\left(\Theta^* \times T_{-i} \times \tilde{S}_{-i}\right)$ such that (**R1**)-(**R3**) hold with respect to $t_{i,n}$ for each n. Since $d^{\mathrm{p}}\left(t_n, t^\theta\right) \to 0$, $\{t_{i,n}\}$ is relatively compact. Since $\Theta^* \times T_{-i}$ is Polish, by Prohorov's Theorem, $\{t_{i,n}\}$ is tight. Hence, for each $\varepsilon > 0$, there is some compact set $K_\varepsilon \subset \Theta^* \times T_{-i}$ such that $\kappa_{t_{i,n}}\left[K_\varepsilon\right] > 1 - \varepsilon$ for every n. It follows from (**R2**) that $\mu_n\left[K_\varepsilon \times \tilde{S}_{-i}\right] > 1 - \varepsilon$. That is, $\{\mu_n\}$ is also tight. Again, by Prohorov's Theorem, $\{\mu_n\}$ is relatively compact. Hence, $\{\mu_n\}$ has a limit point $\mu \in \Delta\left(\Theta^* \times T_{-i} \times \tilde{S}_{-i}\right)$. Let $\pi \equiv \mathrm{marg}_{U_i \times \tilde{S}_{-i}} \mu$.

First, by (**R3**) of μ_n, we know that

$$\mu_n\left(\left\{\left(\theta^*, t_{-i}, \tilde{s}_{-i}\right) : \tilde{s}_{-i} \in R_{-i}^{k-1}\left(t_{-i}, \tilde{g}\right)\right\}\right) = 1.$$

By finiteness of $\tilde{S}_{-i}$, the induction hypothesis implies that if $d^{\mathrm{p}}\left(t_n, t^\theta\right) \to 0$, there is some subsequence, say itself, such that $R_{-i}^k\left(t_{-i,n}, \tilde{g}\right) \subset R_{-i}^k\left(\theta, \tilde{g}\right)$. It follows from (**R2**), i.e. that $\mathrm{marg}_{\Theta \times T_{-i}}\, \mu_n = \kappa_{t_{i,n}}$ and finiteness of $\overline{T}$ that

$$\limsup_n \mu_n\left(\left\{\left(\theta^*, t_{-i}, \tilde{s}_{-i}\right) : \tilde{s}_{-i} \in R_{-i}^{k-1}\left(\theta, \tilde{g}\right)\right\}\right) = 1.$$

Since $\left\{\left(\theta^*, t_{-i}, \tilde{s}_{-i}\right) : \tilde{s}_{-i} \in R_{-i}^{k-1}\left(\theta, \tilde{g}\right)\right\}$ is closed and μ is a limit point of μ_n, Portmanteau theorem implies that

$$\mu\left[\left\{\left(\theta^*, t_{-i}, \tilde{s}_{-i}\right) : \tilde{s}_{-i} \in R_{-i}^{k-1}\left(\theta, \tilde{g}\right)\right\}\right] = 1.$$

[22]Note that, as is standard, a different metric than the standard Euclidean metric is needed for the open subset U of $\mathbb{R}^k$ to be complete. Let $d(\cdot,\cdot)$ denote the standard Euclidean metric, and $C = \mathbb{R}^{|A|} \backslash U$. Consider the metric $\hat{d}(x,y) = d(x,y) + \left|\frac{1}{d(x,C)} - \frac{1}{d(y,C)}\right|$. Note that $\hat{d}$ preserves d-open sets, but prevents sequences in U that converge to a point in C from being Cauchy.

Second, since $d_i^{\mathrm{p}}\left(t_{i,n}, t_i^{\theta}\right) \to 0$ and μ_n satisfies (**R2**) with respect to $t_{i,n}$, it follows that $\pi\left[U_i^{\theta} \times R_{-i}^{k-1}(\theta, \tilde{g})\right] = 1$, where U_i^{θ} is the set of utility functions consistent with θ. Finally, since μ_n satisfies (**R1**) and $u_i\left(\tilde{s}_i', \cdot, \cdot\right)$ is continuous in $\left(\tilde{s}_{-i}, \theta^*\right)$,

$$\tilde{s}_i \in \arg\max_{\tilde{s}_i'} \int_{\Theta^* \times \tilde{S}_{-i}} u_i\left(\tilde{g}(\tilde{s}_i', \tilde{s}_{-i}), \theta^*\right) \operatorname{marg}\mu_{\Theta^* \times \tilde{S}_{-i}}\left[d\theta^*, \tilde{s}_{-i}\right].$$

Since $\pi \equiv \operatorname{marg}_{U_i \times \tilde{S}_{-i}} \mu$, it follows that

$$\tilde{s}_i \in \arg\max_{\tilde{s}_i'} \int_{U_i^{\theta} \times R_{-i}^{k-1}(\theta, \tilde{g})} u_i\left(\tilde{g}(\tilde{s}_i', \tilde{s}_{-i}), \theta\right) \pi\left[u_i, \tilde{s}_{-i}\right].$$

Thus, $\tilde{s}_i \in R_i^k(\theta, \tilde{g})$.

($\Rightarrow$): As in the proof of Theorem 3, it suffices to prove the following Lemma.

LEMMA 5. *Let $\varepsilon \in (0,1)$ and $T_{i,0,\varepsilon} \equiv \overline{T}_i$. Then, for each $k \geq 1$, there is a model $\mathcal{T}_{k,\varepsilon} \supset \overline{\mathcal{T}}$ such that $T_{i,k,\varepsilon} \equiv \left(\bigsqcup_{\theta \in \Theta} R_i^k(\theta, \tilde{g})\right) \bigsqcup T_{i,k-1,\varepsilon}$.*

This model $\mathcal{T}_{k,\varepsilon}$ has the property that for any BNE $\tilde{\sigma}$ in the game $U(\tilde{g}, \mathcal{T}_{k,\varepsilon})$ with $\tilde{\sigma}\left(t^{\theta}\right) = \delta_{\tilde{s}^{\theta}}$, we have that for any action $\tilde{s}_i \in R_i^k(\theta, \tilde{g})$, there exists a type $t_{i,k,\varepsilon}(\tilde{s}_i, \theta)$ in $T_{i,k,\varepsilon}$ such that:

(1) $d_i^k\left(t_{i,k,\varepsilon}^k(\tilde{s}_i, \theta), \left(t_i^{\theta}\right)^k\right) < \varepsilon$, and,

(2) $\tilde{\sigma}_i\left(t_{i,k,\varepsilon}^k(\tilde{s}_i, \theta)\right) = \delta_{\tilde{s}_i}$.

PROOF OF LEMMA 5. 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 $k-1$. By definition, each $\tilde{s}_i \in R_i^k(\theta, \tilde{g})$ is a best response against some belief $\pi \in \Delta\left(U_i^{\theta} \times R_{-i}^{k-1}(\theta, \tilde{g})\right)$. Let $\varphi_i : \Theta^* \to U_i^{\theta}$ be a mapping such that $\varphi_i(\theta^*) = u_i(\cdot, \theta^*)$. By the induction hypothesis, there is a one-to-one mapping $\eta_{-i} : R_{-i}^{k-1}(\theta, \tilde{g}) \to T_{-i,k-1,\varepsilon}$ such that

$$\eta_{-i,k-1,\varepsilon}(\tilde{s}_{-i}) = t_{-i,k-1,\varepsilon}(\tilde{s}_{-i}, \theta).$$

Then, define $\kappa_{t_{i,k,\varepsilon}(\tilde{s}_i)} \in \Delta\left(\Theta^* \times T_{-i,k,\varepsilon}\right)$

$$\kappa_{t_{i,k,\varepsilon}(\tilde{s}_i)} = (1-\varepsilon)\left(\pi \circ (\varphi_i \times \eta_{-i})^{-1}\right) + \varepsilon \delta_{\left(s_i, t_{-i}^{s_i}\right)}.$$

That is, with probability $(1-\varepsilon)$, type $t_{i,k,\varepsilon}(\tilde{s}_i, \theta)$ believes that the cardinal state and opponents' types are distributed according to that induced from π through $\varphi_i \times \eta_{-i}$ (where each $t_{-i,k-1,\varepsilon}(\tilde{s}_{-i}, \theta)$ plays $\tilde{\sigma}_{-i}\left(t_{-i,k-1,\varepsilon}(\tilde{s}_{-i}, \theta)\right) = \tilde{s}_{-i}$ by the induction hypothesis); with probability ε, type $t_{i,k,\varepsilon}(\tilde{s}_i, \theta)$ believes that the state is some s_i from the equivalent class $\tilde{s}_i$ and that the opponents have common belief about the state being s_i (and thereby plays $\tilde{\sigma}_{-i}\left(t_{-i}^{s_i}\right) = \delta_{\tilde{s}_{-i}^{s_i}}$). Since $\tilde{s}_i$ is a best response against π and the unique best response against

$\left(s_i, \tilde{s}^{s_i}_{-i}\right)$ in $\tilde{g}$, it follows that $\tilde{\sigma}_i\left(t_{i,k,\varepsilon}\left(\tilde{s}_i,\theta\right)\right) = \delta_{\tilde{s}_i}$. Moreover, since

$$d_i^{k-1}\left(t^{k-1}_{-i,k-1}\left(\tilde{s}_{-i},\theta\right), \left(t^{\theta}_{-i}\right)^{k-1}\right) < \varepsilon$$

for each $\tilde{s}_{-i} \in R^{k-1}_{-i}\left(\theta,\tilde{g}\right)$, it follows that $d_i^k\left(t^k_{i,k,\varepsilon}\left(\tilde{s}_i,\theta\right), \left(t^{\theta}_i\right)^k\right) < \varepsilon$. ■

As before, 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 5. Further, by Lemma 5 for any message $\tilde{s}_i \in R_i^{\infty}(\theta,\tilde{g})$ we can construct a sequence of types in T_i converging to the complete information type t_i^{θ} such that the unique BNE action in $\tilde{g}$ is $\tilde{s}_i$. Truthful continuous implementability of f then implies that $\tilde{s}_i = \tilde{s}_i^{\theta}$ concluding our proof. ■

PROOF OF LEMMA 3. 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.,

$$\sum_{(\theta,t_{-i})\in\Theta\times\overline{T}_{-i}}\left[u_i\left(g(t'_i,t_{-i}),\theta\right) - u_i\left(g(t_i,t_{-i}),\theta\right)\right]\overline{\kappa}_{t_i}\left[(\theta,t_i)\right] \geq 0 \tag{8}$$

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

$$u_i\left(g(t'_i,t''_{-i}\right) - u_i\left(g(t_i,t''_{-i}\right) > 0. \tag{9}$$

Since $\left|\overline{T}\right| < \infty$ and f is truthfully continuously implementable by a DRM g with respect to d^{uw}, for every agent j we have $\sigma_j\left(t_{j,n}\right) = t_j$ for sufficiently large n. Hence, under σ_{-i}, by reporting t'_i, agent i with type $t_{i,n}$ gets the interim expected payoff equal to

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

Then, by (8) and (9), for agent i with type $t_{i,n}$, reporting t'_i is strictly better than reporting t_i. This is a contradiction to $\sigma_i\left(t_{i,n}\right) = t_i$. ■

PROOF OF THEOREM 6. ($\Rightarrow$) 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 3 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.

($\Leftarrow$)Let g be a DRM that truthfully continuously implements f in the sense of Definition 10. 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

$$(1-\varepsilon) \sum_{(\theta,t_{-i})\in\Theta\times\overline{T}_{-i}} \left[u_i\left(g(t_i,t_{-i}),\theta\right) - u_i\left(g(t_i',t_{-i}),\theta\right)\right] \overline{\kappa}_{t_i}\left[(\theta,t_{-i})\right] > \varepsilon D. \tag{10}$$

Here D is defined as

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

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 $\left(d_i^{\text{uw}},\varepsilon\right)$-ball around t_i does not overlap with the $\left(d_i^{\text{uw}},\varepsilon\right)$-ball around t_i', i.e. these balls are disjoint; and secondly,

$$\begin{aligned} &d_i^{\text{uw}}\left(t_i'',t_i\right) < \varepsilon \\ \Longrightarrow\ &\kappa_{t_i''}\left[\left\{(\theta,t_{-i})\right\}^{\varepsilon}\right] \in \left[(1-\varepsilon)\kappa_{t_i}\left[(\theta,t_i)\right],(1+\varepsilon)\kappa_{t_i}\left[(\theta,t_i)\right]\right] \end{aligned} \tag{11}$$

where $\left\{(\theta,t_{-i})\right\}^{\varepsilon}$ denotes the $\left(d_{-i}^{\text{uw}},\varepsilon\right)$-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 $(d_{-i}^{\text{uw}},\varepsilon)$ close to t_{-i}.

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

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

Furthermore, σ is a BNE in the original game $U\left(\mathcal{M},\mathcal{T}\right)$. 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 (10) and (11).

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

PROOF OF THEOREM 7. ($\Leftarrow$): 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 $U\left(g,\mathcal{T}\right)$. 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 identified with the set of equivalent messages 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

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

Since $\tilde{\sigma}$ is a BNE in $\tilde{g}$, it follows that $\tilde{\sigma}_i(t_i') = \delta_{\tilde{t}_i}$ for any $t_i' \in T_i$ with $d_i^{\mathrm{p}}(t_i', t_i) < \varepsilon$. 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

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

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 10.

($\Rightarrow$): 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 6 and Corollary 8, f is implementable in strict BNE in $\tilde{g}$.

The following lemma will be useful.

LEMMA 6. *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 $\tilde{t}_i' \in R_i^k(t_i, \tilde{g})$),

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

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

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 $\tilde{t}_i' \in R_i^k(t_i, \tilde{g})$, 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, $t_{i,k,\varepsilon}\left(\tilde{t}'_i, t_i\right)$ that is ε-close to t_i in their k-th-order beliefs; moreover, player i of type $t_{i,k,\varepsilon}\left(\tilde{t}'_i, t_i\right)$ must play $\tilde{t}_i$ under the BNE $\tilde{\sigma}$.

Before we present the proof of Lemma 6, let us conclude the now routine proof of Theorem 7. 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 6.

Since f is truthfully continuously implementable w.r.t. d^{P}, there is a BNE σ in the game $U\left(g, \mathcal{T}\right)$ such that requirements (a) and (b) in Definition 10 hold. Again, σ induces a BNE $\tilde{\sigma}$ in $\tilde{g}$. Since $\sigma\left(t\right) = \delta_{\tilde{t}}$ by requirement (b) of Definition 10, we have $\tilde{\sigma}\left(t\right) = \delta_{\tilde{t}}$.

Thus, it follows from Lemma 6 that for each $\tilde{t}'_i \in R_i^{\infty}\left(t_i, \tilde{g}\right)$, for each k, there is a type $t_{i,k,\frac{1}{k}}\left(\tilde{t}'_i, t_i\right) \in T_i$ such that

$$d_i^k\left(t_{i,k,\frac{1}{k}}\left(\tilde{t}'_i, t_i\right), t_i^k\right) \leq \frac{1}{k}, \tag{12}$$

and

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

It follows from (12) that $d_i^{\mathrm{P}}\left(t_{i,k,\frac{1}{k}}\left(\tilde{t}'_i, t_i\right), t_i\right) \to 0$. Since σ satisfies requirement (b) in Definition 10, we know that it must be the case that $\sigma_i\left(t_{i,k,\frac{1}{k}}\left(\tilde{t}'_i, t_i\right)\right) = \delta_{\tilde{t}_i}$ for any k large enough. Hence, $\tilde{t}'_i = \tilde{t}_i$.

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

PROOF OF LEMMA 6. 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 $\tilde{T}_i \equiv \left\{\tilde{t}_i : t_i \in \overline{T}_i\right\}$. By definition, each $\tilde{t}'_i \in R_i^k\left(t_i, \tilde{g}\right)$ is a best response to some belief $\mu_{-i} \in \Delta\left(\Theta \times \overline{T}_{-i} \times \tilde{T}_{-i}\right)$ such that:

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

By the induction hypothesis, there is a mapping $\eta_{-i,k-1,\varepsilon}$ from each $t_{-i} \in \overline{T}_{-i}$ and $\tilde{t}'_{-i} \in R_{-i}^{k-1}\left(t_{-i}, \tilde{g}\right)$ to a type $t_{-i,k-1,\varepsilon}\left(\tilde{t}'_{-i}, t_{-i}\right)$ such that (1) and (2) in Lemma 6 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 $\kappa_{t_{i,k,\varepsilon}\left(\tilde{t}'_i, t_i\right)} \in \Delta\left(\Theta \times T_{-i,k,\varepsilon}\right)$

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

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 8), 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

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

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

PROOF OF LEMMA 4. 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 $d_j^{\text{uw}}\left(t_{j,n}^{\left(t_j,\theta',t'_{-j}\right)},t_j\right) \to 0$. 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

$$g(\sigma(t)) = f(t), \forall t \in \overline{T}. \tag{13}$$

$$g\left(\sigma_i\left(t_{i,n}^{\left(t_i,\theta',t'_{-i}\right)}\right),\sigma_{-i}(t_{-i})\right) \to f(t_i,t_{-i}), \forall t_{-i} \in \overline{T}_{-i} \tag{14}$$

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.,

$$\sum_{(\theta,t_{-i})\in\Theta\times\overline{T}_{-i}}\left[u_i\left(g\left(m'_i,\sigma_{-i}(t_{-i})\right)\right),\theta) - u_i\left(g\left(\sigma_i(t_i),\sigma_{-i}(t_{-i})\right)\right),\theta)\right]\overline{\kappa}_{t_i}[(\theta,t_{-i})] \geq 0. \tag{15}$$

and for some t'_{-i} and θ',

$$u_i\left(g\left(m'_i,\sigma_{-i}(t_{-i})\right)\right),\theta) - u_i\left(g\left(\sigma_i(t_i),\sigma_{-i}(t_{-i})\right)\right),\theta) > 0. \tag{16}$$

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

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

where the equality follows from (13). Second, by (14) 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,

$$\begin{aligned}\sigma_i\left(t_{i,n}^{\left(t_i,\theta',t_{-i}'\right)}\right)\left[M_i^{t_{-i}}\right] &\geq 1-\frac{1}{2\left|\overline{T}_{-i}\right|};\\ g\left(m_i,\sigma_{-i}\left(t_{-i}\right)\right) &= f\left(t_i,t_{-i}\right)=g\left(\overline{\sigma}_i\left(t_i\right),\overline{\sigma}_{-i}\left(t_{-i}\right)\right)),\forall m_i\in M_i^{t_{-i}}.\end{aligned}\tag{17}$$

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

$$\sigma_i\left(t_{i,n}^{\left(t_i,\theta',t_{-i}'\right)}\right)\left[\bigcap_{t_{-i}\in\overline{T}_{-i}}M_i^{t_{-i}}\right]\geq\sum_{t_{-i}\in\overline{T}_{-i}}\sigma_i\left(t_{i,n}^{\left(t_i,\theta',t_{-i}'\right)}\right)\left[M_i^{t_{-i}}\right]-\left(\left|\overline{T}_{-i}\right|-1\right)>0\tag{18}$$

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

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

where the equality follows from (17). Then, by (15), (16) and (18), agent i of type $t_{i,n}^{\left(t_i,\theta',t_{-i}'\right)}$ 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 8. ($\Rightarrow$) Consider a mechanism where the message space of each agent i is restricted to $\overline{\sigma}_i(\overline{T}_i)$. By Lemma 4 and Theorem 6 such a mechanism clearly truthfully continuously implements f. The implication now follows from Corollary 8.

($\Leftarrow$) By Corollary 8, an scf f satisfying this condition is truthfully continuously implementable and therefore trivially, also continuously implementable. ■

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.

BERGEMANN, D., S. MORRIS, AND O. TERCIEUX (2011): "Rationalizable implementation," *Journal of Economic Theory*, 146(3), 1253–1274.

BERNHEIM, B. D. (1984): "Rationalizable strategic behavior," *Econometrica*, 52(4), 1007–1028.

BÖRGERS, T. (1993): "Pure strategy dominance," *Econometrica*, 61(2), 423–430.

——— (1995): "A note on implementation and strong dominance," in *Social Choice, Welfare, and Ethics: Proceedings of the Eighth International Symposium in Economic Theory and Econometrics*, vol. 8, p. 277. Cambridge University Press.

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.

MULLER, E., AND M. A. SATTERTHWAITE (1977): "An impossibility theorem for voting with a different interpretation," *Journal of Economic Theory*, 14, 412–418.

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.

PEARCE, D. G. (1984): "Rationalizable strategic behavior and the problem of perfection," *Econometrica*, 52(4), 1029–1050.

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.

SAIJO, T. (1987): "On constant Maskin monotonic social choice functions," *Journal of Economic Theory*, 42(2), 382–386.

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. (2016): "The Effect of Changes in Risk Attitude on Strategic Behavior," *Econometrica*, 84(5), 1881–1902.

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.

XIONG, S. (2017): "Designing Referenda: An Economist's Pessimistic Perspective," .

APPENDIX B. SUPPLEMENTARY/ NOT FOR PUBLICATION APPENDIX

B.1. *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 $\text{supp}\overline{\kappa}_{t_i} =$. 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 9. *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}.*

PROOF. 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 4. Then, we must have some BNE σ which continuously implements f in $U\left(\mathcal{M}, \mathcal{T}''\right)$ (and there by in $U\left(\mathcal{M}, \mathcal{T}\right)$). It follows from Lemma 4 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 $\sigma_i|_{\overline{T}}\left(t_i\right)$ is not strategically equivalent to m_i', the message m_i' must yield strictly lower payoff than $\sigma_i|_{\overline{T}}\left(t_i\right)$ 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 $U\left(\mathcal{M}, \mathcal{T}'\right)$. Hence, $\tilde{\mathcal{M}}$ strictly continuously implements f w.r.t. 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 13. *Let $\mathcal{T} = (T, \kappa)$ be a model. Denote by $W_i^{\infty}\left(t_i, \mathcal{M}\right)$ the set of (interim correlated)* strictly rationalizable *messages of type t_i in $U\left(\mathcal{M}, \mathcal{T}\right)$ defined as follows:*

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

R1: $\{m_i\} = \arg\max_{m_i'} \sum_{\theta, m_{-i}} u_i\left(m_i', m_{-i}, \theta\right) \operatorname{marg}\ \mu_{\Theta \times M_{-i}}[\theta, m_{-i}]$;
R2: $\operatorname{marg}_{\Theta \times T_{-i}} \mu_{-i} = \kappa_{t_i}$;
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 14. *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 10. *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.*

Proof. Since $\mathcal{M} = (M, g)$ continuously implements f w.r.t. d^{p}, by Theorem 9, 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 7. *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 7, let us conclude the now routine proof of Theorem 10. 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 7. Since $\mathcal{M}$ strictly continuously implements f w.r.t. d^{p}, there is some BNE σ which strictly continuously implements f in $U(\mathcal{M}, \mathcal{T})$. Hence, $\sigma|_{\overline{T}}$

is a pure-strategy strict BNE in $U\left(\mathcal{M}, \overline{\mathcal{T}}\right)$. It follows from Lemma 7 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

$$t_{i,k}^k\left(m_i, t_i\right) = t_i^k \tag{19}$$

and

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

It follows from (19) 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 7. 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 7 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 $t_{i,k}^k\left(m_i, t_i\right) = t_i^k$. ■

The result clarifies the trade-off between Oury and Tercieux (2012) and our paper. The trade-off is actually not that they allow indirect mechanism whereas we focus on direct revelation mechanisms but rather that of the two assumptions.

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 10 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.

B.2. *Strict Nash Implementation*

Strict nash implmentation is well know to be be without loss as long as there are at least 3 agents for whom the social choice function is never pessimal. The following corollary characterize truthful continuous implementation if there are fewer than 3.

A few standard definitions will be helpful. Agent i is a *strict dictator* for f if for all θ, $f(\theta)$ is the unique solution to the problem $\max_{a \in f(\Theta)} u_i(a, \theta)$. We say f satisfies *strict self-selection* for agents i and j if for every pair of states θ, θ', there exist a and a' such that:

$$u_i(f(\theta), \theta) > u_i(a, \theta) \text{ and } u_j(f(\theta'), \theta') > u_j(a, \theta'),$$
$$u_i(f(\theta'), \theta') > u_i(a', \theta') \text{ and } u_j(f(\theta), \theta) > u_j(a', \theta).$$

COROLLARY 9. *(a) If the set of agents for whom f is never pessimal is empty, then a DRM g that satisfies Assumption 1 truthfully continuously implements f if and only if f is constant.*

(b) If there is only one agent for whom f is never pessimal, then a DRM g that satisfies Assumption 1 truthfully continuously implements f iff this agent is a strict dictator for f.

(c) If there are two agents for whom f is never pessimal, the social choice function f is truthfully continuously implementable if f satisfies strict self-selection for these two agents.

PROOF. (1) and (2) The if part is obvious. The only if part follows immediately from Corollary 2.

(3) Consider the DRM g defined among these agents such that $g(s^\theta) = f(\theta)$, further for every pair θ and θ', define:

$$g(s_i^{\theta'}, s_j^{\theta}) = a,$$
$$g(s_i^{\theta}, s_j^{\theta'}) = a',$$

where a and a' are the alternatives identified in the definition of strict self-selection. By construction, g strictly rewards unanimity for both agents and each pair of states. ■

This corollary implies that truthful continuous implementation, while permissive, still rules out social choice functions that may be of interest. For example, consider a setting where the state $\theta = (\hat{\theta}, \hat{\mathbf{u}})$ has two components: the first component identifies some fundamental state of uncertainty, and the second identifies the tuple of cardinal utilities of all agents over the set of alternatives. To remain within the general assumptions of our model assume that the space of fundamental uncertainty $\hat{\Theta}$ is finite and that the space $\hat{U}$ of cardinal utility tuples on A is finite, but every possible profile of ordinal rankings over A has one representation. The state space Θ then equals $\hat{\Theta} \times \hat{U}$.

Suppose now that the social planner cares only about the fundamental state of the world and not about the preferences of the agents. In other words, assume that the social choice function is *preference invariant*, i.e. for every $\hat{\theta} \in \hat{\Theta}$ we have $f(\hat{\theta}, \hat{\mathbf{u}}) = f(\hat{\theta}, \hat{\mathbf{u}}')$ for all $\hat{\mathbf{u}}, \hat{\mathbf{u}}' \in \hat{U}$.

COROLLARY 10. *Suppose that Assumption 2 holds. A truthful continuously implementable social choice function that is preference invariant must be constant.*

The Corollary follows from Corollary 4 and the fact that preference invariance of the social choice function and the assumption on $\hat{U}$ directly imply that no agent is never pessimal for f. Therefore, truthful continuous implementation fails.