TRUTHFUL CONTINUOUS IMPLEMENTATION

YI-CHUN CHEN[a], MANUEL MUELLER-FRANK[b], AND MALLESH M. PAI[c]
JANUARY 8, 2018

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 (here, common knowledge of preferences) 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 Nash 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 rationalizable strategies in the initial model, and further that there be a unique rationalizable strategy. This is a very restrictive requirement. 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.

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
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-1101389.

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 this 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. 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. Our paper substantially builds off their work, we defer a fuller discussion of the details of this (and other related papers) to Section 5, after we have formally stated our own results.

For the sake of motivation, consider two applied settings that exemplify two different ways that agents could be close to complete information. Both are variants of, say, 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 have not received some of these surveys. Alternately, consider the variant that with some small probability, some of the agents might have commissioned additional, accurate surveys 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.[1] 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.

In other words, when the principal is unsure about the agents' higher order beliefs, she cannot construct a game such that agents robustly reveal the state to the principal. Unless one of the agents' preferences aligns with the principal's state by state, the principal cannot robustly implement her desired social choice function.

We also introduce a modeling innovation not previously considered in this literature. In particular, this literature normally considers settings where the uncertainty agents have is about each others' cardinal utilities over alternatives. At complete information, therefore, agents know each others' cardinal preferences exactly. In Section 4, we consider a setting where even at complete information, agents only know each others' ordinal preferences. In settings without transfers, this may be more natural to envisage. [2] Our results in this setting are similar to those in the original, described above.

[1] 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).

[2] Indeed, 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 letter).

1.1. *Model and Results*

Let us now be slightly more formal in describing the setting and our results. To repeat from earlier, in our 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. The baseline model/ information structure is that the state is common knowledge among agents.

We restrict attention to mechanisms where each agent's message space is the set of states, i.e. a direct revelation mechanism in the baseline model. Of course, restricting attention to this space of mechanisms is with loss if we consider richer models.[3] However, we feel that this is a natural restriction to consider since direct revelation mechanisms are without loss under the baseline model due to the Revelation Principle.

One way to interpret our restriction is that it formalizes conditions under which a principal who believes the state of the world is common knowledge among agents 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. Further, conceptually, 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.

Before we proceed, a qualifier is in order. The most general form of our results requires a non-standard assumption (Assumption 1). Formally, we characterize when a mechanism truthfully continuously implements the social choice function under the assumption that the mechanism has a well-defined reduced normal form. That is to say if any strategically equivalent messages for any agents were merged into a single message, the resulting mechanism is well-defined. This is an assumption that may be satisfied by various combinations of conditions on the environment, the social choice function to be implemented, and the mechanism itself. For full generality, we state our results in terms of this non-standard assumption. However we recognize that this assumption may not be fully satisfactory for readers. To this end, we point out that simple richness assumptions on the primitives are sufficient to satisfy this assumption.

The cleanest sufficient condition is purely on the environment— informally, that for any two alternatives, and any agent, there exists some state at which the agent is not indifferent between these two alternatives (Assumption 2). This sort of rich domain assumption is standard and weaker than the full domain assumptions that some impossibility results

[3]Formally, there exist social choice functions that are continuously implementable by general mechanisms but not truthfully continuously implementable in the sense of Definition 2—see e.g. the example 4.1 in the working paper version of Oury and Tercieux (2012).

require (e.g. Gibbard-Satterthwaite). Nevertheless, we recognize may be strong in some settings. For instance, in a private-good allocation setting, agent 1 may be always indifferent between the alternatives "agent 2 gets the good" and "agent 3 gets the good." WE should point out, however, that even in such settings our results have bite—for example, the social choice function corresponding to efficient allocation is such that any mechanism that implements this cannot have any strategically equivalent messages (at some profile of others' reports, an agent's report affects whether they get the good), and therefore satisfies our more general assumption. We refer the reader to Section 2.3 for more details.

We use the standard formulation of incomplete information introduced in Harsanyi (1967) and developed in Mertens and Zamir (1985). We require that in any (epistemic) model that embeds our baseline model, there is an equilibrium of the direct revelation mechanism that yields the desired social choice function at all types close to the initial common knowledge 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 the universal type space. 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 common knowledge, i.e. types close to the common knowledge types in this topology still have approximately common knowledge of the state in the following sense: for some p close to 1, they assign probability no smaller than p to the state, assign probability no smaller than p to the event that the state obtains and the other players assign probability no smaller than p to the state, and so forth, *ad infinitum*. 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 Nash Equilibrium in the baseline common knowledge model (Corollary 1, Theorem 2). We argue that implementation in Strict Nash Equilibrium is a permissive solution concept: for example with three or more agents, it is enough that the social choice function never attempts to implement the worst alternative of at least three agents (corollaries in Section 3.2).

The second, the product topology, used in Weinstein and Yildiz (2007) places no restrictions on agents' higher order beliefs. Formally, types close to the common knowledge types in the product topology attach large probability to the state up to arbitrarily high but finite order. 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 common knowledge model, each agent has a unique rationalizable action, and the desired alternative of the social choice function obtains if each agent plays this unique rationalizable action (Theorem 3).

To argue that this is a very demanding solution concept in the strongest possible form, we consider the ordinal setting considered in the early implementation literature. In other words, we only assume agents know each others' ordinal preferences over alternatives, not the cardinal utilities. Rationalizability therefore involves considering all cardinalizations that are consistent with the ordinal preferences. We characterize truthful continuous implementation with respect to the product topology in this setting to be requiring that the social choice function be uniquely ordinal rationalizable implementable in the original common knowledge model (Theorem 5).[4] 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 9). No appeal to a full or rich domain is required.

To get some intuition for our characterization result in the product topology, recall the work of Weinstein and Yildiz (2007). They consider a *given* game of incomplete information. They assume a form of richness: for each player, and each action of that player, there exists a "crazy type" whose preferences make that action strictly dominant. Their main result is to show that for any action a that is rationalizable for a (normal) type in the game, there exist close-by types in the product topology for whom that action is the unique rationalizable action. The possibility of aforementioned crazy types is used to start a contagion process, with the strict dominance used to break ties.

In an implementation setting, this assumption of crazy types is not well grounded, since the game form is chosen by the planner and therefore not fixed a priori. Further, we are after a partial equilibrium result, i.e. there exists one equilibrium of the game with the desired properties.[5]

Instead our result in the product topology builds off of our result in the uniform-weak topology. Closeness in the uniform-weak topology implies closeness in the product topology. By our results in the former topology, we know that the social choice function must 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

[4]Our definition of ordinal rationalizable differs slightly from that in Börgers (1993). See Definition 8 and the discussion that follows for details.

[5]In this sense, there is a tighter connection between our results and those of Weinstein and Yildiz (2004), we discuss the details after we introduce our formal result, Theorem 3. See also Weinstein and Yildiz (2011).

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 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 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, Theorem 3.

The organization of the rest of the paper is as follows. Section 2 defines the model, including the notion of continuous implementation under three notions of robustness. The first (Section 2.1) is a baseline model where each agent receives a private signal realization about the true state of the world and agents share a common prior over states and signal realizations. This common prior is such that the true state is close to common knowledge among agents, but the prior is unknown to the principal. Studying this restricted setting helps us build intuition. The second and third notions consider types close to the common knowledge type in the universal type space, and consider closeness in the uniform-weak and product topologies (Section 2.2). Section 3 characterizes truthful continuous implementation in each of these three notions. Section 4 introduces the ordinal model, and characterizes truthful continuous implementation in this model. Section 5 then discusses the related literature and connections. Finally Section 6 discusses some critical assumptions, highlights possible extensions and concludes.

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, and $\sim_{i,\theta}$ 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 $s_i \in S_i \equiv \Theta$, $S \equiv \prod_i S_i$.

These signals are drawn according to a 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 $\theta \in \Theta$.[6]

A mechanism/ game form, denoted (M, g) is a message space M_i for each player i, and an outcome function $g : M \to A$. A mixed 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 $m \in supp\ \sigma(s)$ 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), with two major differences. First we restrict attention to "direct revelation mechanisms," i.e. where the message space equals the set of possible states of the world Θ. Secondly, of course, this definition is adapted to the baseline 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 $n \geq \underline{n}$.[7]

[6] 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.

[7] 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.

2.2. *Universal type space and topologies*

To study continuous implementation under general perturbations, we introduce the following from Oury and Tercieux (2012), Section 4.2. 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 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 $\overline{\kappa}_{t_i^\theta}\left[(\theta, t_{-i}^\theta)\right] = 1$ for each $\theta \in \Theta$.

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 $E \subset \Theta \times T_{-i}$.

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

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

We can compute the entire hierarchy of beliefs $(t_i^1, t_i^2, \ldots, t_i^k, \ldots)$ by proceeding in this way.

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 $(\Delta(\Theta))$.[8] 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 $\kappa_i^* : T_i^* \to \Delta(\Theta \times T_{-i}^*)$.

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

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

Moreover, write $d^{\text{uw}}(t_n, t) \to 0$ if $d_i^{\text{uw}}(t_{i,n}, t_i) \to 0$ for each i.[9]

[8] For a metric space (X, ρ), the Prohorov distance between any two $\mu, \mu' \in \Delta(X)$ is
$$\inf\{\gamma > 0 : \mu'(E) \leq \mu(E^\gamma) + \gamma \text{ for every Borel set } A \subseteq X\},$$
where $E^\gamma = \{x \in X : \inf_{y \in E} \rho(x, y) < \gamma\}$.

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

We say that a sequence of types $\{t_{i,n}\}_{n=1}^{\infty}$ converges in the product topology to a type t_i if

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

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

The following adapts Definition 1 to now consider continuity in these topologies defined above. 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 $\mathcal{T} \supset \overline{T}$, there is a (possibly mixed) BNE σ in the game $U\left(\mathcal{M}, \mathcal{T}\right)$ 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) $g(s^\theta) = f\left(\theta\right)$, and,
(b) $\sigma\left(t_n\right) = \delta_{s^\theta}$ for any n large enough.

Definition 2 is directly comparable to Definition 2 of Oury and Tercieux (2012). There are two major differences. First, as discussed before, we restrict attention to implementation by what we term direct revelation mechanisms. Second 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 $g(s^\theta) = f\left(\theta\right)$. 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. The following definition is the standard definition of strategic equivalence applied 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.:*

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

In light of this we could consider defining the reduced normal-form of our 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. There is then an indeterminacy as to which outcome should be selected in the reduced normal form mechanism $\tilde{g}$ (formally 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, introduced by 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, our 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.*

As we said in the Introduction, 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 where possible. The reader may mentally substitute the stronger/ sufficient, but also more elegant Assumption 2 if they prefer.

3. Results in the Cardinal Model

3.1. *Common Prior Perturbations*

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

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'.*

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', 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_{-i}^T 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 reporting s_i^θ for every n. Thus, truth-telling is not a BNE of g under P_n for every n. This is a contradiction. ■

Thus, we obtain our main characterization of truthful continuous implementation as 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. ($\Rightarrow$) Observe that when a DRM g strictly rewards unanimity at θ over θ', then θ' cannot always weakly dominate θ. Thus, 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

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

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. ■

3.2. *Characterizing Implementation*

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. Thus, it follows from Theorem 1 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 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.*

Truthful continuous implementation therefore only has bite if there are fewer than three agents for whom f is never pessimal. The following corollary characterize truthful continuous implementation in these cases.

A few standard definitions will be helpful. We say that 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 alternatives 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 4. *(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 if and only if 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.

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 5. *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.

3.3. *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 ex-ante stage, together with the fact that there is a common prior, makes the proofs of Theorem 1 easier but do 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

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

where D is defined as in (4).

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 3.2. 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).

3.4. *Product topology*

Finally, we consider truthful continuous implementation in the product topology. As before, we still consider what we term direct revelation mechanisms, i.e. game forms where the message space of each player i is $S_i = \Theta$.

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 $_{\Theta \times M_{-i}} \mu [d\theta, m_{-i}]$;
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) $R^\infty\left(t^\theta, \tilde{g}\right) = \left\{\tilde{s}^\theta\right\}$.

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 truthfully continuously 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 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 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. From the fact that 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

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

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 $\tilde{s}_i = \tilde{s}_i^\theta$.

This almost concludes the proof. For completeness, 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.[10] Moreover, $g\left(s^\theta\right) = f(\theta)$ and by construction $\overline{\sigma}$ also satisfies requirement (2) 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 $\tilde{g}$.

The following Lemma will be useful.

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

[10] 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.

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, t_i^θ; such that 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 $T_i = \bigsqcup_{k=1}^{\infty} T_{i,k,\frac{1}{k}}$ and $\mathcal{T}_{k,\frac{1}{k}}$ is given as in Lemma 2.

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 $t_{i,k,\frac{1}{k}}(\tilde{s}_i, \theta) \in T_i$ such that

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

and

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

It follows that $d_i^{\mathrm{p}}\left(t_{i,k,\frac{1}{k}}(\tilde{s}_i, \theta), t_i^\theta\right) \to 0$ and $\tilde{\sigma}_i\left(t_{i,k,\frac{1}{k}}(\tilde{s}_i, \theta)\right) = \delta_{\tilde{s}_i}$. Since σ satisfies requirement (1) in Definition 2, we know that it must be the case that $\sigma_i\left(t_{i,k,\frac{1}{k}}(\tilde{s}_i, \theta)\right) = \delta_{s_i^\theta}$ 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 equilibrium actions possible 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 one-to-one 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 $\kappa_{t_{i,k,\varepsilon}(\tilde{s}_i)} \in \Delta\left(\Theta \times T_{-i,k,\varepsilon}\right)$

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

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 $t_{-i}^{s_i}$ has common belief about the state being s_i (and thereby plays $\tilde{\sigma}_{-i}\left(t_{-i}^{s_i}\right) = \tilde{s}_{-i}^{s_i}$). Since $\tilde{s}_i$ is a best response against λ_{-i} and the strict/ unique best response against $\left(s_i, \tilde{s}_{-i}^{s_i}\right)$ 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

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

it follows that $d_i^k\left(t_{i,k,\varepsilon}^k\left(\tilde{s}_i,\theta\right),\left(t_i^\theta\right)^k\right)<\varepsilon$. ■

As in the case of Section 3.1, our results so far characterizing truthful continuous implementation in the product topology are in terms of the game form g or the reduced normal form game $\tilde{g}$. 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, comparable to Assumption 1 imposed here. However, they strengthen the requirement to strict continuous implementation, and show that any f that is strictly continuously implementable is strictly Maskin monotonic. To be precise, a social choice function $f:\Theta\to A$ is strictly Maskin monotonic if, for every pair of states θ and θ',

(1) $f(\theta)=f(\theta')$ whenever

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

for all i and a;

or, equivalently,

(2) $f(\theta)\neq f(\theta')$ implies

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

for some i and a.

COROLLARY 6. *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 7. *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 6 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 8. *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). ■

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 7 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, the full domain assumption 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.

4.1. *An Ordinal Model*

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 assess how a mechanism fares with information perturbations. We model cardinal utilities 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

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

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 $u_i : A \times \theta \to [0,1]$.[11]

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 $\kappa_{t_i} \in \Delta(\Theta^* \times T_{-i})$ 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. Hereafter, 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.

4.2. *Results*

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 3.2 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

[11] 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.

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 9.

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) $R^\infty(\theta, \tilde{g}) = \left\{\tilde{s}^\theta\right\}$.

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.[12] However, the core 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

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

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 7, the following result requires a substantially weaker domain richness condition and requires neither $|A| \geq 3$ nor f be onto.

COROLLARY 9. *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. ■

5. 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.[13] 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 goal therefore is to implement the social choice function on all models the planner 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, 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, i.e. local 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 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).[14] 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)).

[13] 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).

[14] 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.

They do not consider the uniform-weak topology but do hint at similar results in one direction (see, e.g., Footnote 16 of their paper). Our results on the uniform-weak topology thus both strengthen their results, and also constitute a key intermediate step to our characterization in the product topology. Further, they do not consider the case where only ordinal preferences are common knowledge (i.e., Section 4 of the present paper). They do consider the more general case of Bayesian implementation, which we omit.

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 with the product topology is as demanding as implementation in rationalizability.

To argue that implementation in rationalizability (Theorem 5, Corollary 9) 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 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. Here continuity is with respect to small perturbations of the initial endowments, and is meant 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.[15] In this sense, considering only truthful continuous implementation is no more permissive than general robust implementation.

6. Conclusion

This paper revisited the question of when a social choice function is continuously implementable, i.e. partially implementable both when the model corresponds to the state of nature being common knowledge, and also when it is "close" to common knowledge. Specifically, we restrict attention to mechanisms which correspond to direct revelation mechanisms in the initial model, and study what we term truthful continuous implementation, i.e. continuous implementation by only such a mechanism. We show that the specific topology by which we formalize this notion of proximity is critical. When considering closeness between types in the uniform-weak topology, or other topologies that preserve higher order beliefs, continuous implementability roughly boils down to whether the social choice function can be implemented in Strict Nash Equilibrium when the state is common knowledge among agents. As we argue in Section 3.2, this is a very weak requirement. If however, we use the product topology, continuous implementation becomes extremely demanding: it must be implementable in the unique rationalizable action. By considering a model where only ordinal preferences over alternatives

[15] 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).

are common knowledge, we show that a very weak richness condition implies that only dictatorial social choice functions are continuously implementable. The restrictiveness of continuous implementation, thus, boils down to whether nearby types preserve higher order beliefs.

APPENDIX A. OMITTED PROOFS

A.1. *Proofs from Section 3.2*

PROOF OF COROLLARY 3. Consider the DRM g among agents for whom f is never pessimal defined as:

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

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 COROLLARY 4. (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. ■

A.2. *Proofs from Section 4*

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

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

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 $R_i^\infty\left(t_i, \mathcal{M}\right)$ the set of interim correlated rationalizable messages.[16] Observe that for every ordinal complete-information type $R_i^\infty\left(t_i^\theta, \mathcal{M}\right) \subset R_i^\infty\left(\theta, \mathcal{M}\right)$.

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 $\Theta \times \left([0,1]^{|A|}\right)^I$ 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^*.[17]

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 $\pi \equiv \mathrm{marg}_{U_i \times \tilde{S}_{-i}} \mu$.

[16] 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.

[17] 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.

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

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

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 $R_{-i}^k(t_{-i,n}, \tilde{g}) \subset R_{-i}^k(\theta, \tilde{g})$. 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

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

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

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

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 $(\tilde{s}_{-i}, \theta^*)$,

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

Since $\pi \equiv \mathrm{marg}_{U_i \times \tilde{S}_{-i}}\mu$, 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 3. *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 3. Formally, fix $\varepsilon \in (0,1)$ and we prove the claim by induction. First, the claim trivially holds for $k = 0$. Now we prove the claim for $k \geq 1$, assuming that it holds for $k-1$. 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

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

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

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

That is, with probability $(1-\varepsilon)$, type $t_{i,k,\varepsilon}\left(\tilde{s}_i,\theta\right)$ believes that the cardinal state and opponents' types are distributed according to that induced from π through $\varphi_i \times \eta_{-i}$ (where each $t_{-i,k-1,\varepsilon}\left(\tilde{s}_{-i},\theta\right)$ plays $\tilde{\sigma}_{-i}\left(t_{-i,k-1,\varepsilon}\left(\tilde{s}_{-i},\theta\right)\right) = \tilde{s}_{-i}$ 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 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 $\left(s_i, \tilde{s}_{-i}^{s_i}\right)$ 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

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

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

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 3. Further, by Lemma 3 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. ■

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.

HARSANYI, J. C. (1967): "Games with Incomplete Information Played by "Bayesian" Players, I-III. Part I. The Basic Model," *Management Science*, 14(3), 159–182.

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.

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," .