

Strategy elimination in games with interaction structures

Andreas Witzel Krzysztof R. Apt
Jonathan A. Zvesper

ILLC, Universiteit van Amsterdam, Netherlands
and CWI, Amsterdam, Netherlands

September 7, 2009

Abstract

We study games in the presence of an interaction structure, which allows players to communicate their preferences, assuming that each player initially only knows his own preferences. We study the outcomes of iterated elimination of strictly dominated strategies (IESDS) that can be obtained in any given state of communication.

We also give epistemic foundations for these “intermediate” IESDS outcomes. This involves firstly describing the knowledge that the players would have in any state of communication, using the framework from Apt et al. [4]. We then prove that when there is common knowledge of rationality, each intermediate outcome is entailed by the knowledge in the relevant state of communication.

1 Introduction

1.1 Background and motivation

There is a substantial amount of research within game theory on the implications of assumptions concerning players’ *knowledge* and *beliefs* [6]. In particular, Tan and Werlang [17] have shown that if payoffs are commonly known and all players are *rational* and commonly believe in each other’s rationality, they will only play strategies that survive iterated elimination of strictly dominated strategies (IESDS). In this context rationality means that one does not choose strictly dominated strategies.

Another line of research stresses the relevance of *locality* in strategic games. For example, in *graphical games* [15] the locality assumption is formalized by assuming a graph structure over the set of players and using payoff functions which depend only on the strategies of players’ neighbors.

In this paper we study a game-theoretic framework which combines *locality* and *interaction*. The locality assumption refers to the *information* about payoffs (or more generally, preferences), rather than to the payoffs themselves. In turn, interaction takes

place by means of *communication* within (possibly overlapping) groups of players. The framework is realized by incorporating the notion of a strategic game into the setting of interaction structures discussed in [4].

An *interaction structure* consists of (possibly overlapping) groups of players within which synchronous communication is possible. We assume that players' preferences are *not* commonly known. Instead, the initial information of each player only covers *his own* preferences, and the players can communicate this information only within the limits of the interaction structure.

More precisely, we make the following assumptions:

- the players initially know their own preferences;
- they are rational;
- they are part of an interaction structure and can communicate their own preferences within any group they belong to;
- communication is truthful and synchronous, as in [4];
- the players have no knowledge other than what follows from these assumptions, and this is common knowledge.

In this setting we then study the outcome of iterated elimination of strictly dominated strategies started in some intermediate state of communication, in particular in the state in which all communication permitted by the interaction structure has taken place. We use the results from our previous work [4] to prove that this outcome can be described by analyzing what the players know in the considered state.

It is important to note that we do *not* examine strategic or normative aspects of the *communication* here. So we do not allow players to lie and do not examine *why* they communicate or *what* they should communicate. Rather, we examine what happens *if* they do communicate, assuming that they are rational and have reasoning powers.

To justify this focus, we can think of a setting in which the strategic aspects of communication are not relevant. One possibility is when communication is not a deliberate act, but rather occurs through observing somebody's behavior. Such communication is certainly more difficult to manipulate and more laborious to fake than mere words. In a sense it is inherently credible, and research in social learning argues along similar lines [9, Ch. 3].

This also helps to explain another assumption we make, corresponding to the framework we examined in [4]: players only communicate their *own* preferences, since information about *others'* preferences is either difficult to obtain or communication about them is not credible. One may also assume that communicating about preferences of third parties is less common for *privacy* reasons. From this perspective the groups of the interaction structure can be viewed as the ones who can commonly observe each other, for example colleagues sharing lunch at work.

In other settings, for example that of artificial agents communicating by means of messages, it may be more difficult to view communication as something non-deliberate. Here, ignoring strategic aspects of communication can be interpreted as bounds on the

players' rationality or reasoning capabilities—they simply lack the capabilities to deal with all the consequences of such an inherently rich phenomenon as communication.

In general, strategic communication is a research topic on its own, with controversial discussions (see, e.g., [16]) and many questions widely open. Crawford and Sobel [11] have considered the topic in a probabilistic setting, and Farrell and Rabin [13] have looked at related issues under the notion of *cheap talk*. Also within epistemic logic, formalizations of the information content of strategic communication have been suggested, e.g., by Gerbrandy [14].

Finally, it is useful to clarify the relation between strategic games with interaction structures and *pre-Bayesian games*, see, e.g., Ashlagi et al. [5]. In these games, too, each player knows his payoff but does not know the payoffs of the other players and makes no assumptions about them. In our setup this private knowledge aspect of pre-Bayesian games can be trivially modelled by the empty interaction structure, or viewed as corresponding to our initial situation. Due to the different nature of these frameworks, however, the questions of interest are also different.

1.2 Plan of the paper

This paper is organized as follows. In the following Section 2, we review the basic definitions concerning strategic games, optimality notions and operators on restrictions of games. Next, in Section 3, we study the outcome of IESDS in the presence of an interaction structure. We first look at the outcome that is arrived at after all communication permitted in the given interaction structure has taken place, and then detail the outcomes obtained in any particular intermediate state of communication. The formulations we consider make no direct use of the notion of knowledge. The connection with knowledge is made in Section 4, where we prove the outcomes we have obtained to be correct with respect to the epistemic framework from [4], in the sense that the outcomes capture exactly what the players can do given their partial knowledge of the game structure in any particular state. Finally, in Section 5, we suggest some future research directions.

2 Preliminaries

Following [3], by a **strategic game with parametrized preferences** (in short, a **game**) for players $N = \{1, \dots, n\}$, where $n > 1$, we mean a tuple $(S_1, \dots, S_n, \succ_1, \dots, \succ_n)$, where for each $i \in N$,

- S_i is the non-empty, finite set of **strategies** available to player i . We write S to abbreviate the set of **strategy profiles**: $S = S_1 \times \dots \times S_n$.
- \succ_i is the strict **preference relation** for player i , so $\succ_i \subseteq S \times S$.

This qualitative approach precludes the use of mixed strategies, but they will not be needed in our considerations.

As usual we denote player i 's strategy in a strategy profile $s \in S$ by s_i , and the tuple consisting of all other strategies by s_{-i} , i.e., $s_{-i} = (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_n)$.

Similarly, we use S_{-i} to denote $S_1 \times \dots \times S_{i-1} \times S_{i+1} \times \dots \times S_n$, and for $s'_i \in S_i$ and $s_{-i} \in S_{-i}$ we write (s'_i, s_{-i}) to denote $(s_1, \dots, s_{i-1}, s'_i, s_{i+1}, \dots, s_n)$. Finally, we use $s'_i \succ_{s_{-i}} s_i$ as a notational alternative for $(s'_i, s_{-i}) \succ_i (s_i, s_{-i})$.

Fix now an *initial* strategic game $\mathcal{G} := (S_1, \dots, S_n, \succ_1, \dots, \succ_n)$. We say that (S'_1, \dots, S'_n) is a **restriction** of \mathcal{G} if each S'_i is a non-empty subset of S_i . We identify the restriction (S_1, \dots, S_n) with \mathcal{G} .

To analyze iterated elimination of strategies from the initial game \mathcal{G} , we view such procedures as operators on the set of restrictions of \mathcal{G} . This set together with component-wise set inclusion forms a complete lattice.

For any restriction $\mathcal{G}' := (S'_1, \dots, S'_n)$ of \mathcal{G} and strategies $s_i, s'_i \in S_i$, we say that s_i is **strictly dominated by s'_i on S'_{-i}** if $s'_i \succ_{s'_{-i}} s_i$ for all $s'_{-i} \in S'_{-i}$. Then we introduce the following abbreviations (ℓ stands for “local” and g stands for “global”; the terminology is from Apt [2]):

- $sd^\ell(s_i, \mathcal{G}')$ which holds iff strategy s_i of player i is not strictly dominated on S'_{-i} by any strategy from S'_i (i.e., $\neg \exists s'_i \in S'_i \forall s'_{-i} \in S'_{-i} s'_i \succ_{s'_{-i}} s_i$),
- $sd^g(s_i, \mathcal{G}')$ which holds iff strategy s_i of player i is not strictly dominated on S'_{-i} by any strategy from S_i (i.e., $\neg \exists s'_i \in S_i \forall s'_{-i} \in S'_{-i} s'_i \succ_{s'_{-i}} s_i$).

So in sd^g , the global version of strict dominance introduced by Chen et al. [10], it is stipulated that a strategy is not strictly dominated by a strategy *from the initial game*.

We call each relation of the form sd^ℓ or sd^g an **optimality notion**. We say then that the optimality notion ϕ used by player i is **monotonic** if for all restrictions \mathcal{G}'' and \mathcal{G}' and strategies $s_i, \mathcal{G}'' \subseteq \mathcal{G}'$ and $\phi(s_i, \mathcal{G}'')$ implies $\phi(s_i, \mathcal{G}')$.

As noted in [8, 2], sd^g is monotonic, while sd^ℓ is not (though in finite games their respective outcomes coincide, as discussed in the proof of Theorem 3.2).

Given an operator T on a finite lattice (D, \subseteq) with the largest element \top and $k \geq 0$, we denote by T^k the k -fold iteration of T , where $T^0 = \top$ (so the iterations start “at the top”) and put $T^\infty := \bigcap_{k \geq 0} T^k$. We call T **monotonic** if for all D', D'' , we have that $D' \subseteq D''$ implies $T(D') \subseteq T(D'')$.

Finally, as in [4], an **interaction structure** H is a *hypergraph* on N , i.e., a set of non-empty subsets of $A \subseteq N$, called *hyperarcs*.

3 Iterated strategy elimination

In this section we define procedures for iterated elimination of strictly dominated strategies. Let us fix a strategic game $\mathcal{G} = (S_1, \dots, S_n, \succ_1, \dots, \succ_n)$ for players N , an interaction structure $H \subseteq 2^N \setminus \{\emptyset\}$, and an optimality notion ϕ . In Section 3.1, we look at the outcome reached after all communication permitted by H has taken place, that is, when within each hyperarc of H all of its members’ preferences have been communicated. In Section 3.2, we then look at the outcomes obtained in any particular intermediate state of communication. We stress that in general there is no relation between the preferences \succ_i and H .

The formulations we give here make no direct use of a formal notion of knowledge. The connection with a formal epistemic model is made in Section 4.

All iterations of the considered operators start at the initial restriction (S_1, \dots, S_n) .

3.1 Completed communication

Let us assume that within each hyperarc $A \in H$, all its members have shared all information about their preferences. We leave the exact definition of communication to Section 3.2 and the epistemic formalization to Section 4, and focus here on an operational description.

For each group of players $G \in N$, let S_G denote the set of those restrictions of \mathcal{G} which only restrict the strategy sets of players from G . That is,

$$S_G := \{(S'_1, \dots, S'_n) \mid S'_i \subseteq S_i \text{ for } i \in G \text{ and } S'_i = S_i \text{ for } i \notin G\}.$$

Now we introduce an elimination operator T_G on each such set S_G , defined as follows. For each $\mathcal{G}' = (S'_1, \dots, S'_n) \in S_G$, let $T_G(\mathcal{G}') := (S''_1, \dots, S''_n)$, where for all $i \in N$,

$$S''_i := \begin{cases} \{s_i \in S'_i \mid \phi(s_i, \mathcal{G}')\} & \text{if } i \in G \\ S'_i & \text{otherwise.} \end{cases}$$

We call T_G^∞ the **outcome of iterated elimination (of non- ϕ -optimal strategies) on G** . We then define the restriction $\mathcal{G}(H)$ of \mathcal{G} as¹ $\mathcal{G}(H) := (\mathcal{G}(H)_1, \dots, \mathcal{G}(H)_n)$, where for all $i \in N$,

$$\mathcal{G}(H)_i := T_{\{i\}} \left(\bigcap_{A: i \in A \in H} T_A^\infty \right)_i.$$

That is, the i th component of $\mathcal{G}(H)$ is the i th component of the result of applying $T_{\{i\}}$ to the intersection of T_A^∞ for all $A \in H$ containing i . We call $\mathcal{G}(H)$ the **outcome of iterated elimination (of non- ϕ -optimal strategies) with respect to H** . Note that $\mathcal{G}(H)$ implicitly depends on ϕ .

Let us “walk through” this definition to understand it better. Given a player i and a hyperarc $A \in H$ such that $i \in A$, T_A^∞ is the outcome of iterated elimination on A , starting at (S_1, \dots, S_n) . The strategies of players from outside of A are not affected by this process. This elimination process is performed simultaneously for each hyperarc that i is a member of. By intersecting the outcomes, i.e., by considering the restriction $\bigcap_{A: i \in A \in H} T_A^\infty$, one arrives at a restriction in which all such “groupwise” iterated eliminations have taken place. However, in this restriction some of the strategies of player i may be non- ϕ -optimal. They are eliminated using one application of the $T_{\{i\}}$ operator. We illustrate this process, and in particular this last step, in the following example.

Example 3.1. Consider local strict dominance, sd^ℓ , in the following three-player game \mathcal{G} where the payoffs of players 1 and 2 and those of players 1 and 3 respectively depend on each other’s actions, but the payoffs of player 2 and 3 are independent:

		Pl. 2, 3			
		L, l	L, r	R, l	R, r
Pl. 1	U	1, 1, 1	0, 1, 0	0, 0, 1	0, 0, 0
	D	0, 1, 1	1, 1, 0	1, 0, 1	1, 0, 0

¹Here and elsewhere the outer subscript ‘ i ’ refers to the preceding restriction.

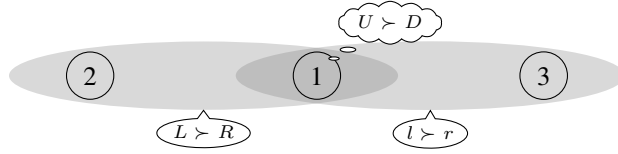


Figure 1: Illustrating Example 3.1. Hyperarcs are shown in gray. Callouts attached to hyperarcs represent communicated, and thus commonly known, information. The thought bubble represents private information, in this case obtained from the combination of information only available to player 1.

So, for example, the payoffs for the strategy profile (U, L, r) are, respectively, 0, 1, and 0. Now assume the interaction structure $H = \{\{1, 2\}, \{1, 3\}\}$. We obtain $T_{\{1,2\}}^\infty = (\{U, D\}, \{L\}, \{l, r\})$ and $T_{\{1,3\}}^\infty = (\{U, D\}, \{L, R\}, \{l\})$. The restriction defined by these two outcomes is $(\{U, D\}, \{L\}, \{l\})$, and in the final step player 1 eliminates his strategy D by one application of $T_{\{1\}}$. The outcome of the whole process is thus $\mathcal{G}(H) = (\{U\}, \{L\}, \{l\})$. See Figure 1 for an illustration of this situation. \square

In this example, the outcome with respect to the given interaction structure coincides with the outcome of the customary IESDS on the fully specified game matrix. We should emphasize that this is not the case in general, and the purpose of this example is simply to illustrate how the operators work. Example 3.3 later on shows in a different setting how the interaction structure can influence the outcome.

Note that when H consists of the single hyperarc N that contains all the players, then for each player i , $\bigcap_{A:i \in A \in H} T_A^\infty$ reduces to T_N^∞ , and this is closed under application of each operator $T_{\{i\}}$. So then, indeed, $\mathcal{G}(H) = T_N^\infty$, that is, $\mathcal{G}(H)$ in this special case coincides with the customary outcome of iterated elimination of non- ϕ -optimal strategies.

In general, this customary outcome is included in the outcome w.r.t. any hypergraph H . This result is established in Theorem 3.2, and Example 3.3 shows a case where the inclusion is proper.

Theorem 3.2. For $\phi \in \{sd^l, sd^g\}$ and for all hypergraphs H , we have $T_N^\infty \subseteq \mathcal{G}(H)$.

The inclusion proved in this result cannot be reversed, even when each pair of players shares a hyperarc. The following example also shows that the hypergraph structure is more informative than the corresponding graph structure.

Example 3.3. Consider the following strategic game with three players. The payoffs of player 1 and 2 depend here only on each other's choices, and the payoffs of player 3 depend only on the choices of player 2 and 3:

		Pl. 2	
		L	R
Pl. 1	U	0, 1	0, 0
	D	1, 0	1, 1

Payoff of players 1 and 2

		Pl. 2	
		L	R
Pl. 3	A	0	1
	B	1	0

Payoff of player 3

So, for example, the payoffs for the strategy profile (U, L, A) are, respectively, 0, 1, and 0. If we assume the hypergraph H that consists of the single hyperarc $\{1, 2, 3\}$, then the outcome of iterated elimination of non- ϕ -optimal strategies w.r.t. H is the customary outcome which equals $(\{D\}, \{R\}, \{A\})$. Indeed, player 1 can eliminate his strictly dominated strategy U , then player 2 can eliminate L , and subsequently player 3 can eliminate B .

In contrast, if the hypergraph consists of all pairs of players, so $H = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\}$, then the outcome of iterated elimination of non- ϕ -optimal strategies w.r.t. H equals $(\{D\}, \{R\}, \{A, B\})$.

Informally, the reason for this difference is that in the latter case, player 3 can eliminate B only using the fact that player 2 eliminated L , but this information is available only to players 1 and 2. \square

To familiarize ourselves further with our definitions, we establish the following intuitive monotonicity result. We say that H' extends H if for each $A \in H$ there is $A' \in H'$ such that $A \subseteq A'$.

Proposition 3.4. *If H' extends H and T is monotonic, then $\mathcal{G}(H') \subseteq \mathcal{G}(H)$.*

3.2 Intermediate states

The setting considered in Section 3.1 corresponds to a state in which in all hyperarcs all players have shared all information about their preferences. Given the game \mathcal{G} and the hypergraph H , the outcome $\mathcal{G}(H)$ there defined thus reflects which strategies players can eliminate if initially they know only their own preferences and they communicate all their preferences in H . We now define formally what communication we assume possible, and then look at intermediate states, where only certain preferences have been communicated.

Each player i can communicate his preferences to each $A \in H$ with $i \in A$. We take a **message** by i to consist of a preference statement $s'_i \succ_{s_{-i}} s_i$ for $s_i, s'_i \in S_i$ and $s_{-i} \in S_{-i}$. We denote such a message by $(i, A, s'_i \succ_{s_{-i}} s_i)$, and require that $i \in A$ and that it is **truthful** with respect to the given initial game \mathcal{G} , that is, indeed $s'_i \succ_{s_{-i}} s_i$ in \mathcal{G} . Note that the fact that i is the sender is, strictly speaking, never used. Thus, in accordance with the interpretation of communication described in Section 1.1, we could drop the sender and simply write “the players in A commonly observe that $s'_i \succ_{s_{-i}} s_i$.” An **intermediate state** is now given by the set M of messages which have been communicated.

We now adjust the definition of an optimality notion to account for intermediate states. An **intermediate optimality notion** $\phi_{G,M}$ (derived from an optimality notion ϕ) uses only information shared among the group G in the intermediate state given by M . So with singleton $G = \{i\}$ only i 's preferences are used, and with larger G only preferences contained in messages to a superset of G are used. Thus in the case of sd^g we have that $sd^g_{G,M}(s_i, \mathcal{G}')$ holds iff

$$\begin{aligned} \neg \exists s'_i \in S_i \forall s_{-i} \in S'_{-i} s'_i \succ_{s_{-i}} s_i & \quad \text{if } G = \{i\} \\ \neg \exists s'_i \in S_i \forall s_{-i} \in S'_{-i} M \upharpoonright_G \models s'_i \succ_{s_{-i}} s_i & \quad \text{otherwise,} \end{aligned}$$

where by $M \uparrow_{G \neq} s'_i \succ_{s_{-i}} s_i$ we mean that $s'_i \succ_{s_{-i}} s_i$ is entailed by those messages in M which G received. Specifically, the **entailment relation**

$$M \uparrow_{G \neq} s'_i \succ_{s_{-i}} s_i$$

holds iff there exist messages $(\cdot, G^k, s_i^k \succ_{s_{-i}} s_i^{k+1}) \in M$ for $k \in \{1, \dots, \ell - 1\}$ such that $G^k \supseteq G$, $s_i^1 = s'_i$ and $s_i^\ell = s_i$.

We now define a generalization of the T_G operator by:

$$T_{G,M}(\mathcal{G}') := (S''_1, \dots, S''_n),$$

where $\mathcal{G}' = (S'_1, \dots, S'_n)$ and for all $i \in N$,

$$S''_i := \{s_i \in S'_i \mid \phi_{G,M}(s_i, \mathcal{G}')\}.$$

Note that, as before, S'_i remains unchanged for $i \notin G$, since then $\phi_{G,M}(s_i, \mathcal{G}')$ always holds. Indeed, for it to be false, there would have to be some message $(i, G, \cdot) \in M$, which would imply $i \in G$.

Similarly, we now define the **outcome of iterated elimination (of non- ϕ -optimal strategies) with respect to H, M** to be the restriction $\mathcal{G}(H, M)$, where for $i \in N$

$$\mathcal{G}(H, M)_i := T_{\{i\}, M} \left(\bigcap_{A: i \in A \in \overline{H}} T_{A, M}^\infty \right)_i.$$

Here \overline{H} denotes the closure of H under non-empty intersection. That is,

$$\overline{H} = \{A_1 \cap \dots \cap A_k \mid \{A_1, \dots, A_k\} \subseteq H\} \setminus \{\emptyset\}.$$

The use of \overline{H} is necessary because certain information may be entailed by messages sent to different hyperarcs. For example, with $(j, A, s'_j \succ_{s_{-j}} s'_j), (j, A', s'_j \succ_{s_{-j}} s_j) \in M$, the combined information that $s'_j \succ_{s_{-j}} s_j$ is available to the members of $A \cap A'$.

Again, let us “walk through” the definition of $\mathcal{G}(H, M)$. First, a local elimination process is run on each hyperarc of \overline{H} , using only information which has been communicated there (which now no longer covers all members’ preferences, but only the ones according to the intermediate state M). Then, in the final step, each player combines his insights from all hyperarcs of which he is a member, and eliminates any strategies that he thereby learns not to be optimal.

It is easy to see that in the case where the players have communicated all there is to communicate, i.e., for

$$M_H^{\text{all}} := \{(i, A, s'_i \succ_{s_{-i}} s_i) \mid i \in N, A \in H, s_i, s'_i \in S_i \text{ with } s'_i \succ_{s_{-i}} s_i \text{ in } \mathcal{G}\},$$

the intermediate outcome coincides with the previously defined outcome, i.e.,

$$\mathcal{G}(H, M_H^{\text{all}}) = \mathcal{G}(H).$$

This corresponds to the intuition that $\mathcal{G}(H)$ captures the elimination process when all possible communication has taken place. In particular, all entailed information has also been communicated in M_H^{all} , which is why we did not need to consider \overline{H} in Section 3.1.

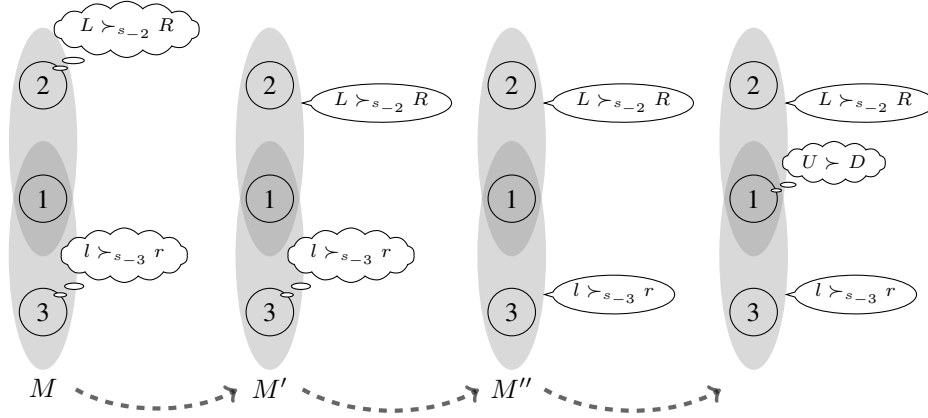


Figure 2: Illustrating Example 3.5.

Example 3.5. The process described in this example is illustrated in Figure 2. Consider again the game \mathcal{G} from Example 3.1, and the initial state where $M = \emptyset$. We have $T_{A,M}^\infty = \mathcal{G}$ for all $A \in \overline{H}$, that is, without communication no strategy can “commonly” be eliminated. However, players 2 and 3 can “privately” eliminate one of their strategies each, since each of them knows his own preferences, so

$$\begin{aligned} T_{\{1\},M}(\bigcap_{A:1 \in A \in \overline{H}} T_{A,M}^\infty) &= (\{U, D\}, \{L, R\}, \{l, r\}), \\ T_{\{2\},M}(\bigcap_{A:2 \in A \in \overline{H}} T_{A,M}^\infty) &= (\{U, D\}, \{L\}, \{l, r\}), \\ T_{\{3\},M}(\bigcap_{A:3 \in A \in \overline{H}} T_{A,M}^\infty) &= (\{U, D\}, \{L, R\}, \{l\}), \end{aligned}$$

This elimination cannot be iterated upon by other players and the overall outcome is $\mathcal{G}(H, M) = (\{U, D\}, \{L\}, \{l\})$.

Consider now the intermediate state $M' = \{(2, \{1, 2\}, L \succ_{s-2} R) \mid s_{-2} \in S_{-2}\}$, that is, a state in which player 2 has shared with player 1 the information that for any joint strategy of players 1 and 3, he prefers his strategy L over R . Then only the result of player 1 changes:

$$T_{\{1\},M'}(\bigcap_{A:1 \in A \in \overline{H}} T_{A,M'}^\infty) = (\{U, D\}, \{L\}, \{l, r\}),$$

while the other results and the overall outcome remain the same. If additionally player 3 communicates all his information in the hyperarc he shares with player 1, that is, if the intermediate state is $M'' = M' \cup \{(3, \{1, 3\}, l \succ_{s-3} r) \mid s_{-3} \in S_{-3}\}$, then player 1 can combine all the received information and obtain

$$T_{\{1\},M''}(\bigcap_{A:1 \in A \in \overline{H}} T_{A,M''}^\infty) = (\{U\}, \{L\}, \{l\}).$$

This is also the overall outcome $\mathcal{G}(H, M'')$, coinciding with the outcome $\mathcal{G}(H, M_H^{\text{all}})$ where all possible information has been communicated. \square

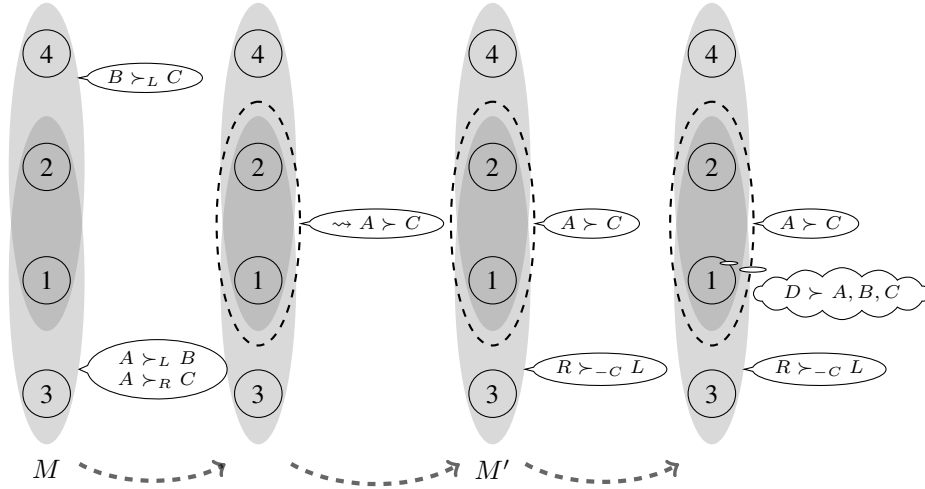


Figure 3: Illustrating Example 3.6. Strategies of the dummy players are omitted. $A \succ C$ stands for $A \succ_{s-1} C$, and \succ_{-C} combines \succ_α for $\alpha \in \{A, B, D\}$. Note that in the first step, information is not explicitly communicated but deduced.

Let us now illustrate the importance of using entailment in the intermediate optimality notions and \overline{H} (rather than H) in the definition of $\mathcal{G}(H, M)$.

Example 3.6. We look at a game involving four players, but we are only interested in the preferences of two of them. The other two players serve merely to create different hyperarcs. The strategies and payoffs of player 1 and 2 are as follows:

		Pl. 2	
		L	R
Pl. 1	A	3, 0	1, 1
	B	2, 0	1, 1
	C	1, 1	0, 0
	D	0, 0	5, 1

For players 3 and 4 we assume a “dummy” strategy, denoted respectively by X and Y . Consider the hypergraph $H = \{\{1, 2, 3\}, \{1, 2, 4\}\}$ and the intermediate state

$$M = \{(1, \{1, 2, 3\}, A \succ_{LXY} B), \\ (1, \{1, 2, 4\}, B \succ_{LXY} C), \\ (1, \{1, 2, 3\}, A \succ_{RXY} C)\}.$$

The fact that player 1, independently of what the remaining players do, strictly prefers A over C is not explicit in these pieces of information, but it is *entailed* by them, since $A \succ_{LXY} B$ and $B \succ_{LXY} C$ imply $A \succ_{LXY} C$. However, this combination of information is only available to the players in $\{1, 2, 3\} \cap \{1, 2, 4\}$.

Player 2 can make use of this fact that C is dominated, and eliminate his own strategy L . If we now look at a state in which player 2 has communicated his relevant preferences, so $M' = M \cup \{(2, \{1, 2, 3\}, R \succ_{\alpha XY} L) \mid \alpha \in \{A, B, D\}\}$, we notice that player 1 can in turn eliminate A and B , but only by combining information available to the players in $\{1, 2, 3\} \cap \{1, 2, 4\}$. There is no single hyperarc in the original hypergraph which has all the required information available. It thus becomes clear that we need to take into account iterated elimination on intersections of hyperarcs.

The whole process is illustrated in Figure 3. \square

4 Epistemic foundations

In this section, we provide epistemic foundations for our framework. The aim is to prove that the definition of the outcome $\mathcal{G}(H, M)$ correctly captures what strategies the players can eliminate using all they “know”, in a formal sense.

We proceed as follows. First, in Section 4.1, we briefly introduce an epistemic model formalizing the players’ knowledge. In Section 4.2, we give a general epistemic formulation of strict dominance and argue that it correctly captures the notion. Section 4.2 also contains the main result of our epistemic analysis, namely that the outcome $\mathcal{G}(H, M)$ indeed yields the outcome stipulated by the epistemic formulation. We rely on the basic framework and results from [4], which we recall in Appendix A.

We focus on the global version of strict dominance, sd^g , mainly because the presentation is then more concise. However, our results carry over to the local version sd^l due to the equivalence result mentioned in the proof of Theorem 3.2.

4.1 Epistemic language and states

Again, we assume a fixed game \mathcal{G} with non-empty set of strategies S_i for each player i , and a hypergraph H representing the interaction structure. Analogously to [4], we use a propositional **epistemic language** with a set At of **atoms** which is divided into disjoint subsets At_i , one for each player i , where $\text{At}_i = \{s'_i \succ_{s_{-i}} s_i \mid s_i, s'_i \in S_i, s_{-i} \in S_{-i}\}$.

The set At_i describes all possible strict preferences between pairs of strategies of player i , relative to a joint strategy of the opponents. We consider the usual **connectives** \wedge and \vee (but not the negation \neg), and a **common knowledge** operator C_G for any group $G \subseteq N$ of players. As in [4], we write K_i for $C_{\{i\}}$. By \mathcal{L}^+ we denote the set of formulas built from the atoms in At using these two connectives and knowledge operators.

A **valuation** V is a subset of At such that for each $s_{-i} \in S_{-i}$, the restriction $V \cap \{\cdot \succ_{s_{-i}} \cdot\}$ is a strict partial order.

Intuitively, a valuation consists of the atoms assumed true. Each specific game \mathcal{G} *induces* exactly one valuation which simply represents its preferences. However, in general we also need to model the fact that players may not have full knowledge of the game. The restriction imposed on the valuations ensures that each of them is induced by some game.

So for example $\{s \succ_a t\}$ is a valuation (given a game with appropriate strategy sets), while $\{s \succ_a t, t \succ_a u\}$ and $\{s \succ_a t, t \succ_a s\}$ are not.

Recall from Section 3.2 that a **message** from player i to a hyperarc $A \in H$ has the form $(i, A, s'_i \succ_{s_{-i}} s_i)$, where $i \in A$, $s_i, s'_i \in S_i$, and $s_{-i} \in S_{-i}$. We say that a message (\cdot, \cdot, p) is **truthful** with respect to a valuation V if $p \in V$. A **state**, or **possible world**, is a pair (V, M) , where V is a valuation and M is a set of messages that are truthful with respect to V .

This setting is an instance of the framework defined in [4], and the formal **semantics** is as defined there (see Appendix A for a brief summary). We repeat here only the intuition that $C_G\varphi$ means that φ is *common knowledge* among G , that is, everybody in G knows φ , everybody knows that everybody knows φ , etc. In particular, $K_i\varphi$ means that player i *knows* φ . We assume that each player i initially knows the true facts in At_i entailed by the initial game \mathcal{G} and that the basic assumptions from Section 1.1 are commonly known among the players.

4.2 Correctness result

We use results from [4], summed up in Appendix A, in order to prove that the T_G operator defined in Section 3 is correct with respect to an epistemic formulation of our setting.

We start by giving an epistemic formula describing the global version of iterated elimination of strictly dominated strategies. In contrast to the formulation in Section 2, this formula states *player i knows that a strategy is strictly dominated*.

We define, for $i \in N$ and $s_i \in S_i$,

$$\begin{aligned} \text{dom}^1(s_i) &:= K_i \bigvee_{s'_i \in S_i} \bigwedge_{s_{-i} \in S_{-i}} s'_i \succ_{s_{-i}} s_i, \\ \text{dom}^{\ell+1}(s_i) &:= K_i \bigvee_{s'_i \in S_i} \bigwedge_{s_{-i} \in S_{-i}} (s'_i \succ_{s_{-i}} s_i \vee \bigvee_{j \in N \setminus \{i\}} \text{dom}^\ell(s_j)). \end{aligned}$$

That is, in the base case, player i knows that s_i is strictly dominated if i knows that there is an alternative strategy s'_i which, for all joint strategies of the other players, is strictly preferred. Furthermore, after iteration $\ell+1$, i knows that s_i is strictly dominated if i knows that there is an alternative strategy s'_i such that, for all joint strategies s_{-i} of the other players, either s'_i is strictly preferred or some strategy s_j in s_{-i} is already known by player j to be strictly dominated after iteration ℓ .

We restrict our attention to formulas $\text{dom}^\ell(s_i)$ with $\ell \in \{1, \dots, \hat{\ell}\}$, where $\hat{\ell} = \sum_{i \in N} |S_i|$. By the semantics of the considered formulas, there is some ℓ within this range such that for all $\ell' \geq \ell$, $\text{dom}^{\ell'}$ is equivalent to dom^ℓ . To reflect the fact that this can be seen as the outcome of the iteration, we denote $\text{dom}^{\hat{\ell}}$ by dom^∞ .

As a first connection with the T_G operator defined in Section 3, we have the following epistemic counterpart of Proposition 3.4. Intuitively, this is due to the fact that if we look at the states in which all communication allowed by a given hypergraph has taken place, then knowledge (of positive formulas) can only grow as that hypergraph grows.

Proposition 4.1. *If H' extends H , then for all $i \in N$ and $s_i \in S_i$,*

$$(V, M_H^{\text{all}}) \models \text{dom}^\infty(s_i) \text{ implies } (V, M_{H'}^{\text{all}}) \models \text{dom}^\infty(s_i),$$

where M^{all} is defined as in Section 3.2.

Proof. The claim follows from Lemma A.2 and the fact that $\text{dom}^\infty(s_i) \in \mathcal{L}^+$. \square

We now proceed to the main result of the paper. We prove that the non-epistemic formulation of iterated elimination of non- sd^g -optimal strategies, as given in Section 3, coincides with the epistemic formulation of strict dominance.

Theorem 4.2. *For any strategic game \mathcal{G} , hypergraph H , set of messages M truthful with respect to \mathcal{G} , and $i \in N$,*

$$\mathcal{G}(H, M)_i = \{s_i \in S_i \mid (V, M) \not\equiv \text{dom}^\infty(s_i)\},$$

where V is the valuation induced by \mathcal{G} .

5 Conclusions

We studied here strategic games in the presence of interaction structures. We assumed that initially the players know only their own preferences, and that they can truthfully communicate information about their own preferences within their parts of the interaction structure. This allowed us to analyze the consequences of locality, formalized by means of an interaction structure, on the outcome of the iterated elimination of strictly dominated strategies. To this end we appropriately adapted the framework introduced in [4] and showed that in any given state of communication this outcome can be described by means of epistemic analysis.

We plan to extend our analysis in a number of ways by:

- Allowing players to send information about the preferences of other players that they learned through interaction. The abstract epistemic framework of [4] includes already this extension,
- Allowing other forms of messages, for example, messages containing information that a strategy has been eliminated, or containing epistemic statements, such as knowing that some strategy of *another player* has been eliminated,
- Considering strategic aspects of communication, even if truthfulness is required (should one send some piece of information or not?)
- Considering formation or evolution of interaction structures, given strategic advantages of certain interaction structures over others.

The last point could connect our research with that on network formation games, see, e.g., [7].

Finally, let us mention that in [4] we already abstracted from the framework considered here and studied a setting in which players send messages that inform a group about some atomic fact that a player knows or has learned. We clarified there, among others, under what conditions common knowledge of the underlying hypergraph matters. The framework there considered could be generalized by allowing players to

arrive jointly at some conclusions using their background theories, by means of an interaction through messages sent to groups. From this perspective IESDS could be seen as a metaphor of such a conclusion. Through its focus on the form of allowed messages and background knowledge, this study would differ from the line of research pursued by Fagin et al. [12], where the effects of communication are considered in the framework of distributed systems.

Acknowledgements

We thank Rohit Parikh and Willemien Kets for discussion and helpful suggestions. The first and third authors were supported by a GLoRiClass fellowship funded by the European Commission (Early Stage Research Training Mono-Host Fellowship MEST-CT-2005-020841).

References

- [1] Krzysztof R. Apt. Relative strength of strategy elimination procedures. *Economics Bulletin*, 3(21):1–9, 2007.
- [2] Krzysztof R. Apt. The many faces of rationalizability. *The B.E. Journal of Theoretical Economics*, 7(1):Article 18, 2007.
- [3] Krzysztof R. Apt, Francesca Rossi, and Kristen Brent Venable. Comparing the notions of optimality in CP-nets, strategic games and soft constraints. *Annals of Mathematics and Artificial Intelligence*, 52(1):25–54, 2008.
- [4] Krzysztof R. Apt, Andreas Witzel, and Jonathan A. Zvesper. Common knowledge in interaction structures. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge (TARK XII)*, 2009.
- [5] Itai Ashlagi, Dov Monderer, and Moshe Tennenholtz. Resource selection games with unknown number of players. In Hideyuki Nakashima, Michael P. Wellman, Gerhard Weiss, and Peter Stone, editors, *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 819–825, Hakodate, Japan, 2006. ACM Press.
- [6] Pierpaolo Battigalli and Giacomo Bonanno. Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics*, 53(2): 149–225, 1999.
- [7] Francis Bloch and Matthew Jackson. Definitions of equilibrium in network formation games. *International Journal of Game Theory*, 34(3):305–318, 2006.
- [8] Adam Brandenburger, Amanda Friedenberg, and H. Jerome Keisler. Fixed points for strong and weak dominance, 2006. Working paper. Available from <http://pages.stern.nyu.edu/~abranden/>.

- [9] Christophe P. Chamley. *Rational herds: Economic models of social learning*. Cambridge University Press, 2004.
- [10] Yi-Chun Chen, Ngo Van Long, and Xiao Luo. Iterated strict dominance in general games. *Games and Economic Behavior*, 61(2):299–315, 2007.
- [11] Vincent P. Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50(6):1431–1451, 1982.
- [12] Ronald Fagin, Joseph Y. Halpern, Moshe Y. Vardi, and Yoram Moses. *Reasoning about knowledge*. MIT Press, 1995.
- [13] Joseph Farrell and Matthew Rabin. Cheap talk. *The Journal of Economic Perspectives*, 10(3):103–118, 1996.
- [14] Jelle Gerbrandy. Communication strategies in games. *Journal of Applied Non-Classical Logics*, 17(2):197–211, 2007.
- [15] Michael Kearns, Michael L. Littman, and Satinder Singh. Graphical models for game theory. In Jack S. Breese and Daphne Koller, editors, *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*, pages 253–260, Seattle, Washington, 2001. Morgan Kaufmann.
- [16] David Sally. Can I say ”bobobo” and mean ”There’s no such thing as cheap talk”? *Journal of Economic Behavior & Organization*, 57(3):245–266, 2005.
- [17] Tommy Chin-Chiu Tan and Sérgio Ribeiro da Costa Werlang. The bayesian foundations of solution concepts of games. *Journal of Economic Theory*, 45(2):370–391, 1988.
- [18] Andreas Witzel. *Knowledge and Games: Theory and Implementation*. PhD thesis, University of Amsterdam, 2009. ILLC Dissertation Series 2009-05.

A Results used from Apt et al. [4]

As described in Section 4.1, given some game we consider the **basic propositions (atoms)** At to consist of disjoint subsets At_i , one for each player i . For each $s_i, s'_i \in S_i$, and $s_{-i} \in S_{-i}$, At_i contains one atom $s'_i \succ_{s_{-i}} s_i$. A **valuation** is a subset of At , consisting of those atoms that are *true*. We denote valuations by V and require that for each i and each $s_{-i} \in S_{-i}$, the restriction $V \cap \{\cdot \succ_{s_{-i}} \cdot\}$ represents a strict partial order (so it may indeed be the preference order induced by some game).²

A **message** has the form $(i, A, s'_i \succ_{s_{-i}} s_i)$, where $i \in A \in H$, $s_i, s'_i \in S_i$, and $s_{-i} \in S_{-i}$.

A **state** is a tuple (V, M) where V is a valuation and M is a set of truthful messages (i, A, p) , that is, indeed $p \in V$.

² In [4] we did not consider such restrictions on valuations; however, the relevant results can easily be seen to remain correct. See also [18, Chapter 2].

For a set of messages M , $G \subseteq N$, and $p \in \text{At}$, $M \upharpoonright_G \models p$ is defined as in Section 3.2. That is, $M \upharpoonright_G \models p$ means that p is entailed by the messages in M received by G , for example, by transitivity of the represented preference order.

In [4], we defined set operations to act component-wise on states, e.g. $(V, M) \subseteq (V', M')$ iff $V \subseteq V'$ and $M \subseteq M'$. However, the results we consider also hold with a modified inclusion relation, where $M \subseteq M'$ iff for each $(i, A, p) \in M$ there is $(i, A', p) \in M'$ with $A \subseteq A'$.

We define an **indistinguishability relation** between states:

$$(V, M) \sim_i (V', M') \text{ iff } (V_i, M_i) = (V'_i, M'_i).$$

For $G \subseteq N$ the relation \sim_G is the transitive closure of $\bigcup_{i \in G} \sim_i$.

We consider the following positive epistemic **language** \mathcal{L}^+ :

$$\varphi ::= p \mid \varphi \wedge \psi \mid \varphi \vee \psi \mid C_G \varphi,$$

where the atoms p denote the facts in At , \neg , \wedge and \vee are the standard connectives; and C_G is a knowledge operator, with $C_G \varphi$ meaning φ is common knowledge among G . We write K_i for $C_{\{i\}}$; $K_i \varphi$ can be read ‘ i knows that φ ’. For a sequence of players $w = i_1 \dots i_k$, we write K_w to abbreviate $K_{i_1} K_{i_2} \dots K_{i_k}$.

The **semantics** is as follows:

Definition A.1.

$$\begin{aligned} (V, M) \models p & \quad \text{iff } p \in V, \\ (V, M) \models \varphi \vee \psi & \quad \text{iff } (V, M) \models \varphi \text{ or } (V, M) \models \psi, \\ (V, M) \models \varphi \wedge \psi & \quad \text{iff } (V, M) \models \varphi \text{ and } (V, M) \models \psi, \\ (V, M) \models C_G \varphi & \quad \text{iff } (V', M') \models \varphi \text{ for each } (V', M') \\ & \quad \text{with } (V, M) \sim_G (V', M'). \end{aligned}$$

Now we are ready to state the following results from [4], slightly adapted to fit our notation.

Lemma A.2 (from [4, Lemma 3.2]). *For any $\varphi \in \mathcal{L}^+$ and states (V, M) and (V', M') with $(V, M) \subseteq (V', M')$,*

$$\text{if } (V, M) \models \varphi, \text{ then } (V', M') \models \varphi.$$

Theorem A.3 (from [4, Theorem 3.5]). *For any $\varphi_1, \varphi_2 \in \mathcal{L}^+$, state (V, M) , and $G \subseteq N$,*

$$(V, M) \models C_G(\varphi_1 \vee \varphi_2) \text{ iff } (V, M) \models C_G \varphi_1 \vee C_G \varphi_2.$$

Lemma A.4 (from [18, Lemma 2.3.8], cf. [4, Lemma 3.7]). *For any $G \subseteq N$ with $|G| \geq 2$, $p \in \text{At}$, and state (V, M) , the following are equivalent:*

- (i) $M \upharpoonright_G \models p$,
- (ii) $(V, M) \models C_G p$

Theorem A.5 (from [4, Theorem 3.8]). *For any $G \subseteq N$, $\varphi \in \mathcal{L}^+$, and state (V, M) ,*

$$(V, M) \models C_G \varphi \text{ iff } (V, M) \models K_w \varphi \text{ for some permutation } w \text{ of } G.$$

B Omitted Proofs

Proof of Theorem 3.2. First, consider $\phi = sd^g$, and $G \subseteq G' \subseteq N$. By definition, this implies that for all restrictions \mathcal{G}' we have $T_{G'}(\mathcal{G}') \subseteq T_G(\mathcal{G}')$. Since ϕ is monotonic, so is the operator T_C for all $C \subseteq N$. Hence by a straightforward induction $T_N^\infty \subseteq T_G^\infty$ for all $G \subseteq N$, and consequently, for all players i ,

$$T_N^\infty \subseteq \bigcap_{A:i \in A \in H} T_A^\infty. \quad (1)$$

Hence, for all $i \in N$,

$$T_N^\infty = T_{\{i\}}(T_N^\infty) \subseteq T_{\{i\}}(\bigcap_{A:i \in A \in H} T_A^\infty),$$

where the inclusion holds by the monotonicity of $T_{\{i\}}$. Consequently $T_N^\infty \subseteq \mathcal{G}(H)$.

We now prove the same claim for $\phi = sd^\ell$. We need to distinguish the T_C operator for $\phi = sd^\ell$ and $\phi = sd^g$. In the former case we write $T_{C,\ell}$ and in the latter case $T_{C,g}$. The reason that we use the latter operators is that they are monotonic and closely related to the former operators. Namely, as noted in [1], $T_{N,\ell}^\infty = T_{N,g}^\infty$, and the proof carries over for N replaced by an arbitrary $C \subseteq N$. Now fix an arbitrary $i \in N$, then

$$\bigcap_{A:i \in A \in H} T_{A,g}^\infty = \bigcap_{A:i \in A \in H} T_{A,\ell}^\infty,$$

and by (1) for $\phi = sd^g$, $T_{N,g}^\infty \subseteq \bigcap_{A:i \in A \in H} T_{A,g}^\infty$, so

$$T_{N,\ell}^\infty = T_{N,g}^\infty \subseteq \bigcap_{A:i \in A \in H} T_{A,\ell}^\infty. \quad (2)$$

Further, we have $T_{N,\ell}^\infty = T_{N,g}^\infty$ and $T_{N,g}^\infty = T_{\{i\},g}(T_{N,g}^\infty)$, so $T_{N,\ell}^\infty = T_{\{i\},g}(T_{N,\ell}^\infty)$. Hence, by (2) and monotonicity of $T_{\{i\},g}$,

$$T_{N,\ell}^\infty = T_{\{i\},g}(T_{N,\ell}^\infty) \subseteq T_{\{i\},g}(\bigcap_{A:i \in A \in H} T_{A,\ell}^\infty).$$

Also, for all $i \in N$ and all restrictions \mathcal{G}' we have, by definition,

$$T_{\{i\},g}(\mathcal{G}') \subseteq T_{\{i\},\ell}(\mathcal{G}'),$$

so by the last inclusion

$$T_{N,\ell}^\infty \subseteq T_{\{i\},\ell}(\bigcap_{A:i \in A \in H} T_{A,\ell}^\infty).$$

Consequently, $T_{N,\ell}^\infty \subseteq \mathcal{G}(H)$, as desired. \square

Proof of Proposition 3.4. We prove the following stronger proposition: If H' extends H and for each $G \in \{\{i\} \mid i \in N\} \cup H \cup H'$, T_G is monotonic, then $\mathcal{G}(H') \subseteq \mathcal{G}(H)$.

First note that for all restrictions \mathcal{G}' :

$$G \subseteq G' \text{ implies } T_{G'}(\mathcal{G}') \subseteq T_G(\mathcal{G}'). \quad (3)$$

To see this, suppose that $G \subseteq G'$. Then for all $i \in N$, either

- $i \in G$, in which case $i \in G'$, so $T_G(\mathcal{G}')_i = T_{G'}(\mathcal{G}')_i$, or
- $i \notin G$, in which case $T_G(\mathcal{G}')_i = \mathcal{G}'_i \supseteq T_{G'}(\mathcal{G}')_i$.

In each case, $T_{G'}(\mathcal{G}')_i \subseteq T_G(\mathcal{G}')_i$.

From (3) and the monotonicity of T_G and $T_{G'}$, it follows that:

$$G \subseteq G' \text{ implies } T_{G'}^\infty \subseteq T_G^\infty. \quad (4)$$

Now we show that if H' extends H , then $\bigcap_{A':i \in A' \in H'} T_{A'}^\infty \subseteq \bigcap_{A:i \in A \in H} T_A^\infty$: Fix $i \in N$ and take some $s_i \notin (\bigcap_{A:i \in A \in H} T_A^\infty)_i$. Then there is $A \in H$ such that $i \in A$ and $s_i \notin (T_A^\infty)_i$. Then since H' extends H , there is $A' \in H'$ such that $A \subseteq A'$, so by (4), $s_i \notin (T_{A'}^\infty)_i \supseteq \bigcap_{A':i \in A' \in H'} T_{A'}^\infty$.

So, since each $T_{\{i\}}$ is monotonic,

$$\mathcal{G}(H')_i = T_{\{i\}} \left(\bigcap_{A':i \in A' \in H'} T_{A'}^\infty \right)_i \subseteq T_{\{i\}} \left(\bigcap_{A:i \in A \in H} T_A^\infty \right)_i = \mathcal{G}(H)_i.$$

□

Preparing the proof of Theorem 4.2. In order to prove this result, we need some preparatory steps.

Lemma B.1. For any $\ell \geq 1$, $i \in N$, $s_i \in S_i$, and state (V, M) ,

$$\begin{aligned} & (V, M) \models \text{dom}^{\ell+1}(s_i) \\ \text{iff } & (V, M) \models \bigvee_{s'_i \in S_i} \bigwedge_{s_{-i} \in S_{-i}} ((K_i s'_i \succ_{s_{-i}} s_i) \vee \bigvee_{A:i \in A \in \overline{H}} \bigvee_{j \in A \setminus \{i\}} C_A \text{dom}^\ell(s_j)). \end{aligned}$$

Proof. We have

$$\begin{aligned} & (V, M) \models \text{dom}^{\ell+1}(s_i) \\ \text{iff (by definition)} & (V, M) \models K_i \bigvee_{s'_i \in S_i} \bigwedge_{s_{-i} \in S_{-i}} (s'_i \succ_{s_{-i}} s_i \vee \bigvee_{j \in N \setminus \{i\}} \text{dom}^\ell(s_j)) \\ \text{iff (by Theorem A.3)} & (V, M) \models \bigvee_{s'_i \in S_i} \bigwedge_{s_{-i} \in S_{-i}} (K_i s'_i \succ_{s_{-i}} s_i \vee \bigvee_{j \in N \setminus \{i\}} K_j \text{dom}^\ell(s_j)) \\ \text{iff } & (V, M) \models \bigvee_{s'_i \in S_i} \bigwedge_{s_{-i} \in S_{-i}} (K_i s'_i \succ_{s_{-i}} s_i \vee \bigvee_{A:i \in A \in \overline{H}} \bigvee_{j \in A \setminus \{i\}} C_A \text{dom}^\ell(s_j)). \end{aligned}$$

The last step holds by Lemma A.4 and Theorem A.5 since $\text{dom}^\ell(s_j) = K_j(\dots)$. □

Lemma B.2. For any $\ell \geq 1$, $i \in A \in \overline{H}$, $s_i \in S_i$, and state (V, M) ,

$$s_i \notin T_{A,M}^\ell(S_1, \dots, S_n)_i \text{ iff } (V, M) \models C_A \text{dom}^\ell(s_i).$$

Proof. By induction on ℓ . The base case follows straightforwardly from the definitions. Now assume the claim holds for ℓ . Then, focusing on the interesting case where $A \neq \{i\}$, we have the following chain of equivalences:

$$\begin{aligned}
& s_i \notin T_{A,M}^{\ell+1}(S_1, \dots, S_n)_i \\
& \text{iff (by definition)} \\
& s_i \notin T_{A,M}^{\ell}(S_1, \dots, S_n)_i \text{ or } \neg sd_{A,M}^g(s_i, T_{A,M}^{\ell}(S_1, \dots, S_n)) \\
& \text{iff (by monotonicity of } sd^g) \\
& \neg sd_{A,M}^g(s_i, T_{A,M}^{\ell}(S_1, \dots, S_n)) \\
& \text{iff (by definition)} \\
& \exists s'_i \in S_i \forall s_{-i} \in T_{A,M}^{\ell}(S_1, \dots, S_n)_{-i} M \upharpoonright_A \models s'_i \succ_{s_{-i}} s_i \\
& \text{iff } \exists s'_i \in S_i \forall s_{-i} \in S_{-i} M \upharpoonright_A \models s'_i \succ_{s_{-i}} s_i \text{ or} \\
& \quad s_{-i} \notin T_{A,M}^{\ell}(S_1, \dots, S_n)_{-i} \\
& \text{iff } \exists s'_i \in S_i \forall s_{-i} \in S_{-i} M \upharpoonright_A \models s'_i \succ_{s_{-i}} s_i \text{ or} \\
& \quad \exists j \in A \setminus \{i\} s_j \notin T_{A,M}^{\ell}(S_1, \dots, S_n)_j \\
& \text{iff (by induction hypothesis)} \\
& \exists s'_i \in S_i \forall s_{-i} \in S_{-i} M \upharpoonright_A \models s'_i \succ_{s_{-i}} s_i \text{ or} \\
& \quad \exists j \in A \setminus \{i\} (V, M) \models C_A \text{dom}^{\ell}(s_j) \\
& \text{iff (by Lemma A.4)} \\
& \exists s'_i \in S_i \forall s_{-i} \in S_{-i} (V, M) \models C_A s'_i \succ_{s_{-i}} s_i \text{ or} \\
& \quad \exists j \in A \setminus \{i\} (V, M) \models C_A \text{dom}^{\ell}(s_j) \\
& \text{iff } (V, M) \models \bigvee_{s'_i \in S_i} \bigwedge_{s_{-i} \in S_{-i}} (C_A s'_i \succ_{s_{-i}} s_i \vee \bigvee_{j \in A \setminus \{i\}} C_A \text{dom}^{\ell}(s_j)) \\
& \text{iff (by Theorem A.3)} \\
& (V, M) \models C_A \bigvee_{s'_i \in S_i} \bigwedge_{s_{-i} \in S_{-i}} (s'_i \succ_{s_{-i}} s_i \vee \bigvee_{j \in A \setminus \{i\}} \text{dom}^{\ell}(s_j)) \\
& \text{iff } (V, M) \models C_A \text{dom}^{\ell+1}(s_i).
\end{aligned}$$

□

We are now ready to prove the main result.

Proof of Theorem 4.2. We have:

$$\begin{aligned}
& s_i \notin \mathcal{G}(H, M)_i \\
& \text{iff (by definition)} \\
& s_i \notin T_{\{i\}, M}(\bigcap_{A:i \in A \in \overline{H}} T_{A, M}^\infty)_i \\
& \text{iff } \neg sd_{\{i\}, M}^g(s_i, \bigcap_{A:i \in A \in \overline{H}} T_{A, M}^\infty(S_1, \dots, S_n)) \\
& \text{iff } \exists s'_i \in S_i \forall s_{-i} \in \bigcap_{A:i \in A \in \overline{H}} T_{A, M}^\infty(S_1, \dots, S_n)_{-i} s'_i \succ_{s_{-i}} s_i \\
& \text{iff } \exists s'_i \in S_i \forall s_{-i} \in S_{-i} s'_i \succ_{s_{-i}} s_i \text{ or} \\
& \quad s_{-i} \notin \bigcap_{A:i \in A \in \overline{H}} T_{A, M}^\infty(S_1, \dots, S_n)_{-i} \\
& \text{iff } \exists s'_i \in S_i \forall s_{-i} \in S_{-i} s'_i \succ_{s_{-i}} s_i \text{ or} \\
& \quad \exists A : i \in A \in \overline{H} s_{-i} \notin T_{A, M}^\infty(S_1, \dots, S_n)_{-i} \\
& \text{iff } \exists s'_i \in S_i \forall s_{-i} \in S_{-i} s'_i \succ_{s_{-i}} s_i \text{ or} \\
& \quad \exists A : i \in A \in \overline{H} \exists j \in A \setminus \{i\} : s_j \notin T_{A, M}^\infty(S_1, \dots, S_n)_j \\
& \text{iff (by Lemma B.2)} \\
& \quad \exists s'_i \in S_i \forall s_{-i} \in S_{-i} s'_i \succ_{s_{-i}} s_i \text{ or} \\
& \quad (V, M) \models \bigvee_{A:i \in A \in \overline{H}} \bigvee_{j \in A \setminus \{i\}} C_A \text{dom}^\infty(s_j) \\
& \text{iff (since } s'_i \succ_{s_{-i}} s_i \in \text{At}_i) \\
& \quad \exists s'_i \in S_i \forall s_{-i} \in S_{-i} (V, M) \models K_i s'_i \succ_{s_{-i}} s_i \text{ or} \\
& \quad (V, M) \models \bigvee_{A:i \in A \in \overline{H}} \bigvee_{j \in A \setminus \{i\}} C_A \text{dom}^\infty(s_j) \\
& \text{iff } (V, M) \models \bigvee_{s'_i \in S_i} \bigwedge_{s_{-i} \in S_{-i}} ((K_i s'_i \succ_{s_{-i}} s_i) \vee \bigvee_{A:i \in A \in \overline{H}} \bigvee_{j \in A \setminus \{i\}} C_A \text{dom}^\infty(s_j)) \\
& \text{iff (by Lemma B.1)} \\
& \quad (V, M) \models \text{dom}^\infty(s_i). \quad \square
\end{aligned}$$