Games and Economic Behavior 119 (2020) 189–196

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Dominance rationality: A unified approach

John Hillas [a,*], Dov Samet [b,1]

[a] Department of Economics, Sir Owen G Glenn Building, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
[b] Tel Aviv University, Israel

ARTICLE INFO

Article history:
Received 9 April 2018
Available online 18 November 2019

Keywords:
Dominance rationality
Strict dominance
Weak dominance
Order independence
Flaws
Common knowledge of rationality
Iterative elimination

ABSTRACT

There are four types of dominance depending on whether domination is strict or weak and whether the dominating strategy is pure or mixed. Letting d vary over these four types of dominance, we say that a player is *d-dominance rational* when she does not play a strategy that is d-dominated relative to what she knows. For weak dominance by a mixed strategy, Stalnaker (1994) introduced a process of iterative maximal elimination of certain profiles that we call here *flaws*. We define here, analogously, d-flaws for each type of dominance d, and show that for each d, iterative elimination of d-flaws is order independent. We then show that the characterization of common knowledge of d-dominance rationality is the same for each d. A strategy profile can be played when d-dominance rationality is commonly known if and only if it survives an iterative elimination of d-flaws.

1. Introduction

There is a fairly substantial literature on the processes of the iterative elimination of dominated strategies and some literature on the justification of such processes in terms of common knowledge of rationality. Either explicitly or implicitly the justification is in terms of dominance relative to what the player knows or believes. We say that a player is dominance rational if she does not play a strategy that is dominated relative to the set of strategy profiles of the other players that she considers possible. This paper is concerned with this kind of dominance rationality and the characterization of its common knowledge in terms of elimination processes that are the same for all types of dominance.

The notion of dominance we have just discussed is a general term that refers to four different types of dominance: strict dominance by a mixed strategy (sm), strict dominance by a pure strategy (sp), weak dominance by a mixed strategy (wm), and weak dominance by a pure strategy (wp). Correspondingly, there are four notions of rationality which we denote by d-dominance rationality, for d in $\{sm, sp, wm, wp\}$.

We start with a description of various processes of elimination and their *informal* justification by common knowledge of d-dominance rationality. Iterative elimination of strictly dominated strategies (for pure or mixed domination) can be informally justified by common knowledge of strict dominance rationality. It is well known that similar claim, *mutatis mutandis*, cannot be made for iterative elimination of weakly dominated strategies. For some games, the argument that justifies this process by common knowledge of weak dominance rationality is inconsistent. In the next three subsections we demonstrate this inconsistency and show that the right way to avoid it is by eliminating flawed profiles, to be defined, rather than eliminating weakly dominated strategies.

* Corresponding author.
E-mail addresses: j.hillas@auckland.ac.nz (J. Hillas), dovsamet@gmail.com (D. Samet).

1 Samet acknowledges financial support of the Israel Science Foundation, grant 1827/14, and the Henry Crown Institute of Business Research in Israel.

https://doi.org/10.1016/j.geb.2019.11.001
0899-8256/© 2019 Published by Elsevier Inc.

1.1. Strict dominance

Iterative elimination of strictly dominated strategies can be justified by common knowledge of strict-dominance rationality. In each iteration of the elimination there is a smaller restricted game that the players know they are playing. Because players are strict-dominance rational, some strategies that are dominated relative to the set of profiles known to be played cannot be chosen by the players, and they are eliminated. Since strict-dominance rationality is commonly known, the elimination is known to the players, and the knowledge of the even smaller game starts the next iteration.

1.2. Weak dominance

Common knowledge of weak dominance rationality may fail to justify an iterative elimination of weakly dominated strategies. The simple example shown in Fig. 1 demonstrates this well known failure, and will lead us to a process that can be justified, at this stage informally, by common knowledge of weak dominance rationality.

		Player II	
		L	R
Player I	T	$2, 1$	$3, 0$
	B	$2, 0$	$2, 1$

Fig. 1. Iterative elimination of weakly dominated strategies.

At first glance, the argument for iterative elimination of weakly dominated strategies can be justified by common knowledge of weak dominance rationality analogously to the case of strict dominance. Thus, if Player I is weak dominance rational she should not play B. Knowing that, Player II should not play R. Thus common knowledge of weak dominance rationality should imply that they play (T, L) and this profile should be commonly known to be played.

The argument, however, is inconsistent. The elimination of the weakly dominated strategy B is justified if Player I does *not* exclude the possibility that Player II plays R. But the conclusion of the argument is that Player I *does* know that Player II plays L and not R. But, if this conclusion is right, then the elimination of B was unjustified.

The iterative elimination of strictly dominated strategies does not suffer from this inconsistency because of the property of strict dominance that we call *monotonicity*: if a strategy of a player is strictly dominated relative to some known set of strategy profiles of the other players, then it is strictly dominated also when the player knows more, that is, a subset of that set. This property does not hold for weakly dominated strategies. In the above example, strategy B is weakly dominated if all that Player I knows is that Player II may play either L or R, but is not dominated if Player I knows that Player II plays L.

Note that it is the iterative elimination of weakly dominated strategies that is inconsistent when it tries to capture common knowledge of weak dominance rationality. However, the idea that players commonly know that they avoid playing strategies that they know to be weakly dominated is coherent and meaningful, and so is the question of what they might play in this case.

1.3. Flaws

Understanding why common knowledge of weak dominance rationality fails to justify the process of elimination of weakly dominated strategies helps us to describe the right process that can be thus justified. The elimination of strategy B means that both profiles (B, L) and (B, R) are eliminated. It is justified only if Player I does not exclude the possibility that R is played. If she does exclude this possibility, then B is not weakly dominated given Player I's knowledge and hence the profile (B, L) may be played even when Player I is weak dominance rational. In contrast, the profile (B, R) cannot be played if Player I is weak dominance rational, because if it is played, Player I does not exclude the possibility that R is played, and then B *is* weakly dominated given Player I's knowledge. This will remain true no matter how much more knowledge she has at the end of the process.

Thus at this point, since we don't yet know what the state of knowledge of the players is when weak dominance rationality is commonly known, all we can say is that the profile (B, R) cannot be played if Player I is weak dominance rational. We call (B, R) a I-*weak flaw*. The "weak" refers to the fact that the strategy B, which is used to define the flaw (B, R) is *weakly*, rather than strictly dominated. We eliminate only this profile rather than the two profiles (B, R) and (B, L). Recall that it was the elimination of (B, L) that rendered the process of elimination of weakly dominated strategies inconsistent.

If Player II knows that Player I is weak dominance rational, she knows that the profiles that can be played are the three profiles other than (B, R). Therefore, if Player II plays R she knows that Player I plays T. But then, given this knowledge, L dominates R and she should not play R. Thus the II-flaw (T, R) is eliminated. We are left with the two profiles (T, L) and (B, L) that can be played when weak dominance rationality is commonly known.

This example demonstrates how we are led to the elimination of profiles rather than strategies in order to avoid the pitfalls of the inconsistency in eliminating weakly dominated strategies, the inconsistency being the result of the lack of monotonicity of weak dominance, as discussed in the example.

The flaws in this example are defined for weak dominance. The iterative maximal elimination of such flaws (maximal in the sense that in each iteration *all* flaws are eliminated) was introduced by Stalnaker (1994) to characterize certain Bayesian models. We discuss his work in the last subsection of the introduction.

Here, we extend this definition and define, analogously, d-flaws for all four types of dominance. We then show:

Order Independence. *For each d, all processes of iterative elimination of d-flaws terminate in the same set of profiles.*

For different d the terminal sets of profiles of the four processes may be all different.

The iterative elimination of strictly dominated strategies (by pure or mixed strategy) is a special case of iterative elimination of strict flaws, thus, the order independence of iterative elimination of strict flaws implies:

Strict dominance. *A profile survives the iterative elimination of strict flaws if and only if it survives the iterative elimination of strictly dominated strategies.*

1.4. Common knowledge of rationality

In order to define formally the notions of knowledge and common knowledge, which we previously discussed, we use the standard *partition structure* introduced by Aumann (1976), which consists of a state space and a partition for each player that defines her knowledge.

A *model* of the game is a partition structure where each state specifies the strategy profiles played in the state. Such models were used in Aumann (1995) and Aumann (1998) to study dominance rationality in extensive form games of perfect information.[2] In such a model we are able to define the event that a player is d-dominance rational as follows. At a given state ω, a player knows that the other players play one of the profiles specified in the states in the element of her partition that contains ω. The player is *d-dominance rational* at ω if the strategy she plays at ω is not d-dominated given the other players' profiles that she considers possible at ω.

We can now state our main result which characterizes common knowledge of all types of dominance rationality, and show, somewhat surprisingly, that all of them are completely analogous.

The unified theory of dominance rationality. *A strategy profile can be played when there is common knowledge of d-dominance rationality if and only if it survives the iterative elimination of d-flaws.*

1.5. Comparison to the prior literature

1.5.1. Flaws

Stalnaker (1994) introduced the insightful notion that he called an inferior profile and we call here an *sm*-flaw to maintain consistency with our other notions of flaws. Here we extended his notion to all four types of dominance which enables us to unify the characterization of common knowledge of dominance rationality for all types of dominance in terms of flaws. Stalnaker (1994), and the works that followed his idea, for example, Bonanno (2008) and Trost (2013) considered only the process of maximal elimination of *sm*-flaws, and thus there was no room in their work to study, let alone prove the order independence of the general processes of elimination of flaws.

1.5.2. Monotonicity and order independence

Gilboa et al. (1990), Apt (2011), Trost (2014), and others studied order independence of various processes of elimination. In these studies, a monotonicity property of strategy domination, and a weaker notion of hereditary, play an important role. However, in all these papers the processes concern an elimination of strategies or some relation on restrictions of the original game that are products of individual sets of strategies. We study here more general processes in which profiles, rather than strategies, are eliminated. Such processes involve subsets of strategy profiles that are not necessarily products. Our proof that monotonicity implies order independence for such processes is elementary and short. It does not follow by the order independence in the above mentioned studies. The opposite implication does hold: Our proof that iterative elimination of strict flaws is order independent implies the order independence of iterative elimination of strictly dominated strategies.

1.5.3. The relationship between various processes of elimination

Stalnaker (1994, p. 61) remarks that the process of maximal elimination of *wm*-flaws falls between the processes of iterative elimination of weakly and strictly dominated strategies. We prove this claim. To show that elimination of weakly

[2] The first paper that used a model for a game was Aumann (1987) that studied common knowledge of Bayesian rationality in a Bayesian model with a common prior.

dominated strategies is more demanding we use the fact that elimination of weak flaws is monotonic. To show that elimination of strictly dominated strategies is less demanding we use the claim that elimination of strict flaws and strictly dominated strategies end in the same terminal set.

Except for mentioning the set theoretic relation between elimination of weak flaws and weakly dominated strategies, Stalnaker (1994) does not provide any intuitive explanation of the relation between these two processes. Nor is such an explanation given in Bonanno (2008) or Trost (2013). The explanation is clear when we consider the monotonicity property of flaws presented here. The reason for the inconsistency of iterative elimination of weakly dominated strategies is the lack of monotonicity of that process. A weakly dominated strategy can be eliminated at some stage, and then, at a later stage, in a smaller restricted game, it is no longer weakly dominated, which reflects the lack of monotonicity. Thus we cannot conclude that a player will not use such a strategy before we know what exactly she knows when there is common knowledge of rationality. What we can conclude is that when the player is using a weakly dominated strategy, it must be the case that certain profiles cannot be played, exactly those profiles that cannot be played even in later stages. These are exactly the weak flaws. Thus eliminating of weak flaws is the way to avoid the lack of monotonicity of eliminating weakly dominated strategies.

1.5.4. The unified approach

We now discuss each of the four instances of the main theorem with regard to the existing literature. Characterizing common knowledge of strict-dominance rationality in terms of elimination of *strict flaws* is what enables the unification of the theory along the strict-weak axis. Since such flaws have never been defined or discussed no such a characterization can be found in the literature. But some works are related to characterization of common knowledge of strict dominance rationality in terms of elimination *strictly dominated strategies*.

For strict domination by pure strategies, Chen et al. (2007) characterized common knowledge of *sp*-dominance rationality by iterative elimination of strategies that are strictly dominated by pure strategies. Chen, Long and Luo used, like us, a model of the game based on a partition structure.

We are unaware of any work that dealt with the non-Bayesian notion of *sm*-rationality in a non-Bayesian model, and characterized its common knowledge in terms of elimination of strictly dominated strategies. However, all these non-Bayesian notions have equivalent Bayesian notions for which such a characterization does exist. This equivalence is based on the observation that a strategy is strictly dominated by a mixed strategy if and only if it is not a best response against any probabilistic belief about the profiles of the other players (see, Pearce, 1984).[3] Thus the set of profiles that survive iterative elimination of strategies dominated by a mixed strategy is the same as the set of profiles that survive iterative elimination of strategies which are not a best response, known as correlatedly rationalizable strategies (see, Brandenburger and Dekel, 1987).[4] Thus the non-Bayesian characterization of common knowledge of *sm*-dominance rationality can be derived, in a roundabout way, from their characterization. This is done by adding the right probabilistic beliefs to a non-Bayesian model and using the characterization of common knowledge of Bayesian rationality. Our non-Bayesian characterization is simpler conceptually and mathematically, and enables us to use the very same proof for both pure and mixed domination.

The case of weak dominance by pure strategies was proved by Bonanno (2008) and by Trost (2013). The first used a syntactic approach and the second, a type space with preference relations on the state space associated with states.[5] The result for weak dominance by mixed strategies was never stated and cannot be so easily derived from Bayesian models as could be done in the case for strict dominance discussed above. The natural candidate for comparison is Stalnaker (1994). As opposed to the approach used here and in other papers to fix the model of the game and vary the notion of rationality, Stalnaker deals with only Bayesian rationality and varies the family of Bayesian models by various properties of beliefs. Thus, in one family of Bayesian models common knowledge of Bayesian rationality is characterized by the elimination of strategies strictly dominated by mixed strategies, and in another family of models by the elimination of *sm*-flaws. Even if a derivation of our non-Bayesian result can be reached from Stalnaker's model it is not of much interest in light of our simple non-Bayesian derivation which is common to both pure and mixed domination.

2. Domination and flaws

Let G be a game with a finite set of players I, and a finite set of strategies S_i for each player i. The set of strategy profiles is $S = \times_i S_i$, and the set of the profiles of the players other than i, is $S_{-i} = \times_{j \neq i} S_j$. The payoff function for i is

---

[3] Probabilities are used in this equivalence on both sides; the dominating strategy is a mixed strategy and the beliefs about the play of the other players is a probability distribution over their strategy profiles. However, the probability of mixed strategy requires only a "roulette" to choose the pure strategy, while subjective probabilistic beliefs require heavier machinery to derive, like that in Savage (1954) or in Anscombe and Aumann (1963). Thus non-Bayesian notions of rationality seem to be conceptually more elementary than Bayesian rationality even when they are mathematically equivalent.

[4] Brandenburger and Dekel (1987) do not actually use the term correlatedly rationalizable strategies and nor do other authors who have followed them in using this concept. They do refer to correlated rationalizability and also to correlated rationalizable payoffs. But in the last phrase, the adjective 'correlated' modifies the noun 'payoffs', while it is the rationalizability which is correlated. Hence 'correlatedly' in our phrase which modifies the adjective 'rationalizable'.

[5] As was shown by Bonanno and Tsakas (2018), common belief, rather than knowledge, of weak dominance rationality is not characterized by elimination of flaws.

$h_i\colon S \to \mathbb{R}$. We denote by Σ_i the set of player i's mixed strategies. The payoff functions h_i are extended to $\Sigma = \times_i \Sigma_i$ by taking expectations. As usual, we embed S_i as a subset of Σ_i.

We define four types of dominance as follows.

Definition 1. Let T_{-i} be a nonempty subset of S_{-i}.[6]

1. *Strict domination.* A mixed strategy $\sigma_i \in \Sigma_i$ *strictly dominates* $s_i \in S_i$ *relative to* T_{-i} if $h_i(\sigma_i, t_{-i}) > h_i(s_i, t_{-i})$ for all $t_{-i} \in T_{-i}$. We say that s_i is *sm-dominated (by* σ_i*) relative to* T_{-i}. If σ_i is also pure, we say that s_i is *sp-dominated (by* σ_i*) relative to* T_{-i}.
2. *Weak domination.* A mixed strategy $\sigma_i \in \Sigma_i$ *weakly dominates* s_i *relative to* T_{-i} if $h_i(\sigma_i, t_{-i}) \geq h_i(s_i, t_{-i}))$ for all $t_{-i} \in T_{-i}$, and the inequality is strict for at least one $t_{-i} \in T_{-i}$. We say that s_i is *wm-dominated (by* σ_i*) relative to* T_{-i}. If σ_i is also pure, we say that s_i is *wp-dominated (by* σ_i*) relative to* T_{-i}.

Let $d \in \{sm, sp, wm, wp\}$ be one of the four types of dominance. We define processes of elimination of d-dominated strategies.

Definition 2. A process of *elimination of* d*-dominated strategies* is a strictly decreasing sequence of strategy-profile sets $S = S^0, S^1, \ldots, S^m$, such that for each $k \geq 0$, $S^k = \times_i S_i^k$; for each $k > 0$, S_i^k is obtained by eliminating from S_i^{k-1} some strategies which are d-dominated relative to S_{-i}^{k-1}; and where in the sets S_i^m there are no d-dominated strategies relative to S_{-i}^m. The set S^m is called the *terminal* set of the process.[7]

Next we define flaws.

Definition 3. A profile $s = (s_i, s_{-i})$ in $A \subseteq S$ is an *i-d-flaw of A* if for some strategy σ_i of i,

1. s_i is d-dominated by σ_i relative to $\{t_{-i} \mid (s_i, t_{-i}) \in A\}$;
2. $h_i(\sigma_i, s_{-i}) > h_i(s_i, s_{-i})$.

We say that a profile in A is a *d-flaw* of A if it is an i-d-flaw of A for some i.

It is straightforward to show that flaws have the following monotonicity property.

Claim 1 *(monotonicity of flaws). Let A and B be sets of profiles. If $s \in A \subseteq B$ is an i-d-flaw of B, then it is also an i-d-flaw of A.*

Definition 4. A process of *elimination of* d*-flaws* is a strictly decreasing sequence of strategy profile sets $S = S^0, S^1, \ldots, S^m$, such that for each $k > 0$, S^k is obtained by eliminating from S^{k-1} some strategy-profiles that are d-flaws of S^{k-1}, and such that there are no profiles in S^m that are d-flaws of S^m. The set S^m is called the *terminal* set of the process.

Due to the monotonicity property in Claim 1, processes of elimination of flaws have the desired property of order independence.

Proposition 1. *All processes of elimination of d-flaws have the same terminal set.*

Proof. Let $g(A)$ be the set of all profiles in A which are not d-flaws of A. By Claim 1, if $A \subseteq B$, then $g(A) \subseteq g(B)$.

Let $S^0, S^1, \ldots, S^m$ be a process of elimination of d-flaws. Then, by definition, for each $k > 0$, $g(S^{k-1}) \subseteq S^k$ and $g(S^m) = S^m$. Suppose that $g(T) = T$. We show by induction on k that $T \subseteq S^k$ for each k. As $S^0 = S$, the claim for $k = 0$ is obvious. Suppose that $T \subseteq S^k$, for $k < m$. Then, by the monotonicity of g and the induction hypothesis, $T = g(T) \subseteq g(S^k) \subseteq S^{k+1}$. Thus, terminal sets of different processes should contain each other, and therefore they are all the same.[8] □

[6] We do not assume that T_{-i} is a product set $\times_{j \neq i} T_j$.

[7] The process can be simplified by looking at stage k for strictly dominating strategies in S_i^{k-1} only, rather than in S_i as required here. This simplification is justified for finite games, where processes are finite, because in this case there exists a strictly dominating strategy in S_i if and only if there exists such a strategy in S_i^{k-1}. However, in infinite games, when the process is infinite, this equivalence breaks down. Dufwenberg and Stegeman (2002) studied conditions on infinite games under which the simplified process is order independent. Chen et al. (2007) showed that the full process is order independent.

[8] The process $S, g(S), g^2(S) \ldots$ is the maximal process, in the sense that in each stage *all* flaws are removed. The convergence of the maximal process to the largest fixed point of g is an instance of Kleene's fixed point theorem or Tarski's fixed point theorem for monotonic operators on lattices. Proposition 1 shows that monotonicity also implies that all processes converge to the same limit. The function g is also a contraction, that is, $g(A) \subseteq A$ which implies that the maximal process, is monotonically decreasing and implies also that starting from any event A, not necessarily S, $A, g(A), g^2(A), \ldots$ converges to the largest fixed point of g contained in A.

3. The relations between elimination processes

The relation between the iterative elimination of strict flaws and weak flaws is simple to state and to prove.

Claim 2. *The terminal set of elimination of wm-flaws (wp-flaws) is a subset of the terminal set of elimination of sm-flaws (sp-flaws).*

Proof. Since a strictly dominated strategy is also a weakly dominated strategy, it follows that a sm-flaw is in particular a wm-flaw. Thus, a process of elimination of sm-flaws is the beginning of a process of elimination of wm-flaws. The proof for the case of pure domination is the same. □

Next we consider the relation between elimination of strict flaws and elimination of weak dominated strategies.

Claim 3. *Any process of elimination of sm-dominated strategies (sp-dominated strategies) is a process of elimination of sm-flaws (sp-flaws).*

Proof. If $S^0, S^1, \ldots, S^m$ is a process of elimination of sm-dominated strategies, then for any s_i which is eliminated from S_i^k and any t_{-i} in S_{-i}^k, (s_i, t_{-i}) is an i-sm-flaw of S^k relative to $S_{-i}^k i$. Thus, the elimination of s_i form S_i^k, which is the elimination of all profiles (s_i, t_{-i}) in S^k, is an elimination of sm-flaws from S^k. The argument for pure domination is the similar. □

Thus, in view of the order independence stated in Proposition 1 we conclude:

Corollary 1. *The terminal set of all processes of elimination of sm-flaws (sp-flaws) is the terminal set of all processes of elimination of sm-dominated strategies (sp-dominated strategies).*

Finally, we consider the relation between processes of elimination of weak flaws and processes of elimination of weakly dominated strategies. While the first are order independent, the latter are not. However,

Proposition 2. *The terminal set of the processes of elimination of wm-flaws (wp-flaws) contains all the terminal sets of processes of elimination of wm-dominated strategies (wp-dominated strategies).*

Proof. Let $T = \times_i T_i$ be a terminal set of a process of elimination of wm-dominated strategies, and let $S = S^0, S^1, \ldots, S^m$ be a process of elimination of sm-flaws. We show by induction that for all k, $T \subseteq S^k$. This hold obviously for $k = 0$. Assume that $T \subseteq S_k$. Suppose that $s \in T$ is a wm-flaw of S_k, then by Claim 1 it is a wm-flaw of T. But this means that for some i, s_i is wm-dominated relative to T_{-i} which contradicts the definition of T. Thus, s is not a wm-flaw of S^k and therefore $s \in S^{k+1}$. The same proof holds for pure domination. □

The example in the introduction demonstrates that the terminal set of elimination of weak flaws can be larger than the terminal sets of processes of elimination of weakly dominated strategies. The relation between the terminal sets of the various processes of elimination is summarized in the following table, where weak and strict refer to either wm- and sm-dominance, or wp- and sp-dominance.

Any terminal set of elimination *of weakly dominated strategies*	$\subseteq$	*The* terminal set of elimination of *weak flaws*	$\subseteq$	*The* terminal set of elimination of *strict flaws*	$=$	*The* terminal set of elimination of *strictly dominated strategies*

Among the ways in which they differ, the elimination of weak flaws differs from the elimination of weakly dominated strategies in the way in they which treat pure strategy equilibria. An equilibrium profile can be eliminated in the latter, but not in the first, as, by definition, an equilibrium profile is never a flaw. Thus, all pure strategy equilibrium profiles are contained in the terminal set of the iterative elimination of weak flaws.[9]

[9] In fact, we can make a much stronger statement. In Hillas and Samet (2019) we show that no strategy profile played with positive probability in a correlated equilibrium is a d-flaw for any d.

4. Common knowledge of dominance rationality

To express formally rationality and its common knowledge we use a *model for the game* G.[10] The model is given by a knowledge structure and a description of the strategy profiles played in each of its states. The knowledge structure consists of a finite state space Ω with a partition Π_i for each player i. At a state ω player i knows all the events that contain $\Pi_i(\omega)$, the element of i's partition that contains ω. The meet of the partitions Π_i is the partition which is the finest among all partitions that are coarser than each Π_i. The event E is *common knowledge* at ω if the element of the meet that contains ω is a subset of E. Thus, the event that E is common knowledge is the union of all the meet's elements that are contained in E. (See Aumann, 1976.)

The strategic choices of the players are given by a function $\mathbf{s}\colon \Omega \to S$ which determines which strategy profile is played in each of the states. The strategy played by i in each state is given by the function $\mathbf{s}_i\colon \Omega \to S_i$, which satisfies $\mathbf{s}_i(\omega) = (\mathbf{s}(\omega))_i$. We further assume that each player knows which strategy she plays. This means that $\mathbf{s}_i$ is measurable with respect to Π_i, or in other words, for each player i and state ω, i plays the same strategy in all the states in $\Pi_i(\omega)$. For any event E we write $\mathbf{s}(E)$ for $\{\mathbf{s}(\omega) \mid \omega \in E\}$ and $\mathbf{s}_{-i}(E)$ for $\{\mathbf{s}_{-i}(\omega) \mid \omega \in E\}$.

Note, that $T_{-i} = \mathbf{s}_{-i}(\Pi_i(\omega))$ is the set of profile strategies of the players other than i played in $\Pi_i(\omega)$. Thus, the event that the players other than i play a profile in T_{-i} contains the event $\Pi_i(\omega)$. Therefore, i knows at ω that the other players play a strategy profile in T_{-i}. We can now define the event that a player is rational.

Definition 5. Player i is *d-dominance rational* in state ω if the strategy she plays in ω is not d-dominated relative to the set of profiles of the other players which she considers possible at ω. That is, there is no strategy of hers that d-dominates $\mathbf{s}_i(\omega)$ relative to the set $\mathbf{s}_{-i}(\Pi_i(\omega))$.[11]

Definition 6. A strategy profile s is *compatible with common knowledge of d-dominance rationality*, if there is a model of the game, and a state ω in the model, such that $\mathbf{s}(\omega) = s$, and it is common knowledge at ω that all players are d-dominance rational.

Theorem 1. *A strategy profile is compatible with common knowledge of d-dominance rationality if and only if it is in the terminal set of the processes of elimination of d-flaws.*

Proof. Let $S^0, S^1, \ldots, S^m$ be a process of elimination of weak (strict) flaws. Suppose that in some model for G it is common knowledge in some state that players are weak-dominance (strict-dominance) rational. By restricting the model to the event that weak-dominance (strict-dominance) is common knowledge, we can assume, without loss of generality, that the players are weak-dominance (strict-dominance) rational in each state.

We show by induction that $\mathbf{s}(\Omega) \subseteq S^k$ for each $k \le m$. This is obvious for $S^0 = S$. Suppose we proved it for k. Observe that for each ω, $\mathbf{s}(\omega) \in \mathbf{s}(\Pi_i(\omega)) \subseteq \mathbf{s}(\Omega) \subseteq S^k$, where the last inclusion is the induction hypothesis. Thus, if, contrary to what we want to show, $\mathbf{s}(\omega) \notin S^{k+1}$, then, for some i, it is a weak (strict) i-flaw of S^k. It follows by Claim 1 that $\mathbf{s}(\omega)$ is a weak (strict) i-flaw of $\mathbf{s}(\Pi_i(\omega))$. But this implies that some strategy $\hat{s}_i$ of i weakly (strictly) dominates $\mathbf{s}_i(\omega)$ relative to $\{t_{-i} \mid (\mathbf{s}_i(\omega), t_{-i}) \in \mathbf{s}(\Pi_i(\omega))\} = \mathbf{s}_{-i}(\Pi_i(\omega))$. This means that i is not weak-dominance (strict-dominance) rational in ω, contrary to our assumption. Thus, $\mathbf{s}(\omega) \in S^{k+1}$ for each ω, that is, $\mathbf{s}(\Omega) \subseteq S^{k+1}$.

For the converse direction we construct a model in which weak-dominance (strict-dominance) rationality holds in all states (and thus is common knowledge in each state) and $S^m = \mathbf{s}(\Omega)$. We take Ω to be S^m and set $\mathbf{s}(s) = s$. For each i and $s \in \Omega$, we define the partition Π_i such that each player knows what she plays, that is, $\Pi_i(s) = \{s' \mid s'_i = s_i\}$. It follows immediately from the fact that for any i there are no weak (strict) i-flaws in S^m, that for each state $s \in S^m$ each player is weak-dominance (strict-dominance) rational at s. $\square$

The order independence of iterative elimination of weak (strict) flaws, which was proved in Proposition 1, is also a corollary of this theorem, as each terminal set of such a process coincides with the same set of profiles that can be played when weak-dominance (strong-dominance) rationality is commonly known. Note, that the proof makes use of the monotonicity of flaws described in Claim 1, which is used to prove directly the order independence in Proposition 1.

References

Anscombe, F.J., Aumann, R.J., 1963. A definition of subjective probability. Ann. Math. Stat. 34, 199–205.
Apt, K.R., 2011. Direct proofs of order independence. Econ. Bull. 1, 105–116.
Aumann, R.J., 1976. Agreeing to disagree. Ann. Stat. 4 (6), 1236–1239.

[10] The model of a game used here is the one described in Chen et al. (2007). It is similar to the one used in Aumann (1987) without the probabilistic beliefs associated with the states. It is also similar to the model of a game in extensive form in Aumann (1995) and Aumann (1998).

[11] Using knowledge operators, this event can be described as the event that a player will not knowingly play a strategy that yields her less than she could have gotten with a different strategy. See Aumann (1995).

Aumann, R.J., 1987. Correlated equilibria as an expression of Bayesian rationality. Econometrica 55, 1–18.
Aumann, R.J., 1995. Backward induction and common knowledge of rationality. Games Econ. Behav. 8, 6–19.
Aumann, R.J., 1998. On the centipede game. Games Econ. Behav. 23, 97–105.
Bonanno, G., 2008. A syntactic approach to rationality in games with ordinal payoffs. In: Bonanno, G., van der Hoek, W., Wooldridge, M. (Eds.), Logics and the Foundations of Game and Decision Theory. Amsterdam University Press, pp. 59–86.
Bonanno, G., Tsakas, E., 2018. Common belief of weak-dominance rationality in strategic-form games: a qualitative analysis. Games Econ. Behav. 112, 231–241. https://doi.org/10.1016/j.geb.2018.09.003.
Brandenburger, A., Dekel, E., 1987. Rationalizability and correlated equilibria. Econometrica 55, 1391–1402. http://www.jstor.org/stable/1913562.
Chen, Y.C., Long, N.V., Luo, X., 2007. Iterated strict dominance in general games. Games Econ. Behav. 61, 299–315.
Dufwenberg, M., Stegeman, M., 2002. Existence and uniqueness of maximal reductions under iterated strict dominance. Econometrica 70, 2007–2023.
Gilboa, I., Kalai, E., Zemel, E., 1990. On the order of eliminating dominated strategies. Oper. Res. Lett. 9, 85–89.
Hillas, J., Samet, D., 2019. Non-probabilistic correlated equilibrium. A manuscript.
Pearce, D.G., 1984. Rationalizable strategic behavior and the problem of perfection. Econometrica 52, 1029–1050.
Savage, L.J., 1954. The Foundations of Statistics. Wiley, New York.
Stalnaker, R., 1994. On the evaluation of solution concepts. Theory Decis. 37, 49–73.
Trost, M., 2013. Epistemic characterizations of iterated deletion of inferior strategy profiles in preference-based type spaces. Int. J. Game Theory 42, 755–776.
Trost, M., 2014. An epistemic rationale for order independence. Int. Game Theory Rev. 16.