WEAK AND STRICT DOMINANCE: A UNIFIED APPROACH

JOHN HILLAS AND DOV SAMET

Abstract. *Strict*-dominance rationality is a non-Bayesian notion of rationality which means that players do not chose strategies they know to be strictly dominated. Similarly, *weak*-dominance rationality means that players do not choose strategies that they know to be weakly dominated. Iterative elimination of *strictly* dominated strategies can be intuitively and formally justified by players having common knowledge of *strict*-dominance rationality. In contrast, common knowledge of *weak*-dominance rationality fails to justify iterative elimination of *weakly* dominated strategies.

Examining the reasons for this failure leads to a characterization of the strategy profiles played when weak-dominance rationality is commonly known. These are profiles that survive a process of iterative elimination of profiles called *weak flaws* that was introduced by Stalnaker (1994) to characterize certain Bayesian models of games. We define, analogously, *strict* flaws, and show that the iterative elimination of either weak or strict flaws is order independent. Our main result is that the case of weak dominance and strict dominance are completely analogous: Common knowledge of weak-dominance or strict-dominance rationality is characterized by iterative elimination of weak or strict flaws correspondingly. Our results hold equally for domination by pure and mixed strategies, which distinguish them from characterizations in Bayesian models that hold only for mixed domination.

1. Introduction

We start with the description of various processes of elimination and their informal justification by common knowledge of rationality. When we refer to dominance in what follows, we mean either domination by pure strategies or domination by mixed strategies.

1.1. **Strict dominance.** We call the avoidance of playing strategies that a player knows to be strictly dominated *strict-dominance rationality.* 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 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.

Date: Draft of April 10, 2018.

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

1.2. **Weak dominance.** *Weak-dominance rationality* of a player is analogously defined as the avoidance of playing a strategy that the player knows to be weakly dominated. However, common knowledge of weak-dominance rationality may fail to justify an iterative elimination of weakly dominated strategies. The following simple example demonstrates this well known failure, and will help us to describe a process that can be justified, at this stage informally, by common knowledge of weak-dominance rationality.

	L	R
T	$2, 1$	$3, 0$
B	$2, 0$	$2, 1$

FIGURE 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 is not inflicted by 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 her opponents' strategy profiles, 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 which is flawed, at least when it tries to capture common knowledge of weak-dominance rationality. However, the idea that players commonly know that they avoid playing strategies which they know to be weakly dominated is coherent and meaningful, and so is the question what they might play in this case.

1.3. **Flaws.** Understanding why common knowledge of weak-dominance rationality fails to explain the process of elimination of weakly dominated strategies, helps us to describe a process that can be thus explained. 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 L 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 (T,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 that we don't know yet 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 *weak* I-*flaw*. The "weak" refers to the weakly dominated strategy to which this flaw belongs. We eliminate only this profile rather than the two profiles (B,R) and (B,T). Recall that it was the elimination of (B,T) 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 (R,L) which can be played when weak-dominance rationality is commonly known.

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

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 model of interactive knowledge, introduced by Aumann (1976). This model consists of a state space where the knowledge of a player is given by a partition.

In order to be able to define the event that a player is rational, each state specifies the strategy profiles played in the state. Thus, at a given state ω, a player knows that her opponents chose one of the profiles that are played in the element of her partition that contains ω. This model for games was introduced in Aumann (1987). In that paper, in each state each player is endowed with probability distribution over the state space, which induces a probability distribution over her opponents' profiles. The paper studies Bayesian rationality, according to which a player is rational in a state if her strategy in the state maximizes her expected payoff given her probabilistic beliefs.

Non-Bayesian dominance rationality, was studied in Aumann (1995) and Aumann (1998) for games in extensive form of perfect information. These papers departed in two ways from Aumann (1987). They moved from Bayesian rationality to non-Bayesian rationality, and from games in strategic form to games in extensive form. It was Chen, Long, and Luo (2007) who moved only one step and studied strict-dominance rationality in games of *strategic form.* A player is *strict-dominance rational* in a state ω if the strategy she plays at ω is not strictly dominated relative to her opponents' profiles that she considers possible, that is, the strategy profiles played by her opponents in the states of her partition element that contains ω.

They showed that a profile of strategies can be played when there is common knowledge of strict-dominance rationality if and only if it survives the iterative elimination of strictly dominated strategies.[1]

Weak-dominance rationality is defined here analogously to strict-dominance rationality in the obvious way. We study common knowledge of both weak and

[1]Chen, Long, and Luo (2007) deal only with dominance by pure strategies in infinite games, but their result and proof apply verbatim to dominance by mixed strategies in finite games.

strict-dominance rationality, and show that both are characterized by the elimination of *flaws*. A profile of strategies (s_i, s_{-i}) in a set of profiles A is called a *weak i-flaw* of A, if some strategy $\hat{s}_i$ in S_i weakly dominates s_i relative to the profiles t_{-i} such that (s_i, t_{-i}) is in A, and $\hat{s}_i$ yields higher payoff than s_i against s_{-i}. The flaw is a *strict i-flaw* if $\hat{s}_i$ strictly dominates s_i. We show that iterative elimination of either strict flaws or weak flaws is order independent.

Order Independence. *All processes of iterative elimination of weak (strict) flaws terminate in the same set of profiles.*

For strict flaws, order independence 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.*

We can now state our main result which characterizes common knowledge of weak and strict-dominance rationality, and show, somewhat surprisingly, that the two cases are completely analogous.

Main characterization. *A strategy profile can be played when there is common knowledge of weak (strict) dominance if and only if it survives the iterative elimination of weak (strict) flaws.*

1.5. **Comparison to existing literature.** A strategy can be dominated by a pure strategy or by a mixed strategy, which we refer to as pure and mixed dominance. The result of iterative elimination of flaws with pure dominance may differ from the result of eliminating flaws with mixed dominance. All our results hold for both types of domination. Thus, our main result is a characterization of, not two, but four cases of common knowledge of rationality: weak pure-dominance, strict pure-dominance, weak mixed-dominance, and strict mixed-dominance rationality. Indeed, we show that all four cases can be described in a single theorem.

The two cases of pure domination are meaningful even when the payoffs in the game are not von Neumann-Morgenstern utility. The well known results obtained for Bayesian rationality cannot be used in these cases. The two cases of mixed domination require that the payoffs are given in terms of von Neumann-Morgenstern utility because expectation of payoffs are compared. However, the notion of mixed strategies is simpler than the notion of probabilistic belief. The first requires only a “roulette” to chose the pure strategy, while subjective probabilistic beliefs require heavier machinery to derive, like the one in Savage (1954) or in Anscombe and Aumann (1963). Thus, non-Bayesian notions of rationality seem to be more elementary than Bayesian rationality.

The case of strict pure-dominance rationality was considered by Chen, Long, and Luo (2007) who showed that common knowledge of this type of rationality is equivalent to iterative elimination of *strategies* that are strictly dominated by pure strategies. This characterization is equivalent to our characterization for this type of rationality, by the claim about strict dominance given above.

Tan and Werlang (1988) showed that common knowledge of Bayesian rationality characterizes the set of profiles that result from iterative elimination of strategies that are not best response. This set was defined by Brandenburger and Dekel (1987) as *correlated rationalizable strategies.* A strategy is a best response with respect to some probabilistic belief about the opponents if and only if it is not dominated by a mixed strategy (Pearce, 1984). Thus the set of profiles characterized by Tan and

Werlang (1988) is the set of profiles that survive iterative elimination of strategies strictly dominated by mixed strategies that we characterize here. However, our characterization of this set is simpler because it does not make use of probabilistic beliefs. More importantly, our characterization is analogous to the characterization of the set of profiles that survive iterative elimination of strategies that are strictly dominated by *pure* strategies, about which Bayesian rationality is silent. Moreover, the proof for the two cases is exactly the same.

Our characterizations of common knowledge of weak pure-dominance and weak mixed-dominance rationality are new. The idea of eliminating profiles, which we call flaws, is due to Stalnaker (1994).[2] He considered only weak mixed-domination and only the maximal process in which in each stage all the weak flaws are eliminated. The restriction to the maximal process is not a limitation since, as we show, processes of elimination of flaws are order independent. However, here, we introduce three more notions of flaws, weak flaws for pure domination and strict flaws for both pure and mixed domination, the elimination of which are all order independent. The order independence implies that iterative elimination of strictly dominated strategies results in the same set of profiles as iterative elimination of strict flaws, for both pure and mixed domination. This enables us to show the similarity of the cases of weak and strict domination in our main result.

The main difference between Stalnaker (1994) and this paper is the notions of rationality. In Stalnaker (1994) it is standard Bayesian rationality and hence have no implication to the cases of pure domination. To account for the difference between weak and strict mixed-domination Stalnaker varies not the type of rationality but rather the family of belief structures. Here, the knowledge structure is fixed and the notion of non-Bayesian rationality varies.

2. Domination

Let G be a game with a finite set of players I, and a finite set of strategies S_i for each player i.[3] The set of strategy profiles is $S = \times_i S_i$, and the set of the profiles of i's opponents is $S_{-i} = \times_{j \neq i} S_j$. The payoff function for i is $h_i \colon S \to \mathbb{R}$.

In order to describe processes of elimination, as well as the notions of dominance rationality, we use the following terminology.

Definition 1. (relative domination) Let T_{-i} be a nonempty subset of S_{-i}. A strategy $\hat{s}_i \in S_i$ *strictly dominates* s_i *relative to* T_{-i} if $h_i(\hat{s}_i, t_{-i}) > h_i(s_i, t_{-i})$ for all $t_{-i} \in T_{-i}$. We say in this case that s_i is *strictly dominated relative to* T_{-i}.[4]

The strategy $\hat{s}_i$ *weakly dominates* s_i *relative to* T_{-i} if $h_i(\hat{s}_i, t_{-i}) \geq h_i(s_i, t_{-i})$ for all $t_{-i} \in T_{-i}$, and the inequality is strict for at least one $t_{-i} \in T_{-i}$. We say in this case that s_i is *weakly dominated relative to* T_{-i}.

Using this terminology we define processes of elimination of dominated strategies.

---

[2] Bonanno and Nehring (1998) pointed out some errors in Stalnaker (1994) and one way of correcting some of them. Stalnaker (1998) discussed the reason for one of the errors and another way of correcting it.

[3] For pure strategy domination, the finiteness of the strategy sets is assumed only for simplicity. All the results can be easily formulated and proved for infinite games, where the order of elimination can be any ordinal. The arguments by induction in the proofs are easily transformed to arguments by transfinite induction.

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

Definition 2. A process of *elimination of strictly (weakly) dominated strategies* is a strictly decreasing sequence of strategy-profile sets $S^0, S^1, \ldots, S^m$, such that for each $k \geq 0$, $S^k = \times_i S_i^k$, where $S_i^0 = S_i$; for each $k > 0$, S_i^k is obtained by eliminating from S_i^{k-1} some strategies which are strictly (weakly) dominated relative to S_{-i}^{k-1}; and where in the sets S_i^m there are no strictly (weakly) dominated strategies relative to S_{-i}^m. The set S^m is called the *terminal* set of the process.[5]

No use of mixed strategies was made in Definitions 1 and 2. Thus, they can be applied to games in which the utilities h_i are ordinal. In case h_i are cardinal utilities then they can be extended to mixed strategy profiles which gives rise to domination by mixed strategies.

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. Domination by mixes strategy is defined as in Definition 1, except that the dominating strategy $\hat{s}_i$ is now taken in Σ_i. Similarly, a process of elimination of strategies dominated by mixed strategies is defined as in Definition 2, where in each stage the eliminated strategies are dominated by mixed strategies. The terminal sets of such processes may differ from the terminal sets of processes in which only strategies dominated by pure strategies are eliminated.

The theory we propose in the sequel applies to the two type of domination: by pure and by mixed strategies. Thus we talk in general about domination which can be either of them.

3. Informal justification of the iterated processes

The iterative elimination of strategies seem to be justified informally by common knowledge of rationality. We call this justification informal because knowledge is not represented rigorously in this process. We refer to two notions of rationality: strict-dominance rationality by which we mean that a player does not play a strictly dominated strategy, and weak-dominance rationality which requires that a player does not play a weakly dominated strategy. In what follows we use “dominated” and “rational” to mean either strictly dominated and strict-dominance rational, or weakly dominated and weak-dominance rational.

Let $S^0, S^1, \ldots, S^m$ be a process of elimination of dominated strategies. Assuming common knowledge of rationality, we conclude in the first step that the profiles played must be in S^1, because the players are rational. As they all know that they are rational, they know that the game played is S^1, and thus, being rational, the strategy profile played must be in S^2, and so on.

We note the following three observations.

(1) Rationality in this argument means that a player does not play a strategy that is dominated *relative to what she knows about the strategy profiles played by her opponents* (see, e.g., Samuelson, 1992, p. 287).

---

[5]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, Long, and Luo (2007) showed that the full process is order independent.

(2) The knowledge of players increases along the process, where in stage k each player i knows that her opponents are playing in S_i^k.
(3) According to this argument, when there is common knowledge of rationality the profile played is in the terminal set S^m, and this is commonly known.

These points seem to be at odds. If common knowledge of rationality implies that each player i knows that her opponents play a strategy in S_{-i}^m why are we justified to eliminate strategies relative to some larger sets, S_{-i}^k, which reflect less knowledge than her actual knowledge? This question can be answered in the case of strict-dominance rationality, because of the following property.

Claim 1. (strict-dominance monotonicity) *If a strategy of i is strictly dominated relative to $T_{-i} \subseteq S_{-i}$ then it is also strictly dominated relative to $T'_{-i} \subseteq T_{-i}$.*[6]

Thus, a strategy of i eliminated in stage k because it is strictly dominated relative to S_{-i}^k, is also strictly dominated relative to S_{-i}^m which is what player i actually knows about her opponents. Hence, the assumption that players know in stage k less than they actually know (in the terminal stage) is innocuous.

However, this monotonicity property does not hold for weakly dominated strategies, and as a result the argument that common knowledge of weak-dominance rationality implies the iterative elimination of weakly dominated strategies is self defeating. It is possible that a strategy of i is weakly dominated relative to S_{-i}^k for some $k < m$, but not relative to S_{-i}^m, and thus its elimination in stage $k+1$ is not justified in light of the purported actual knowledge.

Understanding the defect of the informal argument for the elimination of weakly dominated strategies leads to the right process of elimination. At stage k of the defective process we eliminate a weakly dominated strategy of i, s_i. This amounts to eliminating from S^k all profiles in S^k in which player i plays s_i. As we said, this cannot be justified, because at this stage we still do not know what the actual knowledge of the players is under the assumption that they have common knowledge of weak-dominance rationality. Some of the profiles (s_i, t_{-i}) may be played with the actual knowledge, if relative to this knowledge s_i is not weakly dominated. We can easily tell which profiles *may* be played eventually and which ones may not. If there is a strategy that weakly dominates s_i and yields i a *higher* payoff than s_i against t_{-i}, then no matter how much more i knows about her opponent, the profile (s_i, t_{-i}) cannot be played if i is weak-dominance rational. If, in contrast, all strategies that dominate s_i yield i the same payoff as s_i against t_{-i}, then it is possible that by knowing more, i may conclude that s_i is not weakly dominated given her knowledge. In summary, when s_i is weakly dominated, we cannot eliminate *all* profiles (s_i, t_{-i}) but only those profiles for which there is a strategy that weakly dominates s_i *and* yields i a higher payoff than s_i against t_{-i}. We call these profiles weak flaws. Similarly we define strict flaw for the case that s_i is strictly dominated.

We next define flaws and their iterative elimination and show that the result of iterative elimination of flaws is characterized by common knowledge of rationality in both the strict and weak cases.

[6]A similar claim is made in Gilboa et al. (1990) for the case that T_{-i} and T'_{-i} are product sets, $T_{-i} = \times_{j \neq i} T_j$ and $T'_{-i} = \times_{j \neq i} T'_j$. They refer to this property as hereditary.

4. FLAWS

Definition 3. A profile $s = (s_i, s_{-i})$ in $A \subseteq S$ is a *weak (strict) i-flaw of A* if for some strategy $\hat{s}_i$ of i,

(1) s_i is weakly (strictly) dominated by $\hat{s}_i$ relative to $\{t_{-i} \mid (s_i, t_{-i}) \in A\}$;
(2) $h_i(\hat{s}_i, s_{-i}) > h_i(s_i, s_{-i})$.

We say that a profile in A is a *weak (strict) flaw* of A if it is a weak (strict) i-flaw of A for some i.

It is straightforward to show that flaws have a property, stated next, which is similar to the monotonicity of strictly dominated strategies in Claim 1.

Claim 2. (monotonicity of flaws) *Let A and B be sets of profiles. If $s \in A \subseteq B$ is a weak (strict) i-flaw of B, then it is also a weak (strict) i-flaw of A.*

Definition 4. A process of *elimination of weak (strict) 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 weak (strict) flaws of S^{k-1}, and such that there are no profiles in S^m that are weak (strict) flaws of S^m. The set S^m is called the *terminal* set of the process.

Due to the monotonicty property in Claim 2, processes of elimination of flaws have the desired property of order independence.

Proposition 1. *All processes of elimination of weak (strict) flaws have the same terminal set.*

Proof. Let $g(A)$ be the set of all profiles in A which are not weak (strict) flaws of A. By Claim 2, if $A \subseteq B$, then $g(A) \subseteq g(B)$.

Let $S^0, S^1, \ldots, S^m$ be a process of elimination of weak (strict) 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.[7] □

The relation between elimination of weak and strict flaws is simple.

Claim 3. *The terminal set of elimination of weak flaws is a subset of the terminal set of elimination of strict flaws.*

Proof. Since a strictly dominated strategy is also a weakly dominated strategy, it follows that a strict flaw is in particular a weak flaw. Thus, a process of elimination of strict flaws is the beginning of a process of elimination of weak flaws. □

---

[7]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 latices. 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.

Next we consider the relation between elimination of strict flaws and elimination of strictly dominated strategies. If $S^0, S^1, \dots, S^m$ is a process of elimination of strictly 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 a strict i-flaw of S^k. 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 strict flaws from S^k. Hence, a process of elimination of strictly dominated strategies is an instance of a process of elimination of strict flaws. Thus, in view of the order independence stated in Proposition 1 we conclude:

Corollary 1. *The terminal set of all processes of elimination of strict flaws is the terminal set of all processes of elimination of strictly 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 weak flaws contains all the terminal sets of processes of elimination of weakly dominated strategies.*

Proof. Let $T = \times_i T_i$ be a terminal set of a process of elimination of weakly dominated strategies, and let $T \subseteq \hat{S} \subseteq S$. If $s \in T$ is a flaw of $\hat{S}$, then by Claim 2 it is a flaw of T. But this means that for some i, s_i is weakly dominated relative to T_{-i} which contradicts the definition of T. Thus, in any process of elimination of weak flaws, profiles of T cannot be eliminated. □

We see in Example 1 in the next section that the terminal set of elimination of weak flaws can be larger than the union of all 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.

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*

Elimination of weak flaws differs from elimination of weakly dominated strategy concerning equilibria. An equilibrium profile can be eliminated in the latter, but cannot be, by definition, a flaw. Thus, all equilibria are contained in the terminal set of the iterative elimination of weak flaws.

5. Common knowledge of rationality

To express formally rationality and its common knowledge we use a *model for the game* G.[8] 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

[8]The model of a game used here is the one described in Chen, Long, and Luo (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).

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 element 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 i's opponents played in $\Pi_i(\omega)$. Thus, the event that i's opponents play a profile in T_{-i} contains the event $\Pi_i(\omega)$. Therefore, i knows at ω that her opponents play a strategy profile in T_{-i}. We can now define the event that a player is rational, in agreement with observation (1) in Section 3.

Definition 5. Player i is *strict-dominance (weak-dominance) rational* in state ω if the strategy she plays in ω is not strictly (weakly) dominated relative to the set of her opponents' profiles which she considers possible at ω. That is, there is no strategy of hers that strictly (weakly) dominates $\mathbf{s}_i(\omega)$ relative to the set $\mathbf{s}_{-i}(\Pi_i(\omega))$.[9]

Definition 6. A strategy profile s is *compatible with common knowledge of weak-dominance (strict-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 strict-dominance (weak-dominance) rational.

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

Proof. Let $S^0, S^1, \dots, S^m$ 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 $k \leq m$. 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 2 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

[9]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) and Hillas and Samet (2018).

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. □

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 (stong-dominance) rationality is commonly known. Note, that the proof makes use of the monotonicity of flaws described in Claim 2, which is used to prove directly the order independence in Proposition 1.

	L	C	R
T	1, 0	2, 0	3, 0
M	1, 2	2, 1	2, 1
B	1, 1	0, 3	4, 2

	L	C	R
T	1, 0	2, 0	3, 0
M	1, 2	3, 1	2, 1
B	1, 1	0, 3	

	L	C	R
T	1, 0	2, 0	3, 0
M	1, 2	3, 1	2, 1
B	1, 1		

	L	C	R
T	1, 0	2, 0	3, 0
M	1, 2		
B	1, 1		

Figure 2. Elimination of weak flaws

Example 1. Consider the game on the left side of Figure 2, where Player I is the row player, and Player II the column player. Strategy R is weakly dominated by C. The profile (B,R) is a weak I-flaw. After it is eliminated, (B,L) and (B,C) are the only profiles in which B is played. Since B is weakly dominated relative to $\{\mathrm{L}, \mathrm{C}\}$,(B,C) is a weak II-flaw. After its removal, C and R are weakly dominated relative to $\{\mathrm{T}, \mathrm{M}\}$ and the weak II-flaws, (M,C) and (M,R) are eliminated.

There are several processes of elimination of weakly dominated strategy and they end in different set of strategy profiles. The ending sets of these processes are $\{(\mathrm{T}, \mathrm{L}), (\mathrm{M}, \mathrm{L})\}$, $\{(\mathrm{T}, \mathrm{L}), (\mathrm{T}, \mathrm{C})\}$, and $\{(\mathrm{T}, \mathrm{C})\}$. However, common knowledge of weak-dominance rationality cannot imply that only one of these profiles should be played. Indeed, if this were the case, then it would be commonly known that one of these profiles is played. But, then, strategy R would not be weakly dominated relative to I's knowledge, as the only reason for R not to conform with weak-dominance rationality is the possibility that (B, C) is played, which according to our assumption is commonly known not be the case.

References

Francis J. Anscombe and Robert J. Aumann (1963): "A Definition of Subjective Probability," *The Annals of Mathematical Statistics*, 34, 199–205.

Robert J. Aumann (1976): "Agreeing to disagree," *Ann. Statist.*, 4(6), 1236–1239.

Robert J. Aumann (1987): "Correlated Equilibria as an Expression of Bayesian Rationality," *Econometrica*, 55(1), 1–18.

Robert J. Aumann (1995): "Backward induction and common knowledge of rationality," *Games and Economic Behavior*, 8(1), 6–19.

Robert J. Aumann (1998): "On the centipede Game," *Games and Economic Behavior*, 23, 97–105.

Giacomo Bonanno and Klaus Nehring (1998): "On Stalnaker's notion of strong rationalizability and Nash equilibrium in perfect information games," *Theory and Decision*, 45(3), 291–295.

Adam Brandenburger and Eddie Dekel (1987): "Rationalizability and Correlated Equilibria," *Econometrica*, 55(6), 1391–1402.

Yi-Chun Chen, Ngo Van Long, and Xiao Luo (2007): "Iterated strict dominance in general games," *Games and Economic Behavior*, 61, 299–315.

Martin Dufwenberg and Mark Stegeman (2002): "Existence and Uniqueness of Maximal Reductions Under Iterated Strict Dominance," *Econometrica*, 70, 2007–2023.

Itzhak Gilboa, Ehud Kalai, and Eitan Zemel (1990): "On the order of eliminating dominated strategies," *OR Letters*, 9(2), 85–89.

John Hillas and Dov Samet (2018): "Non-probabilistic correlated equilibrium as an expression of non-Bayesian rationality," a manuscript.

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

Larry Samuelson (1992): "Dominated Strategies and Common Knowledge," *Games and Economic Behavior*, 4, 284–313.

Leonard J. Savage (1954): *The foundations of statistics*, Wiley, New York.

Robert Stalnaker (1994): "On the Evaluation of Solution Concepts," *Theory and Decision*, 37, 49–73.

Robert Stalnaker (1998): "Response to Bonanno and Nehring," *Theory and Decision*, 45(3), 297–299.

Tommy Chin-Chiu Tan and Sérgio Ribeiro da Costa Werlang (1988): "The Bayesian foundations of solution concepts of games," *Journal of Economic Theory*, 45(2), 370–391.

The University of Auckland, New Zealand

Tel Aviv University, Israel