Subjective Uncertainty over Behavior Strategies: A Correction[1]

Eddie Dekel

*Departments of Economics, Northwestern University, Evanston, Illinois 60201;
and Department of Economics, Tel Aviv University, Tel Aviv, Israel*
dekel@northwestern.edu

Drew Fudenberg

Department of Economics, Harvard University, Cambridge, Massachusetts 02138
dfudenberg@harvard.edu

and

David K. Levine

*Department of Economics, University of California—Los Angeles,
Los Angeles, California 90024*
dlevine@ucla.edu

In order to model the subjective uncertainty of a player over the behavior strategies of an opponent, one must consider the player's beliefs about the opponent's play at information sets that the player thinks have probability zero. This corregendum uses "trembles" to provide a definition of the convex hull of a set of behavior strategies. This corrects a definition we gave in [E. Dekel, D. Fudenberg, and D. K. Levine, 1999, *J. Econ. Theory* **89**, 165–185], which led to two of the solution concepts we defined there not having the properties we intended. *Journal of Economic Literature* Classification Numbers: C72, D82, C610. © 2002 Elsevier Science (USA)

Key Words: extensive-form games; behavior strategies; subjective uncertainty; rationalizability; self-confirming equilibrium.

473

1. INTRODUCTION

Rationalizability and related concepts are defined and characterized in terms of sets of strategies for a player that other players think he might use. In the strategic form each player's beliefs about the play of an opponent are given by a probability measure over this set, and each such measure maps to a point in the convex hull of the set of possible strategies, as in Bernheim [3] and Pearce [9]. Consequently, we can take this convex hull as a model of what players might think about other players. In extensive-form models that use behavior strategies, the correct way to model beliefs and map them to strategies is less transparent. G. Asheim has pointed out to us that in Dekel *et al.* [5] we give an incorrect definition of convex combinations of behavior strategies. As a result two of the concepts that we defined (sequential rationalizability and sequentially rationalizable self-confirming equilibrium) do not have the properties that the paper implies and intended. This corregendum uses ''trembles'' to provide a definition of mixtures that, when embedded in our definitions of sequential rationalizability and sequentially rationalizable self-confirming equilibrium, makes them function in the way we intended.[2] In particular, with the corrected definition it will be the case that when a player thinks that only a single behavior strategy is consistent with rational play by his opponent, his beliefs about that opponent correspond to that unique behavior strategy. Consequently, sequential rationalizability implies backward induction in finite games of perfect information with generic payoffs.

Instead of using trembles, Asheim and Perea [1] use lexicographic probability systems (extending Blume *et al.* [4]) to model players' beliefs in extensive-form games; among other things they use these systems to provide a correct definition of sequential rationalizability for two-player games. Battigalli [3] models beliefs in extensive-form games using Myerson's [8] conditional probability systems, to which lexicographic probability systems are closely related. We prefer to use trembles instead because we already used them in another part of our 1999 paper, and because for our purposes it is not necessary to track the relative likelihoods of various zero-probability events.[3]

2. PRELIMINARIES

To save space we will assume that the reader is familiar with most of the notation and terminology of Dekel *et al.* [5], and so we will only restate a

---

[2] As we explain below, the error does not matter for the solution concept of the rationalizable self-confirming equilibrium, which was the primary focus of that paper.

[3] Contemporaneously with this paper, Asheim and Perea developed a trembles-based alternative to their use of lexicographic probability systems.

few of the most relevant definitions. An *assessment* a_i for player i is a probability distribution over nodes at each of his information sets. A *belief* for player i is a pair $b_i \equiv (a_i, \pi_{-i}^i)$, consisting of i's assessment over nodes a_i and i's expectations of opponents' strategies $\pi_{-i}^i = (\pi_j^i)_{j \neq i}$.[4] In that paper we defined *belief closed* as follows.

DEFINITION 2.2. A belief model V is *belief closed* if for every $(\pi_i, (a_i, \pi_{-i}^i)) \in V_i$, π_j^i arises from a mixture over strategies in the set $\{\pi_j' \mid (\pi_j', b_j) \in V_j$ for some belief $b_j\}$.

This definition is silent on what it means to say that a behavior strategy "arises from a mixture" over other behavior strategies; the paper elaborates in Footnote 11, which says:

> A behavior strategy π_j is generated by a mixture $(\alpha, 1 - \alpha)$ over π_j' and π_j'' if for every π_{-j}, the distribution over terminal nodes induced by (π_j, π_{-j}) equals the $(\alpha, 1 - \alpha)$ mixture over the distributions induced by (π_j', π_{-j}) and (π_j'', π_{-j}) respectively.

This "clarification" is incorrect. The problems arise in defining the behavior of π_j^i at information sets for j that are not reachable under any of the strategies in V_j.[5] Such information sets are irrelevant for concepts that place no restrictions on play at information sets that the strategy precludes, but the mistake is important for concepts such as our sequential rationalizability that impose restrictions on play at all information sets.

For example consider the game in Fig. 1. The backwards induction profile is $\pi_1 = $ (in, up) and $\pi_2 = $ (across); the profile ((out, down), (down)) is an imperfect Nash equilibrium. Only up is sequentially rational at player 1's second information set, so the set of sequentially rational behavior strategies for player 1 must contain only strategies that play up at this information set. Consider the sets $V_1 = ($(out, up), down$)$, $V_2 = ($down, (out, down)$)$. Since the strategy (out, down) is equivalent in the strategic form to (out, up), it "arises as a mixture over" the set of player 1's strategies in V_1. Consequently, this pair is sequentially rational and belief closed when "arises from" is defined as in Footnote 11 of [5]. In particular,

---

[4] The assumption that player i's expectations about an opponent's play correspond to a strategy profile incorporates the implicit restriction that opponents randomize independently. Note that what we call an "assessment" is what Kreps and Wilson [8] call a "system of beliefs for player i" and that our "belief" is similar to what they call an "assessment."

[5] That is, an information set is unreachable under p_j if there is no profile p_{-j} for j's opponents such that the information set is reached with positive probability.
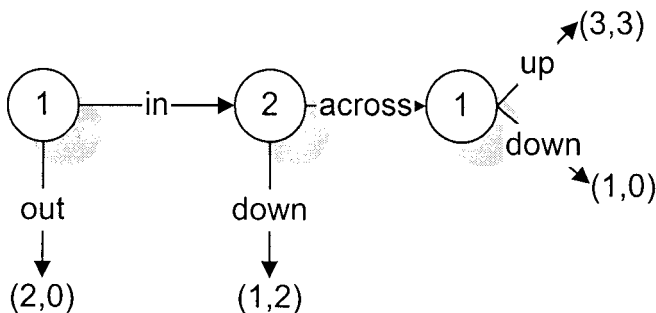
FIGURE 1

sequential rationality and belief-closed with the original definition do not imply backward induction.

3. THE EXTENSIVE-FORM CONVEX HULL

We therefore propose the following definition of the "convex hull" of behavior strategies, which corrects and builds from our previous definition by using "trembles" to make sure that every information set of player i is reachable.[6] When working with strictly positive behavior strategies, there are no unreached information sets. In this case the definition of "generated by a mixture" from our previous paper is adequate.

DEFINITION.   Strategy π_i is in the *extensive-form convex hull* of a set P_i of behavior strategies for player i if there is an integer k, strategies $\{\pi_i^j\}_{j=1,\dots,k}$ in P_i, sequences of strictly positive behavior strategies $\pi_i^{j,n} \to \pi_i^j$, and a sequence $\alpha^n \to \alpha$ of probability distributions on $[1,\dots,k]$, such that the behavior strategies π_i^n generated by the convex combination of $p_1^{1,n}$, $p_i^{2,n}, \dots, p_i^{k,n}$ with weights $\alpha^{1,n}, \alpha^{2,n}, \dots, \alpha^{k,n}$ converge to π_i.

We let α vary along the sequence so that the extensive-form convex hull will be closed. To see why the set would not be closed if the definition used only a fixed α, consider the following one-player game. Player 1 has two moves in a row: The first choice is In or Out; Out ends the game, In gives him a second choice of L or R. Strategy 1 is (Out, L), 2 is (In, R). Now suppose that the definition of the convex hull used only fixed weights, and let α^1, α^2 be the weights on strategies 1 and 2. Since only strategy 2 plays In and enables the move in the second period, the convex combination of

[6] Here and subsequently, we give the space of behavior strategies the norm topology, so that a sequence of behavior strategies converges iff it converges pointwise.

the two strategies with strictly positive weights is $((\alpha^1 \text{ Out}, (\alpha^2) \text{ In}), \text{R})$, which approaches (Out, R) as $\alpha \to 1$. However, this is not a convex combination of strategies 1 and 2, even for $\alpha = (1, 0)$. Consequently, the set of "convex combinations" by this definition is not closed. Our definition of the extensive-form convex hull includes both (Out, L) and (Out, R).

DEFINITION 2.2 (Revised). A belief model V is *belief closed* if, for every $(\pi_i, (a_i, \pi^i_{-i})) \in V_i$, π^i_j is in the extensive-form convex hull of the set $\{\pi'_j \mid (\pi'_j, b_j) \in V_j \text{ for some belief } b_j\}$.

All of the other definitions in our 1999 paper stay unchanged, modulo the change in the definition of belief closed. Note that the difference between the corrected definition of belief closed and the previous one arises when for some players i and j, π^i_j induces the same distribution over outcomes as a mixture over the set $\{\pi'_j \mid (\pi'_j, b_j) \in V_j \text{ for some belief } b_j\}$, but differs from these strategies at an information set that the strategies themselves preclude. For this reason, the changed definition of belief closed has only a minor effect on the concept of "rationalizability at reachable nodes" (Definition 2.3), as this concept does not require that strategies be optimal at information sets that the strategies themselves rule out. In particular, while a given belief model V (such as the pair of singleton beliefs V_1, V_2 in the example of the last section) can be rationalizable under reachable nodes under the old definition but not under the new one (because it is not belief-closed), for any such V we can construct a $\hat{V}$ that is belief-closed by adding to each V_j and to every $\pi_j \in V_j$ every strategy $\hat{\pi}_j$ that agrees with π_j at nodes that are reachable under π_j. Every added strategy is a best response at reachable nodes to the same beliefs that rationalized the original π_j, and since the set V was belief closed under the old definition, $\hat{V}$ is belief closed under the new one.[7] In particular, the change in definitions has no effect on whether a strategy profile $\hat{\pi}$ is a rationalizable self-confirming equilibrium, as this requires that there exists a belief model V that is rationalizable at reachable nodes, such that for all players i every $(\pi_i, b_i) \in V_i$ has the distribution of outcomes induced by $\hat{\pi}$. Thus Theorems 2.1 and 4.1 and all of the examples in Section 3 ([5]) are unaffected by the change.

As we noted earlier, the change in definition does matter for the concept of sequential rationalizability, which requires that strategies in the belief model be rationalized at every information set, and it has a similar impact

---

[7] Note that when a model V is rationalizable at reachable nodes under the old definition, the model formed by enlarging the set of strategies in each V_j to its extensive-from convex hull need not be rationalizable at reachable nodes under the corrected definition. Although the new model will be belief closed, the strategies introduced need not be rational at reachable nodes, and indeed they may be strictly dominated.

on the concept of a sequentially rationalizable self-confirming equilibrium. In particular, because the extensive-form convex hull of a singleton set consists solely of the single strategy in that set, the unique sequentially rationalizable profile in finite games of perfect information with generic payoffs is the one given by backward induction.[8] Despite this change, Theorem 4.2, which is the only result in [5] that refers to concepts using sequential rationality, is correct as stated, since the elaborations used in the proof have a type that is indifferent between all actions at every information set.

Note finally that even if a strategy profile π is sequentially rationalizable as a singleton set (i.e., there are beliefs b_i for each player i such that the sets $V_1 = \{(\pi_1, b_1)\}$, $V_2 = \{(\pi_2, b_2)\}$, etc., are sequentially rationalizable), it need not be a sequential equilibrium. While we have assumed that each player's assessment over nodes in his information sets is consistent in the Kreps–Wilson sense of being derivable from the limit of Bayesian beliefs from full-support strategies, we have not required that all players' assessments be consistent with a *single* sequence of totally mixed strategy profiles, and it is known (see for instance Example 8.5 in Fudenberg and Tirole [6]) that the freedom to use different sequences to derive each player's assessment can allow additional equilibrium outcomes.

REFERENCES

1. G. Asheim and A. Perea, Lexicographic probabilities and rationalizability in extensive-form games, mimeo, 2000.
2. P. Battigalli, Strategic independence and perfect Bayesian equilibria, *J. Econ. Theory* **70** (1976), 201–234.
3. D. Bernheim, Rationalizable strategic behavior, *Econometrica* **52** (1984), 1007–1028.
4. L. Blume, L. A. Brandenburger, and E. Dekel, Lexicographic probabilities and equilibrium refinements, *Econometrica* **59** (1991), 81–98.
5. E. Dekel, D. Fudenberg, and D. K. Levine, Payoff information and self-confirming equilibrium, *J. Econ. Theory* **89** (1999), 165–185.
6. D. Fudenberg and J. Tirole, "Game Theory," MIT Press, Cambridge, MA, 1991.
7. D. Kreps and R. Wilson, Sequential equilibrium, *Econometrica* **50** (1982), 863–894.
8. R. Myerson, Axiomatic foundations of Bayesian decision theory, Working Paper 671, Center for Mathematical Studies in Economics and Management Science, Northwestern University, 1986.
9. D. Pearce, Rationalizable strategic behavior and the problem of perfection, *Econometrica* **52** (1984), 1029–1050.

[8] Bernheim [3] defines subgame rationalizability, and argues that it yields backwards induction in generic game of perfect information.