Referee Report: JET-D-17-00281

Journal of Economic Theory

July 14, 2017

Comprehensive Rationalizability is one of the more interesting ideas related to admissibility I have read about in recent years. Its presence clarifies and raises some natural questions. To start, consider how one might naturally describe in plain language the thought process that gives rise to iterated admissibility:

1. All players are cautious (i.e., put positive probability on "all" strategies).
2. All players maximize expected utility.
3. Admissible strategies are exactly the strategies consistent with "rationality" when "rationality" is defined as caution plus EU maximization.
4. If Ann believes that Bob being rational is infinitely more likely than Bob being irrational, she should start by considering only Bob's admissible strategies and only consider Bob's inadmissible strategies to potentially break ties when the initial consideration yields the conclusion that she has more than one optimal strategy.

If Ann is also rational, then this sort of lexicographically tiered consideration of Bob's admissible and inadmissible strategies implies that she plays a 2-admissible strategy (i.e., one that survives two rounds of deletion of inadmissible strategies). Of course, I am being intentionally imprecise and misleading.

So what is wrong here? The admissible strategies are indeed the *rational* strategies in this context. In the plain-language sense of the word, admissible strategies can be "rationalized" by caution plus EU maximization. However, inadmissible strategies are not the only *irrational* strategies unless the game is trivial![1] That is because a player may choose one admissible strategy *irrationally* when another admissible strategy maximizes EU. In essence, the behavioral implication of irrationality is that "anything goes". So we should amend the description above as follows:

> If Ann believes that Bob being rational is infinitely more likely than Bob being irrational, she should start by considering only Bob's admissible strategies and only *expand her consideration to include* Bob's inadmissible strategies to potentially break ties when the initial consideration yields the conclusion that she has more than one optimal strategy.

[1] On the other hand, inadmissible strategies *are* exactly the strategies that cannot be "rationalized" by caution plus EU maximization.

In other words, she should initially consider a primary hypothesis about all of Bob's admissible strategies and when that results in indecision on her part, move onto a secondary hypothesis about all of Bob's strategies.

However, it is also prima facie very natural (and some might even say more natural) that Ann should start by considering only Bob's admissible strategies and only consider Bob's inadmissible strategies to potentially break ties when the initial consideration yields the conclusion that she has more than one optimal strategy. In fact, that is exactly the kind of thought process captured by *comprehensive rationalizability*. To me, this is the very intriguing starting point of the ideas explored in this paper.

The way I see it, IA (iterated admissibility) and CR (comprehensive rationalizability) could be representative of two opposing philosophical views on the genesis of epistemology. Consider the special status occupied by strategies in the description of states of the world in epistemic games. Strategies are acts in the sense of Savage. These acts are *measurable with respect to the opponents' strategies only* and not other aspects of the states of the world (e.g., opponents' beliefs). With that it mind, I would distinguish the thought processes associated with IA and CR thusly:

An "IA thinker" can and does ascribe multiple levels of rationality to the opponents' IA strategies. Even though Ann might believe that Bob will play an IA strategy because he is hyper rational, she nevertheless does not neglect the possibility—though infinitesimal it may be—that Bob will play that same IA strategy when he is irrational. Due to the inclusion of such infinitesimal possibilities, we might say that the genesis of Ann's beliefs does not seek to explain Bob's strategies but seeks to rank the various explanations of Bob's strategies. On the other hand, a "CR thinker" ascribes a single level of rationality (the maximal one) to the opponents' CR strategies. Due to the exclusion of even infinitesimal possibility of CR strategies as outputs of less-than-hyper rationality, we might say that Ann does seek to explain Bob's strategies (without even infinitesimal hedging) in the process of forming her beliefs. Analogous differences also manifest themselves at lower levels of rationality.

This should not be taken to mean that either approach is superior to the other in some objective way, but rather that there are objective differences over which some subjective preferences may exist. Nevertheless, I disagree with the interpretation offered in **Section 5** of the paper for those reasons. I have quoted the relevant part below:

> The "strategic" in the term "strategic assumption" is the following. We do not only want sufficiently rational types of a player to believe that other players are rational and play rational strategies but also that if a rational strategy is played, it is played for rational reasons. Loosely speaking, if a rational type of a player finds a new manuscript of Dostojevsky novel written on a computer, we want him to believe that it was a Dostojevsky who wrote that novel and not a monkey that just randomly played with the keyboard of the computer typing that manuscript by chance.

First of all, what is said above does not follow from the definition of "strategic assumption" but rather from the fact that "strategic assumption" is applied within the restricted context of epistemic models such that marginal beliefs on strategies are always "lexicographic conjectures" as defined in this paper. The paper calls these epistemic models "lexicographic beliefs type spaces". In other words, a major part of what makes strategic assumption reflect the logic of CR is coming from the way in which these "lexicographic beliefs type spaces" are different from the lexicographic type spaces used by—for example—Brandenburger-Friendenberg-Keisler (2008). This is a subtle point, but one worth making more explicit.

How is this done? For the sake of convenience, I will restrict my following explanation to finite spaces. The sequence of probability measures that make up a lexicographic belief in BFK's type spaces have disjoint supports. The type spaces in this paper add another requirement. The sequence of corresponding marginal probability measures on the opponents' strategies must also have disjoint supports. What this effectively means is that if Ann gives positive weight to two states in which Bob plays b, then those two states must be given positive probability by the exact same measure in the sequence of measures that makes up her lexicographic belief. It follows that, if Bob plays b rationally in one of those states and irrationally in the other, the only way for Ann to consider it infinitely more likely that Bob plays b rationally rather than irrationally is for her to give absolutely zero (not even infinitesimal!) weight to Bob playing b irrationally.

Furthermore, if we are "loosely speaking" as the authors write in the quoted portion, the logic of IA as expounded upon by BFK also says that if a rational type of a player finds a new manuscript of a Dostojevsky novel, she must believe that Dostojevsky and not some monkey wrote it. The subtle difference is that the logic of IA requires that the player to assign infinitesimal but non-zero probability to the monkey scenario while the logic of CR requires that not even infinitesimal probability can be assigned to the monkey scenario. I point this out because, as discussed at the beginning of this report, the logic of IA and CR are quite similar when we are loosely speaking.

Iterative strategy sorting (as opposed to deletion). One of the fascinating implications of these differences is that CR cannot be described by any iterative strategy deletion algorithm that is *memoryless* in the sense that only the reduced game that remains after the last round is needed to perform the next round of deletions. Both iterated admissibility and iterated strong undominance are memoryless in this sense. The first example game provided in the paper is sufficient to demonstrate this.

	a	b	c
x	4,0	4,1	0,2
y	0,0	0,1	4,2
z	3,0	2,2	2,1

The only deletion possible in the first round is a, after which we are left with the following reduced game. In this reduced game, z is optimal only when there is equal weight placed on b and c. However, when z is optimal, the strategies x and y are also optimal.

	b	c
x	4,1	0,2
y	0,1	4,2
z	2,2	2,1

CR breaks this three-way tie by considering *only* the previously deted column of the game matrix, which says that z cannot win the tie breaker in the only scenario where it has a chance to be optimal.

	a
x	4,0
y	0,0
z	3,0

Given that the strategy deletion algorithms that make up the core of this literature are all memoryless in the sense described earlier, I posit that CR belongs in a separate category. I suggest that a more appropriate terminology for categorizing CR is iterative strategy demotion (as opposed to deletion). The strategies capturing higher and higher orders of rationality are iteratively promoted to a lexicographically higher tier in the beliefs. However, the demoted strategies are still needed to continue the next round of promotions.

As the use of lexicographic probabilities increases in epistemic game theory, I suppose that other interesting iterative sorting/promotion algorithms may be conceived by researchers in the field. I point this out for one reason. CR, like Bernheim and Pearce's rationalizability, is a procedure that iteratively eliminates beliefs. However, lexicographic beliefs by their very nature, allows the procedure to "remember" what happened in the previous stages by using extra probability measures. I have no idea whether this is significant at some more fundamental level, but it seems to be at the very least a technical curiosity and yet another demonstration of the extra power afforded by the use of lexicographic beliefs.

Final Comments. Comprehensive rationalizability is a very interesting new idea that deserves some further consideration precisely because of its relation to iterated admissibility. I began this report by pointing out that an informal explanation of how a player might reason toward IA strategies actually lead her to choose CR strategies. It is definitely worth sorting out why exactly that is so. Obviously, I disagreed with some of the explanations and interpretations offered in the paper. While I would like to see these points addressed in the paper, I do not see these disagreements as material to my final judgment on the paper.

After all, I suspect that others in the field will address the points I mentioned in future papers.

Technically, the paper is quite straightforward for a researcher familiar with the literature and the example games make at least the highlights accessible to the general audience. That said, all epistemic game theory papers that use lexicographic beliefs are necessarily heavy in notation. For that reason, I would appreciate it if the authors adopted the kind of notational convention used in BFK and subsequent papers (as opposed to $\bar{\Delta}$, $\ddot{\Delta}$ and the like). It would help reduce the fixed cost needed to read a new paper in this literature.