

Epistemic Permissiveness and Symmetric Games

Brian Weatherson

1 Introduction

Permissivism in epistemology is a family of theses, each of which says that rationality is compatible with a number of distinct attitudes. This paper argues that thinking about symmetric games gives us new reason to believe in permissivism. In some finite games, if permissivism is false then we have to think that a player is more likely to take one option rather than another, even though each have the same expected return given that player's credences. And in some infinite games, if permissivism is false there is no rational way to play the game, although intuitively the games could be rationally played. The latter set of arguments rely on the recent discovery that there are symmetric games with only asymmetric equilibria. It was long known that there are symmetric games with no pure strategy symmetric equilibria; the surprising new discovery is that there are symmetric games with asymmetric equilibria, but no symmetric equilibria involving either mixed or pure strategies.

The permissivist theses that have been the focus on recent philosophical attention vary along two dimensions.¹

The first dimension concerns what we hold fixed when we say that multiple attitudes are rationally permissible. The weakest possible theory just says that two people with distinct attitudes may both be rational. No one really denies this. The strongest theory says that holding every fact about a situation constant, there are two possible rational attitudes. In between we have a number of interesting theses. For instance, we can ask whether multiple attitudes are rationally compatible holding constant the evidence the believer has. And we can ask whether multiple attitudes are rationally compatible holding constant both the evidence and the believer's prior doxastic states. A classic form of subjective Bayesianism answers *yes* to the first question, and *no* to the second. The focus here will be on a thesis very close to the strongest one - whether two people who are alike in all qualitative respects can rationally have different attitudes.

The second dimension concerns whether the folks holding these distinct attitudes can acknowledge that rival attitudes are rational. Some permissivists

¹For a much more thorough introduction to the debate, and especially into the varieties of permissivist theses, see Kopec and Titelbaum (2016). Much of the setup here, including for example the use of the subjective Bayesian as an illustrative example, is from that paper.

hold that distinct attitudes can be rationally compatible with holding fixed evidence or priors or whatever, but the people holding these attitudes cannot acknowledge that attitudes other than theirs are rational. The argument I'm going to offer draws the stronger conclusion that multiple responses are rational, and rational thinkers can acknowledge that alternative responses to theirs are rational.

The negation of a permissive thesis is a Uniqueness thesis. The name suggests that there is precisely one rational attitude to take in a specified situation, but we'll interpret it as the view that there is at most one rational attitude to take so as to ensure each Uniqueness thesis is the negation of a permissive thesis. As with Permissivism, Uniqueness comes in weaker and stronger varieties. The strongest version is the literally incredible view that there is only one doxastic attitude that is rationally permissible. (Presumably it is the view that is certain of all and only truths.) The weakest version, which is still interesting, is that once a situation is described in full detail, there is precisely one doxastic attitude that is rationally permissible. Everyone holds that, since the normative supervenes on the descriptive, that describing a situation in full detail fixes which doxastic states are rationally permissible. The Uniqueness theorist adds the claim that there are 0 or 1 such states.²

As I said, I'm interested in defending a strong, but not maximally strong, version of Permissivism. Equivalently, I'm interested in attacking a weak, but not quite maximally weak, version of Uniqueness. Here is the version of Uniqueness that I want to reject.

- For all kinds K , and evidence E , there is at most one credal distribution that is rational for an agent of kind K with evidence E .

I mean to be fairly liberal over what counts as a 'kind', so if any such theory is false, we have proven that a lot of Uniqueness theories are false. So a kind could be a prior probability function, a set of privileged predicates that one uses for induction, an attitude to inductive risk, and so on. The only assumption I'll make is that kinds are shareable; so there is no such thing as the kind *Being John Malkovich*. In principle we could say that what evidence

²Two small caveats here. Uniqueness theorists may say that it is permissible to not have any attitude towards a proposition. So it is consistent with Uniqueness, as I understand it, to say that it could be both rational to have credence 0.6 in p , and rational to not have any attitude towards p . What the Uniqueness theorist denies is that there are distinct credences towards p one could adopt, each of which would be rational. And of course the Uniqueness theorist thinks it could be rational to have credence 0.6 in p and 0.7 in q . When I say Uniqueness implies that just one state is rational, I mean to quantify over complete credal states, not attitudes towards single propositions.

one has is part of one's kind, but the discussion below will be clearer if we separate out evidence and kinds.

The next two sections set out two symmetric games where Uniqueness leads to surprising results. I think the results are surprising, indeed implausible, enough that we should reject Uniqueness. But even if one doesn't accept that, it's still interesting to see what Uniqueness entails. In all cases we'll assume that the following things are common knowledge among players of the game.

- Each player is rational, so they form rational credences, and maximise expected utility.
- Each player is of kind K .
- Each player knows the payout structure of the game.
- Each player is self-aware; they know their own credences.
- If Uniqueness is true, then each player knows that Uniqueness is true.
- Each player has no other relevant evidence about the game or the players.

Let evidence E , unless otherwise stated, be the evidence specified by those six bullet points. We'll be continually thinking about propositions of the form:

- A rational agent of kind K with evidence E will perform action φ in this game.

Since the games are symmetric, we don't have to ask about which player will make this move; we can think abstractly about what any rational player would do.

2 Chicken

Some finite symmetric games have no symmetric pure-strategy equilibria. One notable example is Chicken. Here's a version of Chicken that will do.³

The symmetric pure-strategy pairs $\langle \text{Stay}, \text{Stay} \rangle$ and $\langle \text{Swerve}, \text{Swerve} \rangle$ are not equilibria; in each case both parties have an incentive to defect. But the game does have a symmetric equilibria. It is that both players play the

³When presenting games in this format, I'll write Row's payout first, then Column's payout. So the top right cell here, for example, says that if Row plays Stay and Column plays Swerve, then Row gets a payout of 1, and Column gets a payout of -1. All payouts are in utils unless otherwise stated. For people unfamiliar with it, the backstory of Chicken is that the players are in cars driving towards each other on a one-lane road. They can stay on the road, possibly winning points for courage and possibly dying, or swerve off.

	Stay	Swerve
Stay	-100, -100	1, -1
Swerve	-1, 1	0, 0

mixed strategy of Stay with probability 0.01, and Swerve with probability 0.99.

Let **Swerve** be the proposition that a rational player of kind K with evidence E will Swerve. And call the players Row and Column. Given our assumptions so far, plus Uniqueness, we can prove that Row's credence in **Swerve** is 0.99. Here's the proof.

1. Let x be Row's credence in **Swerve**.
2. By self-awareness, Row knows that x is her credence in **Swerve**.
3. Since she knows she is rational, Row can infer that x is a rational credence in **Swerve**.
4. Since she knows Uniqueness is true, Row can infer that x is the only rational credence in **Swerve**.
5. Since she knows Column is rational, she can infer that x is Column's credence in **Swerve**.
6. Since all the assumptions so far are common knowledge, she can infer that Column knows that x is her credence in **Swerve**.
7. If $x = 1$, then Row can infer that it is rational for Column to Swerve, while knowing that Row will also Swerve. But this is impossible, since if Column knows Row will Swerve, it is best to Stay. So $x \neq 1$.
8. If $x = 0$, then Row can infer that it is rational for Column to Stay, while knowing that Row will also Stay. But this is impossible, since if Column knows Row will Stay, it is best to Swerve. So $x \neq 0$.
9. So $0 < x < 1$.
10. Since Row knows Column's credence that Row will Swerve (whatever it is), and Row knows Column is rational, but Row does not know what Column will do, it must be that Column is indifferent between Stay and Swerve given her credences about what Row will do.
11. Column is indifferent between Stay and Swerve only if her credence that Row will Swerve is 0.99. (This is a reasonably simple bit of algebra to prove.)
12. So from 10 and 11, Column's credence that Row will Swerve is 0.99.
13. By (known) Uniqueness, it follows that the only rational credence in **Swerve** is 0.99.
14. So since Row is rational, it follows that $x = 0.99$.

Now there is nothing inconsistent in this reasoning. In a sense, it is purely textbook reasoning. But the conclusion is deeply puzzling. We've proven that Column is indifferent between her two options. And we've proven that Row knows this. But we've also proven that Row thinks it is 99 times more likely that Column will choose one of the options over the other. Why is that? It isn't because there is more reason to do one than the other; given Column's attitudes, the options are equally balanced. It is purely because Uniqueness pushes us to a symmetric equilibrium, and this is the only symmetric equilibrium.

I don't think this result will convince many devotees of Uniqueness to give up their view. It's not a particularly novel claim that rational players will end up at the unique Nash equilibrium of a game. And to be sure, if this game was being played repeatedly, much weaker assumptions entail that each player should Stay 1% of the time, with those Stays being randomly distributed across the plays of the game. But it is still odd, at least to me, to see the same conclusion drawn in the single shot game, where each player is known to be indifferent between their choices.

The next case is I think much worse for Uniqueness.

3 Elections

The cases that really inspired this paper come from some recent work on this rather old question,

If a symmetric game has an equilibrium, does it have a symmetric equilibrium?

Over the years, a positive answer was given to various restricted forms of that question. But recently it has been proven that the answer to the general question is no. Mark Fey (2012) showed that there are symmetric positive-sum two-player games that have only asymmetric equilibria.⁴ And Dimitrios Xefteris (2015) showed that there is a symmetric three-player zero-sum that has only asymmetric equilibria. In fact, he showed that a very familiar game, a version of a Hotelling–Downs model of elections, has this property. Here's how he describes the game.

Consider a unit mass of voters. Each voter is characterized by her ideal policy. We assume that the ideal policies of the voters are uniformly distributed in $[0, 1]$. We moreover assume that

⁴Fey also includes a nice chronology of some of the proofs of positive answers to restricted forms of the question.

three candidates A , B and C compete for a single office. Each candidate $J \in \{A, B, C\}$ announces a policy $s_J \in [0, 1]$ and each voter votes for the candidate who announced the policy platform which is nearest to her ideal policy. If a voter is indifferent between two or among all three candidates she evenly splits her vote between/among them. A candidate $J \in \{A, B, C\}$ gets a payoff equal to one if she receives a vote-share strictly larger than the vote-share of each of the two other candidates. If two candidates tie in the first place each gets a payoff equal to one half. If all three candidates receive the same vote-shares then each gets a payoff equal to one third. In all other cases a candidate gets a payoff equal to zero. (Xeferis, 2015, 124)

It is clear that there is no symmetric pure-strategy equilibrium here. If all candidates announced the same policy, everyone would get a payoff of $1/3$. But no matter what that strategy is, if B and C announce the same policy, then A has a winning move available.

What's more surprising, and what Xeferis proves, is that there is no symmetric mixed strategy equilibria either. Again, in such an equilibrium, any player would have a payoff of $1/3$. Very roughly, the proof that no such equilibrium exists is that random deviations from the equilibrium are as likely to lead to winning as losing, so they have a payoff of roughly $1/2$. So there is no incentive to stay in equilibrium. So no symmetric equilibrium exists.

But if Uniqueness is true, and if it is possible to play the game under circumstances of common knowledge of rationality and kind, then there must be a symmetric equilibrium. The reason is that a version of the proof of the previous section still goes through. Whatever credal distribution A has over B 's possible policies, A must also have over C 's policies (since they both adopt the uniquely rational strategy), and she must know that B and C each have over each other's policies and over hers, and these distributions must be consistent with each player having these credal distributions while thinking that the other players have the same distributions and are maximising expected utility. In other words, the assumptions we've made about the game imply that A has a credal distribution F over B 's possible policies only if the mixed strategy triple where each player adopts F as their mixed strategy is itself an equilibrium. And that would be a symmetric equilibrium. But no symmetric equilibrium exists.

But if we drop Uniqueness, it is possible to keep all the other assumptions. As Xeferis points out, the game has asymmetric equilibria. Here is one possible model for the game.

- A plays 0.6 (and wins), B and C each play 0.4 (and lose).
- Each player has a correct belief about what the other players will play.
- But both B and C know they cannot win given the other player's moves, so they pick a move completely arbitrarily.
- Further, each player has a correct belief about why each player makes the move they make.

This is the coherent equilibria that Xefteris describes, but note that it is rather implausible that we'd end up there in a real-life version of the game. It requires two of the players to know that one of the other players will be indifferent between their options, but from this draw a correct inference about what they will do. That's not particularly plausible. So let's note that there is a somewhat more plausible way to get all three players to make those moves.

- A plays 0.6 (and wins), B and C each play 0.4 (and lose).
- The only two rational plays are 0.4 and 0.6, but each of them is permissible.
- In any world that a player believes to be actual, or a player believes another player believes to be actual, or a player believes another player believes another player believes to be actual, etc., the following two conditions hold.
- If a player plays 0.6, they believe the other two players will play 0.4, and hence playing 0.6 is a winning move.
- If a player plays 0.4, they believe the other two players will play 0.6, and hence playing 0.4 is a winning move.

The main difference between this model and Xefteris's is that it allows that players have false beliefs. But why shouldn't they have false beliefs? All they know is that the other players are rational, and rationality (we're assuming) does not settle a unique verdict for what players will do.⁵ So I think this strategy set, where the players have rational (but false) beliefs about the other players, is more useful to think about.

4 Objections

The reductio arguments here have all assumed not just that Uniqueness is true, but that the players know that it is true. What happens if we drop that

⁵To use the game-theory jargon, Xefteris describes a Nash equilibrium of the game, but what I've described is a rationalizable strategy triple (Bernheim, 1984; Pearce, 1984). If Uniqueness is true, then strictly speaking any rationalizable strategy pair for a symmetric game is a Nash equilibrium.

assumption, and consider the possibility that Uniqueness is true but unknowable?

This possibility is a little uncomfortable for philosophical defenders of Uniqueness. If the players in these games do not know that Uniqueness is true, then neither do the authors writing about Uniqueness. And now we have to worry about whether it is permissible to assert in print that Uniqueness is true. I wouldn't make too much of this though. It is unlikely that a knowledge norm governs assertion in philosophical journals.

The bigger worry here is that one key argument for Uniqueness seems to require that Uniqueness is knowable. A number of recent authors have argued that Uniqueness best explains our practice of deferring to rational people.⁶ For instance, Greco and Hedden use this principle in their argument for Uniqueness.

If agent S_1 judges that S_2 's belief that P is rational and that S_1 does not have relevant evidence that S_2 lacks, then S_1 defers to S_2 's belief that P . (Greco and Hedden, 2016, 373).

Similar kinds of arguments are made by Dogramaci (2012) and Horowitz (2014). But the principle looks rather dubious in the case of these games. Imagine that A forms a belief (we'll come back to how) that B believes that a rational thing to do in the Xefteris game is to play 0.6, and so she will play 0.6. She should judge that belief to be rational; as we saw it is fully defensible. But although she does not believe that she has evidence that B lacks, she should not defer to it. At least, she should not act as if she defers to it; believing that B will play 0.6 is a reason to play something other than 0.6.

And that's the general case for these symmetric games with only asymmetric equilibria. Believing that someone else is at an equilibrium point is a reason to not copy them. If it were not a reason to not copy them, then the strategy profile where each player plays the same thing would be a symmetric equilibrium. So thinking about these games doesn't just give us a rebutting defeater for Uniqueness, as described in the previous two sections, but an undercutting defeater, since they also tell against a premise that has been central to recent defences of Uniqueness.

I think there is a somewhat better move available to the Uniqueness theorist. They could simply deny that the Xefteris game, as I've described it, is even possible.⁷ This perhaps isn't as surprising as it might seem.

⁶There is a nice discussion of this argument, including citations of the papers I'm about to discuss, in Kopec and Titelbaum (2016, 195).

⁷This is really just a response to the argument based on that game; I think they just have

Note two things about the Xefteris game. First, it is an infinite game in the sense that each player has infinitely many choices. It turns out this matters to the proof that there is no symmetric equilibrium to the game. Second, we are assuming it is common knowledge, and hence true, that the players are perfectly rational. Third, we are assuming that perfect rationality entails that people will not choose one option when there is a better option available. When you put those three things together, some things that do not look obviously inconsistent turn out to be impossible. Here's one example of that.

A and B are playing a game. Each picks a real number in the open interval (0, 1). They each receive a payoff equal to the average of the two numbers picked.

For any number that either player picks, there is a better option available. It is always better to pick $\frac{x+1}{2}$ than x , for example. So it is impossible that each player knows the other is rational, and that rationality means never picking one option when a better option is available.

So the Uniqueness theorist could say that the same thing is going on in the Xefteris game. Some infinitely games cannot be played by rational actors (understood as people who never choose sub-optimal options); this is one of them. But if this is all the Uniqueness theorist says, it is not a well motivated response. We can say why it is impossible to rationally play games like the open interval game; the options get better without end. But that isn't true in the Xefteris game. The only thing that makes the game seem impossible is the Uniqueness assumption. People who reject Uniqueness can easily describe how the Xefteris game can be played by rational players. Simply saying that it is impossible, without any motivation or explanation for this other than Uniqueness itself, feels like an implausible move.

5 Conclusion

If Uniqueness is true, then the following thing happens in games between people who know each other to be the same kind, and to be rational. When someone forms a belief about what the other person will do, they can infer that this is a rational way to play the game given knowledge that everyone else will do the same thing. But sometimes this is a very unintuitive inference. In Chicken, it implies that we should have asymmetric attitudes to someone who is facing a choice between two options with equal expected value. In the

to say that in Chicken a rational player will rationally think the other player is more likely to make one of the two choices with equal expected payoffs.

election game Xefferis describes, a game that feels consistent turns out to be impossible.

I think the conclusion to draw from these cases of symmetric interactions this is that Uniqueness is false, and hence permissivism is true. Sometimes in such an interaction one simply has to form a belief about the other player, knowing they may well form a different belief about you. Indeed, sometimes only coherent way to form a belief about the other player is to believe that they will form a different belief about you. And that means giving up on Uniqueness.

References

- Bernheim, B. Douglas. 1984. "Rationalizable Strategic Behavior." *Econometrica* 52:1007–1028, doi:10.2307/1911196. (7)
- Dogramaci, Sinan. 2012. "Reverse Engineering Epistemic Evaluations." *Philosophy and Phenomenological Research* 84:513–530, doi:10.1111/j.1933-1592.2011.00566.x. (8)
- Fey, Mark. 2012. "Symmetric Games with Only Asymmetric Equilibria." *Games and Economic Behavior* 75:424–427, doi:10.1016/j.geb.2011.09.008. (5)
- Greco, Daniel and Hedden, Brian. 2016. "Uniqueness and Metaepistemology." *The Journal of Philosophy* 113:365–395, doi:10.1111/phc3.12318. (8)
- Horowitz, Sophie. 2014. "Immoderately Rational." *Philosophical Studies* 167:41–56, doi:10.1007/s11098-013-0231-6. (8)
- Kopec, Matthew and Titelbaum, Michael G. 2016. "The Uniqueness Thesis." *Philosophy Compass* 11:189–200, doi:10.1111/phc3.12318. (1, 8)
- Pearce, David G. 1984. "Rationalizable Strategic Behavior and the Problem of Perfection." *Econometrica* 52:1029–1050, doi:10.2307/1911197. (7)
- Xefferis, Dimitrios. 2015. "Symmetric Zero-Sum Games with Only Asymmetric Equilibria." *Games and Economic Behavior* 89:122–125, doi:10.1016/j.geb.2014.12.001. (5, 6)