Games and the Reason-Knowledge Principle

Brian Weatherson

John Hawthorne and Jason Stanley (2008), defend what they call the "Reason-Knowledge Principle".

> Where one's choice is p-dependent, it is appropriate to treat the proposition that p as a reason for acting iff you know that p. (578)

There have been many attempts in the literature to show that this leads to implausible *actions*. As Jonathan Ichikawa (2012) shows, most of these attempts rest on further, and arguably false, assumptions about the connection between reasons and action. Relatedly, most of those responses concern the role of knowledge and reasons in decision-making. I'll argue that we can formulate a sharper problem for the principle if we focus on game-playing, and say exactly which extra assumptions we are making.

The Reason-Knowledge Principle should have the following implications, at least for cases where S's aim is to produce the best outcome.

(1) If S knows that φ and ψ will produce the same outcome, and S must choose φ or ψ, then it is rationally permissible for S to choose ψ.
(2) If S knows that φ and ψ will produce the same outcome if p, and φ will produce a better outcome if $\neg p$, then it is rationally permissible for S to choose ψ iff she knows p.

The point of (1) is that S can use her knowledge that φ and ψ will produce the same outcome to justify making an arbitrary choice between φ and ψ. And the point of (2) is that the Reason-Knowledge Principle suggests only knowledge that p could justify ignoring the fact that ψ does worse than φ if $\neg p$.

Define a **symmetric** game as having these features:

- The game is purely co-operative; each player gets the same payoffs;
- Each player knows nothing about the other save that it is common knowledge the players are rational, and hence know what each other's rational requirements are;
- Each player has the same moves available; and,
- The payoffs are a function of just which moves are made, not of who makes them.

Assume A and B are playing a symmetric game, and it is common knowledge which symmetric game they are playing. Then the following premise seems hard to dispute:

(3) It is rationally required for A to play φ iff A knows B will play φ.

† Penultimate draft only. Please cite published version if possible. Final version published in *The Reasoner* 6: 6-7.

What makes (3) so compelling is that we can derive it from (4), (5) and (6).

(4) *A* knows that *B* will play φ iff *A* knows that any rational player will play φ.
(5) If *A* knows any rational player will play φ, then *A* is rationally required to play φ.
(6) If *A* is rationally required to play φ, then *A* knows that any rational player will play φ.

We get (4) from the fact that *A* knows nothing about *B* save that she is rational. We get (5) by the factivity of knowledge. And we get (6) by the requirement that the players are rational, and hence know what rationality requires of each player. And these three together entail (3). So (3) is true, and (1) and (2) are entailed by the Reason-Knowledge Principle. Unfortunately, (1), (2) and (3) are inconsistent, as we'll now show.

Informally, in this game *A* and *B* must each play either a green or red card. I will capitalise *A*'s moves, i.e., *A* can play GREEN or RED, and italicise *B*'s moves, i.e., *B* can play *green* or *red*. If two green cards, or one green card and one red card are played, each player gets $1. If two red cards are played, each gets nothing. Each cares just about their own wealth, so getting $1 is worth 1 util. All of this is common knowledge. More formally, here is the game table, with *A* on the row and *B* on the column.

	green	*red*
GREEN	1, 1	1, 1
RED	1, 1	0, 0

Assume *A* knows *B* will play *green*. By (3), it is rationally required that *A* plays GREEN. But *A* can use this knowledge of *B* to deduce that GREEN and RED have the same payoff. So by (1), it is rationally permissible to play RED. Contradiction.

Now assume *A* does not know *B* will play *green*. By (3), it is not a rational requirement that *A* plays GREEN. But *A* knows that GREEN does better than RED unless *B* plays *green*. And since she doesn't know *B* plays *green*, by (2), she's required to play GREEN. Contradiction.

So either assuming that *A* does or does not know that it is rationally required for *B* to play *green* leads to a contradiction given (1), (2) and (3). So these three premises are inconsistent. Since (3) is true, that means (1) or (2) is false. And since the Reason-Knowledge principle entails those two premises, one of which is false, the Reason-Knowledge Principle is false.

I'm not entirely sure which of (1) and (2) is false; both of them do feel plausible. I suspect the problem is (1). Assume *A* deduces from premises she believes that rational players will play a green card. Perhaps she agrees with Robert Stalnaker (1998) that rationality requires avoiding weakly dominated options. Then she knows it doesn't matter to her outcome whether she plays GREEN or RED; she will get $1 either way. But if she plays RED, she is incoherent; she is doing something she thinks no rational player does. And perhaps this incoherence is a bad thing in itself. Niko Kolodny (2005) argues that incoherence is not bad in itself; Jacob Ross (2012)

argues that it is. The suggestion that (1) is the false premise favours Ross's view over Kolodny's. But this conclusion is very speculative; the main thing I wanted to note was the problem this game raises for the Reason-Knowledge Principle.

References

Hawthorne, John and Stanley, Jason, (2008). "Knowledge and Action." *Journal of Philosophy* 105: 571-590, doi:10.5840/jphil20081051022. (1)

Ichikawa, Jonathan, (2012). "Experimentalist pressure against traditional methodology." *Philosophical Psychology* 25: 743-765, doi:10.1080/09515089.2011.625118. (1)

Kolodny, Niko, (2005). "Why be Rational?" *Mind* 114: 509-563, doi:10.1093/mind/fzi509. (2)

Ross, Jacob, (2012). "All Roads Lead to Violations of Countable Additivity." *Philosophical Studies* 161: 381-390, doi:10.1007/s11098-011-9744-z. (2)

Stalnaker, Robert, (1998). "Belief revision in games: forward and backward induction." *Mathematical Social Sciences* 36: 31-56, doi:10.1016/S0165-4896(98)00007-9. (2)