

There are replicas and replicas

Esteban Céspedes

e.cespedes@stud.uni-frankfurt.de

Abstract

In a recent article, José Luis Bermúdez challenged David Lewis's argument about Newcomb's problem (NP) and the prisoner's dilemma (PD) being the same. I show briefly that Bermúdez's counterargument is not sound and that Lewis's original position is correct.

Keywords: Prisoner's dilemma, Newcomb's problem, Decision theory.

According to Bermúdez's (2013) clear reconstruction of Lewis's argument (1979), the following is a fundamental assumption in NP (using Bermúdez's numeration):

(3*) I will receive \$1,000,000 \leftrightarrow It is predicted that I do not take my \$1,000

In a modified version of the PD, which involves exchanging years of prison for bail payments, the following is a fundamental assumption:

(3**) I will receive \$1,000,000 \leftrightarrow you do not take your \$1,000.

Under a general notion of predictive processes, the following assumption permits interpreting both situations as the same general problem:

(A) It is predicted that I do not take my \$1,000 \leftrightarrow you do not take your \$1,000.

The first step in Bermúdez's counterargument is the claim that since the epistemic situation of the players is a special feature of NP, assumption (A) should be reformulated under a high confidence operator (C_p) as follows:

(5) C_p (It is predicted that I do not take my \$1,000 \leftrightarrow you do not take your \$1,000)

This seems to be correct for both NP and the PD. But Bermúdez's second step is to show that holding (5) in a problem already characterised as a PD implies that such a problem is not a PD. He argues that to believe in (5), the agent must have a low degree of belief in the following bi-conditionals:

(6) I take my \$1,000 \leftrightarrow you do not take your \$1,000

(7) I do not take my \$1,000 \leftrightarrow you take your \$1,000

According to Bermúdez, this would imply that Lewis's argument is self-defeating, since the description of the PD would leave only two scenarios open in the extreme case. It seems that if one attributes a low degree of belief to claims (6) and (7), one should also hold the following:

(8) I take my \$1,000 \leftrightarrow you take your \$1,000

(9) I do not take my \$1,000 \leftrightarrow you do not take your \$1,000

Nevertheless, (8) or an epistemic version thereof does not follow from (A), nor from its epistemic formulation (5). That the actions of an agent's replica show a way of predicting the agent's actions does not imply that his actions are the same as those of his replica—i.e. that they represent a reliable prediction. Lewis's (1979: 238) already warned of this, explaining that some replicas have more predictive power than do others. The average degree of belief of claims like (8) and (9) should exceed a low standard of reliability (about 0.5005) if the problem is a genuine disagreement about rationality. (Lewis uses conditional probabilities instead of biconditionals, but such details must not be considered here.) Thus, Bermúdez is right in saying that an assumption like (8) would imply that the game considered as both a NP and a PD is not really a PD, but he is wrong in saying that such an assumption follows from (5) or from anything in Lewis's argument. The problem can still be considered a PD by maintaining (5) and without having to maintain (8). It should be also noticed that Lewis's argument is not *indirectly* self-defeating either, because the confidence operator is not more relevant to the PD than it is in NP.

Now suppose, for the sake of the argument, that the agent is actually committed, as Bermúdez says, to have a high degree of belief in a claim like (8). The problem considered would not be a NP either, which is another reason why Bermúdez's intent of separating NP and the PD fails. Since, according to assumption (A) in the general

problem, the other prisoner's action is interpreted as a predictive process, one could state the following in NP:

(8*) I take my \$1,000 \leftrightarrow It is predicted that I take my \$1,000

This would also transform the problem into a situation that lacks the outcome scenarios that make NP so puzzling. Of course, it would be true only in the extreme situation, which is the case neither for a traditional formulation of the NP nor for any real-life decision problem.

References

- Bermúdez, J. L. (2013). Prisoner's dilemma and Newcomb's problem: why Lewis's argument fails. *Analysis* 73 (3):423-429.
- Lewis, D. (1979). Prisoners' dilemma is a newcomb problem. *Philosophy and Public Affairs* 8 (3):235-240.

This work was supported by CONICYT and by the Konrad Adenauer Foundation.