Philosophia (2023) 51:2543–2557
https://doi.org/10.1007/s11406-023-00694-1

Prisoner's Dilemma and Newcomb's Problem: Two Problems or One?

Emil Badici[1]

Received: 9 January 2023 / Revised: 15 September 2023 / Accepted: 11 October 2023 /
Published online: 3 November 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract

David Lewis argued that Newcomb's Problem and the Prisoner's Dilemma are "one and the same problem" or, to be more precise, that the Prisoner's Dilemma is nothing else than "two Newcomb problems side by side" (Lewis Philosophy and Public Affairs 8:235–240, 1979: 235). It has been objected that his argument fails to take into account certain epistemic asymmetries which undermine the one-problem thesis. Sobel (1985) acknowledges that many tokens satisfy the structural requirements of both problems, while questioning the generality of the thesis. Bermúdez (Analysis 73:423–429, 2013), on the other hand, argues that there is a deeper structural conflict between the two problems in the sense that the epistemic requirements that give Newcomb's Problem its force are precisely those that prevent it from being a Prisoner's Dilemma. I argue that the epistemic asymmetry objections raised by Sobel and Bermúdez fail to undermine the one-problem thesis. A different type of objection raised by Bermúdez, one which relies on the contrast between parametric and strategic choices, fares no better. Although NP is a problem that has been primarily studied for its implications to decision theory, it contains enough game theoretic elements to justify the claim that a juxtaposition of two such problems can be thought of as a strategic game.

Keywords Newcomb's Problem · Prisoner's Dilemma · Decision theory · Game theory

David Lewis argued that Newcomb's Problem and the Prisoner's Dilemma are "one and the same problem" or, to be more precise, that the Prisoner's Dilemma is nothing else than "two Newcomb problems side by side" (Lewis, 1979: 235). According to this view, call it 'the one-problem thesis', the differences in the way the nominally

✉ Emil Badici
emil.badici@tamuk.edu

[1] Texas A&M University-Kingsville, Department of History, Political Science and Philosophy, Kingsville, TX, USA

two problems are typically set up are merely superficial, while their structure is essentially the same. If true, the one-problem thesis is very important because it can be used to show that problems lending support to causal decision theory are not only possible but ubiquitous. Moreover, it would legitimize the exploration of common structural patterns and solutions both in the one-shot case and in the iterated versions of these problems. It has been objected that the argument offered by Lewis fails to take into account certain epistemic asymmetries which undermine the one-problem thesis. Sobel (1985) acknowledges that many tokens satisfy the structural requirements of both problems, while questioning the generality of the thesis. Restricting discussion to what he calls "near-certainty" problems, he argues that although some Prisoner's Dilemmas are Newcomb Problems,[1] not all of them are. Bermúdez (2013), on the other hand, makes a stronger claim. For him, there is a deeper structural conflict between the two problems in the sense that the epistemic requirements that give NP its force are precisely those that prevent it from being a PD. There cannot be an isomorphism between the two problems, he argues, unless the PD players have a very high degree of confidence that they are reliable replicas of one another, in which case the problem is not a genuine PD case.

I argue that the epistemic asymmetry objections raised by Sobel and Bermúdez fail to undermine the one-problem thesis. Although plausible enough, Sobel's conclusion cannot be presented as an objection against the one-problem thesis because Lewis is not actually committed to the view that all near-certainty PD's are near-certainty NP's. As far as the objection raised by Bermúdez is concerned, it is based on a misinterpretation of Lewis' argument, which does not assume, implicitly or explicitly, a high degree of confidence in the reliability of the predictor. A different type of argument proposed by Bermúdez (2015a, b, 2018), one which relies on the contrast between parametric and strategic choices, fares no better. Although NP is a problem that has been primarily studied for its implications to decision theory, it contains enough game theoretic elements to justify the claim that a juxtaposition of two such problems can be thought of as a strategic game. Thus, Lewis' one-problem thesis remains the most plausible option.

1 The Problems Explained

The most popular PD versions describe the pay-offs in terms of prison sentences. Imagine that two prisoners ("players", in the language of game theory), P_1 and P_2, who are held in separate cells and who cannot communicate with one another, are offered a deal. If both confess to an alleged crime, then both get a nine year prison sentence. If none of them confesses, both get a one year sentence. If one confesses and the other does not, then the former is set free, while the latter gets a ten year sentence. The pay-offs can be conveniently summarized in the matrix below.

[1] Henceforth, I will use 'PD' and 'NP' to refer to the aforementioned decision problems.

	P_2 declines	P_2 confesses
P_1 declines	Both get a one year sentence	P_1 gets a ten year sentence. P_2 is set free
P_1 confesses	P_1 is set free P_2 gets a ten year sentence	Both get a nine year sentence

Assuming that each of the two prisoners is rational and aware that the other has been offered the same deal, what choice should they make in order to promote self-interest? According to the dominance principle (DP),

(DP) If there is a partition of states of the world such that relative to it, action A dominates action B, then A should be performed rather than B.[2]

confessing is the rational choice for both prisoners. Since, regardless of what the other prisoner does, each prisoner is better off confessing, confessing dominates declining. The reason why the dilemma is paradoxical is that if both prisoners make what appears to be the rational choice (confess), the outcome is the third best option for each; they would have been better off (the outcome would have been the second best option) if they both had made the "irrational" choice.

NP has been introduced in the philosophical literature by Nozick (1969), who attributes it to the physicist William Newcomb. In this scenario, there are two boxes, a transparent and an opaque box, a reliable predictor and, of course, the player. The player is to make a choice between taking only the opaque box (one-boxing) and taking both (two-boxing). The transparent box contains $1,000 while the content of the opaque box is unknown. However, the player knows that the reliable predictor made a prediction about the player's choice and that the opaque box either contains $1,000,000, if the prediction made is one-boxing, or it is empty, if the prediction made is two-boxing. NP's decision matrix can be represented as follows:

	One-box prediction	Two-box prediction
P takes one box	$1,000,000	$0
P takes two boxes	$1,001,000	$1,000

Nozick uses this problem as an example of a decision problem in which the expected utility principle and the dominance principle, two principles that are otherwise in harmony with one another, are in conflict. While the dominance principle lends support to two-boxing (two-boxing dominates one-boxing), the expected

[2] Nozick (1969:118); the concept that is relevant here is that of weak dominance: action A dominates action B if for any possible outcome A is at least as good as B and for at least one possible outcome A is better than B. In the PD, confessing actually dominates declining in a stronger sense: it is strictly better for all possible outcomes.

utility principle (EUP), stated below, can be used to justify one-boxing as the rational choice.[3]

> (EUP) When faced with a choice between multiple available courses of action, one ought to perform an action with maximal expected utility.

The expected utility of an action A (assuming that there are just two possible outcomes its utility depends on, O_1 and O_2) is calculated as

$$EU(A) = p(O_1/A) \times u(O_1) + p(O_2/A) \times u(O_2),$$

where $u(O_i)$ is the utility of outcome O_i and $p(O_i/A)$ is the probability of O_i given action A. If one-boxing is taken as strong evidence in favor of the opaque box being non-empty and if this means that p(the-opaque-box-is-non-empty/one-boxing) is considerably higher than p(the-opaque-box-is-non-empty/two-boxing), then EU(one-boxing) can be considerably higher than EU(two-boxing), in which case EUP recommends one-boxing. Cases of conflict between the two principles are very important because they reveal the need to carefully distinguish between causal and evidential relations in decision theory. In fact, many authors (including Sobel, whose views are going to be discussed below) take the significance of NP to reside in the fact that it serves as an illustration of the conflict between two ways of calculating expected utility: one that tracks causality and one that tracks evidence. Nozick himself, as a two-boxer, uses NP to argue in favor of a decision theory which tracks causal relations.[4] Since the case for one-boxing makes use of evidential considerations, Nozick argues, one ought to apply DP instead of EUP and choose two-boxing. However, the problem remains controversial and the debate between one-boxers and two-boxers is ongoing.

2 David Lewis on why NP and the PD are one and the Same Problem

On the face of it, the Prisoner's Dilemma and Newcomb's Problem are quite different. The former deals with prison sentences, it involves two players rather than just one, it has nothing to do with a reliable predictor, and it says nothing about reliability. However, Lewis argues, when non-essential aspects are set aside, one can see that the problems are one and the same. The length of prison sentences in the PD, for instance, is not essential. Nor is the fact that the gains and losses are measured in terms of years spent in prison rather than amounts of money or other types of goods. What matters is that for a pay-off matrix of the form.

[3] Adapted from Nozick (1969), which offers a detailed analysis of the dominance principle and the expected utility principle in the context of Newcomb scenarios.

[4] Unlike evidential decision theory, the argument goes, causal decision theory has the resources needed to avoid the conflict. Since the player's choice plays no causal role in the prediction, p(the-opaque-box-is-non-empty/one-boxing) and p(the-opaque-box-is-non-empty/two-boxing) would be treated as equal. As a result, in causal decision theory there would be no conflict between EUP and DP.

	b_1	b_2
a_1	z	x
a_2	w	y

the players' preferences regarding the possible outcomes are ranked as follows:

P_1's ranking: $x < y < z < w$
P_2's ranking: $w < y < z < x$

What this means is that the two players, P_1 and P_2, rank their second and third best option in the same way and the first and fourth in opposite ways. It is easy to see that for an appropriate choice of the pay-offs the pay-off table of the PD can be made to closely resemble the NP's table. In Lewis's version of the PD, each player is offered \$1,000. If both take it, then they get nothing else. If both decline, each gets \$1,000,000. If only one declines, the one who declines gets nothing, while the other player gets an additional \$1,000,000. Seen from the standpoint of player P in NP and player P_1 in the PD, the pay-off tables for the two problems are indeed the same[5]:

	One-box prediction	Two-box prediction
P takes one box	\$1,000,000	\$0
P takes two boxes	\$1,001,000	\$1,000

	P_2 declines	P_2 takes the \$1000
P_1 declines	\$1,000,000	\$0
P_1 takes the \$1,000	\$1,001,000	\$1,000

Lewis (1979: 236) casts each of the nominally two decision problems in the form of three clauses, the first two of which coincide: the PD is characterized by (1), (2) and (3), while NP by (1), (2) and (3'), all quoted below.

(1) I am offered a thousand – take it or leave it.
(2) Perhaps also I will be given a million; but whether I will or not is causally independent of what I do now. Nothing I can do now will have any effect on whether or not I get my million.
(3) I will get my million if and only if you do not take your thousand.
(3') I will get my million if and only if it is predicted that I do not take my thousand.

Lewis argues that although (3) and (3') appear to be very different, each player in PD is faced with essentially the same decision problem as the NP player. His first

[5] By reasons of symmetry, the same can be said about player P_2.

step is to show that if inessential aspects of NP are eliminated, (3') can be replaced by the more general clause (3'').

> (3'') I will get my million if and only if a certain potentially predictive process (which may go on before, during, or after my choice) yields an outcome which could warrant a prediction that I do not take my $1,000.

This generalization is claimed to be legitimate because the time when the prediction is made, the assumption that the predictor is a human being (rather than a machine, for instance), the high degree of reliability,[6] and even the assumption that the prediction actually takes place, are all inessential elements of NP. The second step is to show that (3) is a special case of (3''). This can be easily done, Lewis argues, since simulation can be thought of as a special case of a predictive process. In particular, P_2, the other PD player, can be thought of as a replica of P_1 whose choice between taking and declining the $1,000 is predictive of P_1's own choice. Thus, the dilemma P_1 is confronted with is essentially an NP problem, and since P_2's circumstances are symmetrical, PD is nothing else than "two Newcomb problems side by side" (Lewis, 1979: 235).

3 Prisoner's Dilemmas which are not Newcomb Problems

Sobel argues that the view put forward by Lewis "is not quite right, for though some such Prisoner's Dilemmas are Newcomb problems, *some* are *not* " (Sobel, 1985: 264). Although he grants that there can be NP's in which the player has a fairly low degree of confidence in the reliability of the predictor and PD's in which the players have a fairly low degree of confidence in their similarity with one another, Sobel's discussion is focused on cases in which the confidence degree is very high (which he refers to as "near-certainty" cases). What he argues for is that although all near-certainty NP's are near-certainty PD's, some near-certainty PD's are not near-certainty NP's due to an asymmetry in the epistemic requirements associated with the two problems.

The reason why NP is, for Sobel, of great significance, is that it exemplifies the conflict between the causal and the evidential expected utility theory. The latter, promoted by Jeffrey (1983), is not interested in calculating the expected utility of an action under the assumption that the actions and the states are independent of one another but in calculating what Sobel calls its *news value*, as a measure of the agent's preference ranking of actions when actions and states are not independent of one another. As a result, the conflict between EUP and DP, which NP is often used to illustrate, is replaced by Sobel with the conflict between utility-maximizing and news value maximizing. Near-certainty NP's and PD's, according to Sobel, share in common the following two requirements:

[6] Wolpert & Benford (2013) agree that "the accuracy of the prediction algorithm in Newcomb's paradox … is irrelevant" (1639).

(S_1) They satisfy a matrix of possible outcomes in which a_1 and a_2 are the two possible actions, c_1 and c_2 are the two possible circumstances, and the outcomes are ranked as follows: $w > y > x > z$:

	c_1	c_2
a_1	x	w
a_2	z	y

(S_2) The circumstances are causally independent of the actions performed.

In addition to these, there is a third epistemic clause which is specific to each[7]:

(S_{3n}) In a near-certainty Newcomb Problem circumstances will be nearly maximally dependent on actions epistemically. Formally, $pr(c_{1\,/\,a_1}) \cong 1 \cong pr(c_{2\,/\,a_2})$.[8]
(S_{3p}) In a near-certainty Prisoner's Dilemma, it is nearly certain that the other will do as I do. Formally, $pr((a_1 \,\&\, c_1) \vee (a_2 \,\&\, c_2)) \cong 1$.

Formal clauses (S_{3n}) and (S_{3p}) are not equivalent to one another because the latter can be true even if the former is false. Informally, in the PD circumstances do not need to be epistemically dependent on actions in the sense required for NP. This epistemic asymmetry is the reason why, Sobel argues, although "[e]very near-certainty Newcomb's Problem is a near-certainty Prisoner's Dilemma … not every such Prisoner's Dilemma is a Newcomb Problem" (1985: 267). Among the cases of PD's which are NP's Sobel includes the case of psychological twins. Since the probability that the twin players would make the same choice is very high, (S_{3p}) is satisfied. At the same time, the similarity of their thinking processes guarantees that there is maximal epistemic dependence between the actions of the twins in the sense required by (S_{3n}). The examples of PD's which are not NP's are cases in which P_1 is nearly certain that he would make a specific choice and he could be nearly certain that the other player would make the same choice, but on independent grounds. In the case Sobel imagines, the player is nearly certain that $a_1 \& c_1$ obtains and nearly certain that $a_2 \& c_2$ does not. Nevertheless, circumstances are not epistemically dependent on actions in the way required by NP: while $pr(c_1/a_1)$ is nearly equal to 1, the value of $pr(c_2/a_2)$ is in fact very small. This means that while (S_{3p}) is true, (S_{3n}) is not. For (S_{3n}) to hold, being nearly certain that the other player will act in the same way is not enough. It also matters

[7] I leave out a fourth clause Sobel uses to define NP as it is all but guaranteed by (S_{3n}) and it does not play any role in the arguments presented here:
(S_{4n}) The news value of a_2 exceeds that of a_1.

[8] In other words, in a NP performing action a_1 is near-certain evidence that c_1 is the case and, likewise, a_2 for c_2; on the other hand, a_1 given c_2 and a_2 given c_1 are highly unlikely.

"*why* each prisoner is convinced that the other thinks in much the same way he does" (Sobel, 1985: 264).

What should one make of Sobel's argument? The claim that some near-certainty PD's are not near-certainty NP's is compelling enough, and I am not going to challenge it. However, I will argue that it does not actually undermine Lewis' thesis that all PD's are NP's. Although Lewis does not take the high degree of certainty to be an essential feature of either of the (nominally) two paradoxes, he is indeed committed to saying that there is a correlation between the degree of reliability of the predictor and the degree of similarity between the two PD players (i.e., the degree to which they are reliable replicas of one another). A PD with a very high degree of similarity between the two players is to be thought of as two NPs with a highly reliable predictor. This, however, does not mean that all near-certainty PDs are near-certainty NPs because a near-certainty PD, as defined by Sobel, is not necessarily a PD with a high degree of similarity between the two players. The high degree of similarity has to do with the similarity of reasoning processes but near-certainty is broader in the sense that it requires only similarity of decision. What makes the other player, P_2, a reliable replica of P_1 is not the fact that P_2 is likely to make the same choice but rather the similarity of their reasoning processes. Had Sobel defined near-certainty in terms of reasoning similarity, (S_{3p}) would have been no different from (S_{3n}). Thus, the existence of near-certainty PD's which are less-than-near-certainty NP's is not a counterexample to the thesis defended by Lewis.

4 Bermúdez on the Epistemic Situation of the Players

While Sobel acknowledges that some decision cases satisfy the structural requirements of both PD and NP, Bermúdez (2013) claims that there is a structural conflict which prevents a PD from being a NP. To prove this conclusion, he argues for the following two theses:

(B_1) A PD can be a NP only if the players have "knowledge that [they] are sufficiently similar for each to serve as a simulation of the other" (ibid., 428).
(B_2) When the players have the level of knowledge required by B_1, the dilemma no longer is a genuine example of PD.

The first thesis requires a longer discussion. For ease of exposition, I will first introduce, following Bermúdez, a few abbreviations:

α: I will receive $1,000,000.
β: It is predicted that I do not take my $1,000.
γ: A certain potentially predictive process (which may go on before, during, or after my choice) yields an outcome which could warrant a prediction that I do not take my $1,000.
δ: You do not take your $1,000.

Bermúdez argues that the thesis that a PD is a NP holds only if the PD is restricted in an appropriate way to guarantee that δ and γ stand in the right kind of relation. The restriction is to be carried out in two steps:

i. the players must be sufficiently similar to serve as reliable replicas of one another.
ii. the players must know that they are sufficiently similar to serve as reliable replicas of one another.

To see why this restriction is needed, one needs to return to Lewis' argument. Clauses (3) and (3'') can now be restated formally as the following biconditionals:

(3) $\alpha \leftrightarrow \delta$
(3**) $\alpha \leftrightarrow \gamma$

Lewis' goal is to show that (3) is a special case of (3**), which would mean that (3**) holds in the PD and the PD is just a special case of NP. To do that, Bermúdez argues, Lewis needs to prove that (3) and (3**) entail each other and this is the case if the biconditional (L), which supplies the link between (3) and (3**), is true.

(L) $\delta \leftrightarrow \gamma$

This biconditional holds under those circumstances in which "you (the other prisoner) are sufficiently like me to serve as a reliable replica" (Bermúdez, 2013: 425).[9] This is what justifies the first step of the restriction Lewis would be forced to implement. However, the PD must be restricted even further because Lewis' analysis "leaves out … the epistemic situation of the player(s)" (Bermúdez, 2013: 427). What needs to be taken into account is the degree of confidence each player has that the two-way dependence relations expressed by (3) and (3**) hold. In NP, not only does (3**) hold but (E3**), its epistemic counterpart, must hold as well (Bermúdez introduces 'C_p – ' as an operator which stands for 'Player p has a high degree of confidence that –').

(E3**) $C_p(\alpha \leftrightarrow \gamma)$

This is simply the generalization of the requirement that in the standard NP case the player must know that she gets $1,000,000 if and only if it is predicted that she does not take the $1,000. Since "if Lewis is correct that the PD is an NP, then comparable knowledge is required in the PD" (ibid. 427), (E3**) must hold in the PD. We can assume that (E3)

[9] Strictly speaking, this would ensure the left to right direction of the biconditional. However, the right to left direction is secured as well because "the very same factors that make you a reliable replica of me make me a reliable replica of you" (Bermúdez 2013: 426).

(E3) $C_p(\alpha \leftrightarrow \delta)$

holds in the PD, because the players must know the rules of the game. What is needed for (E3**) to hold, according to Bermúdez, is (E_L), the epistemic counterpart of biconditional (L):

(E_L) $C_p(\delta \leftrightarrow \gamma)$

This is the motivation for the second restriction, ii. Thus, Bermúdez concludes, in order to show that PD is a special case of NP, Lewis would have to restrict the PD not just to cases in which the players are sufficiently similar to serve as reliable replicas of one another, but to cases in which the PD players are highly confident that they are "sufficiently similar for each to serve as a simulation of the other' (ibid. 428).

Consider now the second thesis Bermúdez argues for, namely, that if the epistemic restrictions specified in B_1 are implemented, one can no longer talk about a genuine PD case. An essential aspect of a PD case, Bermúdez argues, is that it presents the players with four alternative scenarios all of which can be taken seriously. When the players have a very high degree of confidence that they will make similar choices though, they cannot take seriously two out of the four alternative scenarios in the decision matrix. The problem is that "if I think that two of the four available scenarios in the pay-off matrix of the PD are to all intents and purposes impossible, then I cannot believe that I am in a PD" (ibid., 428). Thus, the attempt to add restrictions to a PD in order to make it look like a NP turns it into a different decision problem.[10]

I will argue that the objection raised by Bermúdez fails because it is based on a misinterpretation of Lewis' argument. The first thesis, B_1, is false because none of the two restrictions is actually needed in order for Lewis' argument to succeed. Let us examine them one at a time. Recall that, for Bermúdez, biconditional (L) is needed to prove that (3) is a special case of (3**). This however is not how Lewis justifies the claim that (3) is a special case of (3''). Clause (3''), restated here,

> (3'') I will get my million if and only if a certain potentially predictive process (which may go on before, during, or after my choice) yields an outcome which could warrant a prediction that I do not take my $1,000.

should not be formalized as a biconditional but rather as an existentially quantified sentence. While it can be granted that the quantifier 'a certain' is ambiguous and can be interpreted as having either a wide scope or a narrow scope, it is clear

[10] Although the assumption that the players' high degree of confidence that they will make the same choice is incompatible with the PD is questionable (Sobel, for instance, would reject it), I will not pursue this type of response here.

that in the context Lewis uses it it requires a wide scope reading.[11] Consequently, (3'') should be translated as a formula of the following type,

(3″) [∃x: Px] (α ↔ Qx),

where 'Px' stands for 'x is a potentially predictive process' and'Qx' for 'x yields an outcome which could warrant a prediction that I do not take my $1,000'. All that is needed to show that (3) is a special case of (3'') is to see that (3″) can be derived from it by existential generalization, and this is clearly true: I can think of your choice under perfectly symmetric circumstances as predictive of my own choice and there is no need for a high degree of similarity. Even if you are not a very reliable replica, your action would still warrant a (not very reliable) prediction that I do not take my $1,000. Lewis explicitly points out that in the scenario he describes in his paper, in order for there to be a conflict between EUP and DP, the reliability of the predictor does not have to exceed 0.5005 on a scale from 0 to 1, which is quite low.

Let us now consider the second restriction. Bermúdez is right in saying that for the one problem thesis to be true the players must have comparable knowledge, but this means nothing more than what is required for the players to know the pay-off tables and the rules of the game. Recall that for Bermúdez although the knowledge of the PD pay-off table guarantees that (E3), the epistemic counterpart of (3), holds, it does not guarantee that (E3**) holds. For this to be the case, (E_L) must hold as well, i.e., "the player must have a high degree of confidence that the other player will not take the $1,000 if and only if it is predictable that he himself will not take the $1,000" (ibid. 428). This is what leads him to the conclusion that PD must be restricted to cases in which the players are highly confident that they are sufficiently similar to serve as replicas of one another. I will argue that all it takes to see (E3) as a special case of (E3″)[12] is to pay closer attention to the predictive process and to the knowledge that is implicit in the PD scenario. To use Lewis' language, the potentially predictive process we are dealing with, simulation, can yield two possible outcomes: you, the other player, take the $1,000 or you do not. Since the players are aware that they are both rational beings situated in symmetrical circumstances, the later possible outcome is the one that warrants a prediction that I do not take my $1,000. Thus, (E3) expresses the knowledge that I will receive $1,000,000 if and only if the outcome yielded by simulation is the one that could warrant a prediction that I do not take my thousand and for this reason (E3) is indeed a special case of (E3″).[13] It should be noticed that this argument makes no assumptions about the

[11] See Hornstein (1988) for a defense of the wide scope interpretation of 'a certain'. In our case, the narrow scope reading would be implausible because the mere existence of a potentially predictive process whose outcome predicts that I do not take my $1,000 says nothing about the content of the opaque box.

[12] If the analysis of (3″) as an existential generalization provided above is correct, the epistemic counterpart of (3″) that is relevant to our purposes is a formula of the type '[∃x: Px] C_p (α ↔ Qx)'. In other words, the existential quantifier takes wide scope over the epistemic operator.

[13] I am indebted to an anonymous reviewer for insightful comments that prompted me to revise the argument initially developed in this paragraph. One point made by the referee is that since decision problems must be at least partially individuated by the knowledge the players have, to be in a NP I must have the knowledge specific to a NP – in our case, knowledge that your choice is predictive of my choice. While

reliability of the predictive process (i.e., about the degree of similarity between the two players) as my receiving $1,000,000 in a NP is correlated with the making of the prediction rather than with its fulfillment. This suffices to show that none of the two restrictions is necessary in order for Lewis' argument to go through.

Elsewhere, Bermúdez (2015a, b, 2018) makes an attempt to bolster his position on the distinction between NP and the PD with an argument that draws on the distinction between parametric and strategic choices, a distinction which is often used to demarcate decision theory from game theory. Parametric choice problems are studied in decision theory and admit only one dimension of variation, the agent's choice, while the other parameters are set by the environment. Strategic choice problems, on the other hand, are studied in game theory and involve at least two players; whether the choice of a player is rational depends, among other things, "on the choices made by the other players – with the rationality of those choices also being partly determined by the agent's choice" [2015a: p. 130]. Since the PD is a standard example of a strategic problem while NP is not, Bermúdez argues, "[a]ny attempt to assimilate them is doomed from the start" (215b: p. 783). He offers three reasons, quoted exactly below,[14] for thinking that NP is not a problem of strategic choice:

(1) The Predictor to all intents and purposes knows what the player will choose and so is not acting "without any knowledge as to the choices of the other players".[15]
(2) The Predictor does not strictly speaking choose at all. What the Predictor does is directly determined by what the player does.
(3) The Predictor's action is not determined by a preference ordering defined over the possible outcomes and so they do not really qualify as a player.

Bermúdez admits that by restricting the PD in the way required by Lewis one can make it look similar to NP, but he considers this move illegitimate because it turns what is essentially a strategic problem into a parametric problem. If the arguments I offered above are correct, Lewis is not trying to restrict the dimensions of variation in the PD, but the question remains whether there is a parametric/strategic gap between NP and the PD which Lewis fails to appreciate. I will argue that the three reasons Bermúdez uses to deny NP its status as a strategic choice are not compelling enough and that there is no sharp parametric/strategic contrast between NP and the PD. Starting with the first reason, even if the predictor is highly reliable (a requirement which, according to Lewis, can be easily relaxed) that does not necessarily mean that she knows what the player will choose. For that to be the case, not only would one have to assume some sort of reliabilism about justification but one

Footnote 13 (continued)
this is a fair point, I think it is fully addressed by the observation that the players are aware that they are rational beings situated in symmetrical circumstances. To be in a NP, I do not need to explicitly think of the other player's choice as simulation, nor do I need to agree or even be familiar with Lewis' views on this topic. It is enough to have knowledge of the dependencies specific to NP.

[14] See Bermúdez (2015b: p. 793).

[15] The quote used by Bermúdez here is from a classic textbook by Luce and Raiffa (1957) and expresses a requirement on strategic interactions.

would also have to assume that the prediction is true, which cannot be given as part of the game. As far as the second reason is concerned, the prediction can be characterized as a choice because the predictor has the ability to do otherwise and what she does is not causally determined by what the player does (as it is described by Lewis, NP allows the predictive process to take place even after the player makes a choice). The third reason rests presumably on the idea that if, following Dixit et al., strategic games are "interactions between mutually aware players" while decisions are "action situations where each person can choose without concern for reaction or response from others" (Dixit et al., 2009: 18), the predictor's lack of concern for the player's choice would disqualify NP from being a strategic problem. It is not hard to see what the problem is with this line of reasoning. It is true that, when the choices are considered in isolation, the predictor does not have a preference as to what the player actually chooses but this does not mean that she does not have a preference ordering over the possible outcomes. Out of four possible outcomes, two consist of true predictions (one-boxing / one-box prediction; two-boxing / two-box prediction) while the other two are false predictions. Insofar as she is playing the prediction game, the predictor must have a preference for truth (i.e., for one of the two former outcomes). Thus, NP contains enough game theoretic elements to justify the claim that two such problems, side by side, can be seen as a strategic game. As a matter of fact, NP has often been handled as a strategic game which is best examined with the tools provided by game theory (Weber, 2016; Wolpert & Benford, 2013).

5 Concluding Remarks

The discussion above assumes that the PD and NP are individuated at the level of the abstract structures combining the three clauses identified by Lewis. If one is allowed to individuate the problems in alternative ways, one can unsurprisingly draw a conclusion at odds with that of Lewis. For instance, if one stipulates the existence of a conflict between EUP and DP as a defining feature of NP, one can find reasons to reject or at least qualify the one-problem thesis. While in the case described by Lewis the reliability of the predictor does not have to exceed 0.5005 in order for there to be a conflict between EUP and DP, in other cases the degree of reliability must be significantly higher[16] and, as a result, some PD's would be conflict-free Newcomb-style scenarios. Lewis does not take the existence of a conflict-free scenario to be a serious threat to his view because "even this non-problem might legitimately be called a version of Newcomb's Problem, since it satisfies conditions (1), (2), and (3")" [1979: 239]). Since Lewis' thesis is not intended to apply

[16] For instance, if the pay-off matrix of a NP is modified as in the example below, the reliability of the predictor would have to exceed .9545.

	One-box prediction.	Two-box prediction.
P takes one box.	$1,100.	$0.
P takes two boxes.	$2,100.	$1,000.

to such stipulative accounts of the decision problems, the considerations offered above suffice to establish that the one-problem view is not vulnerable to the objections raised by Sobel and Bermúdez and that it remains the most plausible option. It is a significant result as well. Newcomb-style scenarios have been widely used to support causal decision theory against evidential decision theory but it has been argued that such scenarios are artificial and rare or even impossible.[17] However, if Lewis is right, they are not only possible but ubiquitous. While not all of them generate a conflict between EUP and DP, the parameters can be easily adjusted to turn them into problems that lend support to causal decision theory. Moreover, although Lewis was exclusively focused on the one-shot versions of the problems, the one-problem thesis invites further exploration of common structural patterns and solutions that can be extended to the iterated versions of the PD and NP.[18]

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

References

Bermúdez, J. L. (2013). Prisoner's dilemma and Newcomb's problem: Why Lewis's argument fails. *Analysis, 73*, 423–429.

Bermúdez, J. L. (2015a). Prisoner's dilemma cannot be a Newcomb problem. In M. B. Peterson (Ed.), *The Prisoner's Dilemma* (pp. 115–32). Cambridge University Press.

Bermúdez, J. L. (2015b). Strategic vs. parametric choice in Newcomb's problem and the prisoner's dilemma: Reply to Walker. *Philosophia*, *43*, 787–94.

Bermúdez, J.L. (2018). Does Newcomb's problem really exist? In A. Ahmed (Ed.), *Newcomb's Problem (Classic Philosophical Arguments)* (pp. 19–41). Cambridge University Press.

Dixit, A., Skeath, S., & Reiley, D. (2009). *Games of Strategy* (3rd ed.). W.W. Norton & Company.

Hornstein, N. (1988). A certain as a wide-scope quantifier: A reply to Hintikka. *Linguistic Inquiry, 19*(1), 101–109.

Lewis, D. (1979). Prisoners' dilemma is a Newcomb problem. *Philosophy and Public Affairs, 8*, 235–240.

Luce, R. D., & Raiffa, H. (1957). *Games and Decisions: Introduction and critical survey*. John Wiley.

Mackie, J. L. (1977). Newcomb's paradox and the direction of causation. *Canadian Journal of Philosophy, 7*(2), 213–225.

Nozick, R. (1969). Newcomb's problem and two principles of choice. In N. Rescher (Ed.), *Essays in Honor of Carl G. Hempel* (pp. 114–46). Reidel.

Schmidt-Petri, C. (2005). Newcomb's problem and repeated prisoner's dilemma. *Philosophy of Science, 72*(5), 1160–1173.

Sobel, J. H. (1985). Not every prisoner's dilemma is a Newcomb problem. In R. Campbell and L Sowden (Eds.), *Paradoxes of Rationality and Cooperation* (pp. 263–74). Vancouver University Press.

Sorensen, R. (1985). The iterated versions of Newcomb's problem and the prisoner's dilemma. *Synthese, 63*, 157–166.

[17] See Mackie (1977) and Bermudez (2018).

[18] While the one-off PD can be naturally converted into a repeated game, an iterated NP seems to be more contrived and different in important respects from its iterated PD counterpart. Sorensen (1985) however argues that the iterated problems are generated by the same mechanism and "can be solved in much the same way" (1985: 165). See also Schmidt-Petri (2005) for a comparison between the one-shot and the iterated problems.

Weber, T. (2016). A robust resolution of Newcomb's paradox. *Theory and Decision, 81*, 339–356.
Wolpert, D., & Benford, G. (2013). The lesson of Newcomb's paradox. *Synthese, 190*, 1637–1646.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.