PRAGMATIC INTUITIONS
AND RATIONAL CHOICE

Russell Hardin*

I. INTRODUCTION

The literature of rational choice is pervaded with apparent paradoxes. Often the paradoxes are contrived in the following sense. We consider a problem as though it were a realistic or typical choice problem similar to the kinds of problems we normally face in everyday life. Yet the conditions of the problem as specified defy normal experience. Reasoning through the problem under the specified conditions yields one result while boldly attacking the problem from the intuitions developed from our normal experience yields a contrary result. Hence we think there is a paradox.

One way to characterize our conflicting insights in such cases is to note that we work with at least two different kinds of intuition. We have intuitions about logical consistency and we have intuitions from experience about the ways in which things cohere in the world. We might call these apriori and a posteriori or logical and pragmatic intuitions. For example, many of our pragmatic intuitions about our world are equivalent to deductions from the laws of Newtonian physics. Most people do not know these laws well enough to deduce much of anything from them but still almost everyone has firm, reliable intuitions — pragmatic intuitions — about the cruder effects of gravity. Our logical intuitions often come into conflict with our pragmatic intuitions in certain contexts. When this happens we may simply live with the inconsistency or we may become exercised by it. Generally we expect scientists and academics to become exercised by such inconsistencies in their own work.

* Departments of Political Science and Philosophy, The University of Chicago, Chicago, Illinois 60637, USA
In many of the so-called paradoxes of rational choice the underlying problem is one of conflicting logical and pragmatic intuitions. This is the way to characterize the apparent paradox of the single-play Prisoner's Dilemma: that players should choose not to cooperate even though it seems that they should cooperate. In others, the issue is more nearly that certain intuitions are suppressed. The most extreme example of suppressed intuitions in recent literature is in many analyses of Newcomb's Problem, in which decision theorists suppress fundamentally important pragmatic intuitions in reaching apparently firm logical conclusions about how to choose. Because of a similarity in their payoff structures, these two supposed paradoxes are oddly brought together in many writings. I wish to argue that, unlike the Prisoner's Dilemma, Newcomb's Problem is not a serious choice problem, that when we get our pragmatic intuitions straight on it, it collapses. When we get our pragmatic intuitions straight on the Prisoner's Dilemma, it appears fully compelling as a serious choice problem we can and do face.

II. THE PRISONER'S DILEMMA

Let us first discuss the Prisoner's Dilemma, which, especially because of the work of Anatol Rapoport, is so well known that its treatment here can be brief. It is represented in game 1, in which the first payoff in each cell goes to Row and the second to Column. Each player has two choices: cooperate or defect. They both do better in the outcome (1,1) when both cooperate than in the outcome (-1,-1) when both defect. Suppose I am Row and you are Column. If I am sure your choice is contingent on mine and that you will cooperate if I cooperate and defect if I defect, then clearly I will do better if I cooperate. But if I am sure that your choice is unrelated to mine, I can expect to get a higher payoff by defecting no matter what you do: if you cooperate, I get 2 (instead of the 1 I would get from also cooperating); if you defect, I get -1 (instead of the -2 I would get from cooperating). Hence, if our choices are uncoupled, I should defect if I wish to maximize my own payoff, and you should do likewise, so that we end up with the poor outcome (-1,-1) instead of the clearly better outcome (1,1).

The last conclusion seems to many people to be paradoxical. Why? Many students in my classes over many years have insisted they would not choose that way in a real-life Prisoner's Dilemma. Game theorists say that if they did not, they would be irrational in the sense of not seeking their self interest. On the contrary they may be quite rational in real life but they may seldom find themselves in simple single-play 2-person Prisoner's Dilemmas. Whenever they actually do find themselves in 2-person Prisoner's Dilemmas, these are typically in the context of ongoing relationships. In such contexts it is commonly rational to cooperate because in such contexts both players' future actions (especially their future choices in similar Prisoner's Dilemmas with each other) may be made contingent on each others' choices in the present Prisoner's Dilemma. Hence, the conditions of the ideal single-play Prisoner's Dilemma are commonly not met in recognizable contexts. Sensible people therefore find it paradoxical to suppose it rational to defect in the Prisoner's Dilemma: (See further, Hardin, 1982: ch 9.) Intuitions grounded in experienced Prisoner's Dilemma interactions are not easily overcome in trying to understand a Prisoner's Dilemma that is isolated from other social interactions.

<table>
<thead>
<tr>
<th>Game 1: Prisoner's Dilemma</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Column</td>
</tr>
<tr>
<td>Cooperate Defect</td>
</tr>
<tr>
<td>Cooperate</td>
</tr>
<tr>
<td>1, 1</td>
</tr>
<tr>
<td>-2, 2</td>
</tr>
<tr>
<td>Row</td>
</tr>
<tr>
<td>Defect</td>
</tr>
<tr>
<td>2, -2</td>
</tr>
<tr>
<td>-1, -1</td>
</tr>
</tbody>
</table>

Here the problem is one of reliance on intuitions that are well grounded and correct for their proper context but that do not suit the circumstances in which they are now applied. To get clear on the problem, one need merely think through contexts in which relevant intuitions are provoked. For example, one can think of one-time interactions with certain strangers. Still there may be difficulties because we often moralize our behavior. For example, we think it rather moral than self interested that we should keep promises to our associates. Hence, if we suppose one should be cooperative in a one-time interaction with a stranger, that may be for moral reasons. Oddly, if we suppose one should not be cooperative toward strangers, that may also be for moral reasons. For example, Americans commonly moralize their conflict with the Soviet Union. They suppose that the Soviet Union is morally wrong and not merely in a conflict of interests with the United States.

Sorting out these categories may not come readily to most people, but it generally comes clear to anyone who sticks with it. Moreover, there is general agreement on the meaningfulness of the categories, although there may often be disagreement on the facts that characterize a particular interaction. Getting both the categories and the facts straight is typically necessary if we are to rely on intuitions about how to choose in various contexts. Those who find the single-play Prisoner's Dilemma paradoxical seem generally to be judging it from intuitions gained from other kinds of interactions.
III. NEWCOMB’S PROBLEM

The problem for many people in analyzing the single-play Prisoner’s Dilemma seems to be that they fail adequately to imagine themselves into the situation posed by the choice problem when there is no further relationship between the choosees. In Newcomb’s Problem a similar confusion arises. Many sophisticated philosophers evidently think they can imagine themselves into the relevant situation but fail fully to imagine it. They therefore give credit to flawed intuitions.

In Newcomb’s Problem one faces a marvelous genie which knows in advance how one is going to choose between two attractive options. One may choose box 1 or box 2. Box 1 conspicuously contains $1000. If the genie expects one to choose both boxes, it puts nothing in box 2. If it expects one to choose only box 2, it puts $1,000,000 in that box. Since it would be foolish to choose only box 1 or to pass up both boxes, we may suppose that there are two choices: choose only box 2 or choose both boxes. The genie similarly has two choices: put nothing in box 2 or put $1,000,000 in it. We may represent the problem as game 2, in which there are two choosers but in which only the Row player receives a payoff.

<table>
<thead>
<tr>
<th>Game 2: Newcomb’s Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genie</strong></td>
</tr>
<tr>
<td>$1,000,000 in Box 1 Nothing in Box 1</td>
</tr>
<tr>
<td>Box 2 only $1,000,000 0</td>
</tr>
<tr>
<td>Row Both Boxes $1,001,000 $1,000</td>
</tr>
</tbody>
</table>

Tereence Horgan states our apparent choice problem clearly: “Psychologically, Newcomb’s problem is maddeningly paradoxical. Two deep-seated intuitions come head to head, and both refuse to budge: [1] the intuition that it is crazy to choose both boxes in the belief that you will get only $1,000, rather than choosing only box 2 in the belief that you will get $1 million; and [2] the intuition that it is crazy to choose only box 2, since choosing both can’t affect the contents of box 2 ... I don’t think either intuition can be made to go away; thus, insofar as its psychological force is concerned, Newcomb’s problem remains a brain-teaser.” (Horgan, 1981: 341)

Horgan has his own intuitions in the matter: he thinks Row should choose only box 2. He is not alone. Robert Nozick has put the problem to many people. His result? "To almost everyone it is perfectly clear and obvious what should be done. The difficulty is that these people seem to divide almost evenly on the problem, with large numbers thinking that the opposing half is just being silly" (Nozick, 1970: 117, emphasis added). Note that Horgan speaks of conflicting intuitions. But his intuitions here are decision-theoretic. The oversight in his account of the problem is his neglect of concern with his pragmatic intuitions, which are left implicit and taken for granted. It seems inescapable that his implicit pragmatic intuitions and those of many other philosophers who have written on Newcomb’s problem are muddled and without adequate foundation in their greater system of knowledge and, one trusts, in their wider beliefs about the nature of knowledge.

How can people who are relatively skeptical about knowledge claims in general come to believe they find themselves faced with a genuine Newcomb’s Problem? What conceivable evidence would persuade, say, Horgan or Nozick that he faces Newcomb’s genie without simultaneously bringing him to doubt that he is still free to choose? Any intuition about how one should act in a situation that is intuitively implausible cannot be very compelling without remaking part of one’s intuition. We are all familiar with the claim that, in logic, anything can be deduced from a set of logically inconsistent beliefs. Gilbert Harman notes that “this point is sometimes expressed by saying that from a contradiction we may deductively infer anything.” But this is a peculiar use of ‘infer’. “Logic does not tell us that if we discover that our beliefs are inconsistent we may go on to infer or accept anything and everything we may happen to think of. Given the discovery of such inconsistency in our antecedent beliefs, inference should lead not to the acceptance of something more but to the rejection of something previously accepted” (Harman, 1973: 157-158). Newcomb’s problem is not logically absurd but it is pragmatically absurd. And if I become convinced that it is not pragmatically absurd, I must as well be convinced that vast parts of my wider present understanding of my world are absurd.

I want to argue that philosophically Newcomb’s Problem is not a choice problem but a knowledge problem. As typically posed, it is often incompletely specified. Among the specifications that would be crucial to an intelligent choice are how the genie knows one’s likely choice, and more importantly, how one knows the genie knows. In most accounts, it is assumed that one faces the benevolent genie once and once only, although the assumption is sometimes at best implicit. This too is crucial, as the analysis of the iterated Prisoner’s Dilemma suggests, because iteration may recommend a different strategy. I will assume that the opportunity knocks only once. From this assumption we may infer partial answers to the knowledge questions above.

How might the genie know how I will choose? It would have to know how I have chosen in relevantly like circumstances or it would have to command
relevant psychophysical or behavioral laws. Suppose I have faced such a situation before (forgive the absurdity) and the genie knows my record. I may now behave differently (because I am beginning to believe, say) but the genie has acted on my past record. Its choice will not be contingent on my choice this time but only on my past record. If I am never to face the genie again, or any other offering a similar temptation, I should now simply maximize my expected payoff and take both boxes. (If I am to face the genie again, I should perhaps choose otherwise in order to give it new information about me which would induce it to put the $1,000,000 in next time around.) If the genie is merely predicting from my past behavior, my problem is therefore not very interesting. Suppose then that it commands relevant laws and the data to apply them.

If there are very powerful psychophysical behavioral laws governing my choices, clearly I am unaware of their working. Hence, I cannot know which possible choice follows the relevant laws and which violates them and I therefore cannot know which choice outsmarts the genie. The genie knows an important fact about me that I do not know. This raises the second question above in the specification of the problem: How do I know the genie knows such laws and the relevant data for applying them? I do not know what it would take for me to believe there were such laws without telling me what they were and letting me see their power. Suppose I were told the laws and were able to deduce what choice they predicted for me, and suppose the laws were powerful but not deterministically perfect. I now know what choice the genie has made (by hypothesis, it has used the laws to decide). I am now free to maximize my return given the genie’s choice — which is to say that I am free to choose both boxes in the full knowledge that I thereby will get $1000 more than if I choose only box 2. This follows because the genie’s action was contingent only on the predictions from the relevant laws and not on my actual choice this time. Suppose contrariwise that the laws are all astrophysics to me but that I have been convinced by relevant experts (faced with the genie, even Donald Davidson now accepts that there is at least one psychophysical or behavioral law) that the genie knows such a law and that it decides according to what that law predicts I will do. And again suppose the law is very powerful but not perfectly deterministic. It still follows that the genie’s decision is contingent on the prediction of that law, not on what I actually decide. Hence, again I am free to maximize when I actually do choose.

What if the genie’s law is, and I believe it to be, perfect? In that case I cannot consistently also believe that I am free to choose. I may have the illusion of making a choice but I am as deterministically programmed as the nearest computer. Horgan may give me advice on how to choose, but he is whistling in the wind and, sad to say, he may be unable to stop because he may be programmed to whistle. But surely one cannot go so far as to believe that the genie has command of a perfect law in this realm in the first place. As Davidson (1980: 217, emphasis added) argues, “if we were to find an open sentence couched in behavioral terms and exactly coextensive with some neutral predicate, nothing could reasonably persuade us that we had found it. We know too much about thought and behaviour to trust exact and universal statements linking them. Beliefs and desires issue in behaviour only as modified and mediated by further beliefs and desires, attitudes and attendings, without limit. Clearly this illusion of the mental realm is a clue both to the autonomy and to the anomalous character of the mental.” From our experience with the boxes of magicians, we should far sooner believe our genie plays tricks with its boxes after we choose than that it has command over a law to predict our choices. (Quite apart from the problem of finding such a law we may wonder how it could be applied to deduce behavior. How would our genie tap one’s mind enough to apply the law to predicting one’s choice? Wouldn’t the effort to make relevant measurements partly destroy what was being measured?)

David Lewis (1970: 240) notes that “Some have tended off the lessons of Newcomb’s Problem by saying: ‘Let us not have, or let us not rely on, any intuitions about what is rational in goofball cases so unlike the decision problems of real life’.” He has it all wrong. There are no lessons for rational choice from Newcomb’s Problem because its suppositions are not credible. To accept those suppositions would be to undo much of one’s understanding of the world. As Wittgenstein (1967, ¶ 303, my translation) notes, “One can easily imagine, and describe in full detail, events which, if we saw them occur, would cause us to lose confidence in all judgments.” This is surely an apt characterization of Newcomb’s imaginary problem.

The interest in Newcomb’s Problem should be in what it could mean actually to have such a genie. This is an issue in the foundations of knowledge, not in rational choice. Newcomb’s Problem has been persuasively redefined as a choice problem by facing us with the supposedly mundane task of choosing between two boxes. Intuitions that start from assuming its character as a choice problem are of no value in our understanding of rational choice.

Numerous philosophers have failed to lose confidence, with Wittgenstein, in all their judgments — especially of how they would choose in the face of Newcomb’s genie — only because they have not yet been put before the genie. If they were told they were before it, they would not soon reason as they have done in print. Rather, they would first wonder where the genie gets its money, whether its money is counterfeit, what is the trick in the shell game with the boxes, who else is involved in the deceit, and so forth.

Yet, on the evidence of the extraordinary voluminous literature on Newcomb’s Problem (see several contributions and the bibliography of work in Campbell and Sowden, 1985), one may conclude that many very sophisticated philosophers think they have intuitions about how they would choose even in situations
wildly different from any they have ever faced, indeed, for many of these philosophers in situations in which they could not believe they could ever find themselves even in principle. But there is something contradictory in believing a situation to be impossible in principle because it confounds all one's understandings of the world and nevertheless claiming to know how one would act in the situation.

The persuasive redefinition of Newcomb's Problem as a choice problem has tapped one set of intuitions while suppressing another in the minds of surprisingly many sophisticated philosophers. The intuitions that have been tapped are irrelevant to the problem while those that have been suppressed are destructive of it, at least as a supposed choice problem. Yet the tapped intuitions are given such weight that, as in Nozick's trials, to almost everyone it is perfectly clear and obvious what should be done. Furthermore, the intuitions that have been given great weight are far less fundamental than those that have been suppressed, in the following sense. The suppressed intuitions govern virtually everything one actually does while those intuitions that have been tapped are about hypothetical situations that are utterly unfamiliar. Those unaccustomed to decision theory and to philosophical argument might be incapable of having the latter intuitions. Yet one might suppose that, for those accustomed to decision theory and philosophical reasoning, the tapped intuitions are easier to deal with and think through. They involve simple stepwise deduction while the suppressed intuitions involve holistic interpretations of manifold considerations taken all at once.

IV. CONCLUSION

In effect I have argued that the Prisoner's Dilemma is a serious problem in rational choice whereas Newcomb's Problem is not a problem in choice but is rather, if anything, a test of our theory of knowledge. Lewis (1979) asserts on the contrary that the Prisoner's Dilemma is a Newcomb's Problem. It is instructive to see why this claim is false. Clearly, the structures of payoffs to Row in Newcomb's Problem (game 2) and in the Prisoner's Dilemma (game 1) are the same in the sense that they have the same rank order. An obvious difference between the two games is that Column in game 1 has payoffs while the genie in game 2 has no payoffs. But the difference that is of significance is that in Newcomb's Problem Row is supposed to believe that Newcomb's genie makes choices that are contingent on Row's own. In particular, the genie "cooperates" (it puts $1,000,000 in box 2) if Row cooperates (by choosing only box 2) and "defects" if Row defects.

If, in a Prisoner's Dilemma, you believe your adversary's choices are simi-

larly contingent on your own, obviously you should cooperate. For if your belief is correct then the payoffs in the lower left and upper right cells will not occur. The only plausible payoffs will be in the mutual cooperation and mutual defection cells. It follows that, because your choice of cooperation or defection determines the full outcome, then you should cooperate. This is, of course, exactly the opposite of the usual conclusion for single-play Prisoner's Dilemma. When choices are contingent in relevant ways, as they are if Column behaves as Newcomb's genie does, game 1 poses no dilemma.

Earlier I argued that part of the sense of paradox that many people have when faced with an account of the single-play Prisoner's Dilemma may follow from their tendency to generalize from experience to an odd situation. The usual experience is of ongoing or socially regulated interactions, in which cooperation makes good sense even on a narrowly self-interested account. Yet it is easy to design actual experiments in which the Prisoner's Dilemma is played once only by players who are likely to expect to have no other interactions with each other. Once one grasps the nature of such an interaction and perhaps overcomes moral impulses to think one must — and not merely should — be cooperative, one can easily understand that non-cooperation is the self-interested choice in such an interaction. It is no paradox that individually rational choices may produce sub-optimal results, even though this realization may first come as a novel and perhaps distressing insight. The sense of paradox fades as we associate the problem with the relevant pragmatic intuitions and cut off the irrelevant intuitions. It fades the more readily because we can eventually associate the single-play Prisoner's Dilemma with actual experience of interactions with strangers and with strongly learned intuitions that of course we would not deal with all strangers the way we deal with closer associates.

In this respect the Prisoner's Dilemma is radically different from Newcomb's Problem, because the latter becomes increasingly clouded as it is further considered as a choice problem. It is only by bringing pragmatic, rather than a priori decision-theoretic, intuitions to bear on the problem that we finally see it clearly. Once these intuitions cease to be suppressed, the supposed decision problem disappears.

REFERENCES


