Rational dilemmas

Graham Priest

1. Dilemmas

A dilemma for a person is a situation in which they are required to do incompatible things. That, at least, is one natural meaning of the word. Dilemmas (in this sense) may be only prima facie. It may turn out, on reflection, that one of the requirements overrides the other. On the other hand, some dilemmas may be genuine.

There are many kinds of dilemmas, depending on what, exactly, it is that is doing the requiring. We can be required by morality, required by law, required by rationality, and maybe also by prudence and other things. It is often difficult to show that particular moral dilemmas are more than prima facie. This is because the precise dictates of morality are themselves moot. Much more certain are the dictates of law. And for that reason, we can be sure that there are genuine legal dilemmas. A person contracts, say, to go to a certain place under certain circumstances; they also contract (maybe in the same contract) to go to a different place under certain other circumstances. Against all expectations, both circumstances arise at the same time. The person is then legally obligated to do the impossible.[1]

What of rationality? Can people be rationally required to do the impossible? It might be thought not. How could rationality be so stupid? But, in fact, there may be rational dilemmas. The point of this note is to argue for this conclusion.

2. A self-referential dilemma

A dilemma is not a contradiction, of the form ϕ and $\sim\phi$. Let us use the operator O, 'It is obligatory that', from standard deontic logic. Then the paradigm dilemma is of the form: $O\varphi$ and $O\sim\varphi$, where φ is a statement to the effect that something be done. More generally, in a dilemma there are two such statements ϕ, φ, such that $\sim(\phi\ \&\ \varphi)$ is necessarily true, yet $O\phi$ and $O\varphi$.

As a first example of a rational dilemma, consider a claim of the form, 'It is irrational to believe this claim', that is, something of the form:

It is irrational to believe α

where it, itself, is α. Supposes that you believe α. Then you believe something, and at the same time believe that it is irrational to believe it. This, presumably, is irrational. Hence, you ought not to believe α: $O\sim B\alpha$. But

---

[1] For a further discussion of this, see Priest 1987, ch. 13.

we have just shown that α is true! Hence, you ought to believe α, $OB\alpha$. This is a version of the Irrationalist's Paradox.[2]

One might think that something fishy is going on here. Self-reference is, after all, a funny thing. I do not think so, but rather than pursue this, let us look instead at some other plausible examples of rational dilemmas which do not employ self-reference. These are, in fact, versions of arguments that are well known in the literature on game theory.[3] Standardly, commentators accept one or other horn of the dilemma and try to dispose of the other. I want to set things up in such a way that if you accept one, you should accept both.[4]

3. Game-theoretic dilemmas

I will give two arguments of this kind. Both depend on a certain principle of rationality, which is as follows. Suppose that a person has to choose between two alternatives, and they know that, if they choose one alternative, the benefit derived will be a certain amount; and if they choose the other alternative the benefit derived will be a lesser amount. Then they ought to choose the first alternative. Call this principle R. We may formalize it thus:

$$C(\gamma, \delta)$$
$$M\gamma \rightarrow Gc_\gamma$$
$$M\delta \rightarrow Gc_\delta$$
$$\underline{c_\gamma \geq c_\delta}$$
$$OM\gamma$$

The premises are above the line; the conclusion is below. $C(\gamma, \delta)$ means that you have a choice between making γ true and making δ true; $M\gamma$ means that you make γ true; Gx means that you gain x. $\rightarrow$ denotes the indicative conditional. Strictly speaking, the premises should be within the scope of an epistemic operator, K (it is known that), but in what follows this will go without saying.

Now to the first dilemma: Newcomb's paradox. This comes in different versions; let me spell out the one that I have in mind. There are two boxes, a and b, and you are to choose between taking either the contents of both boxes, or the contents of just one box, box a (the aim being to maximize your financial gain). b is transparent, and you can see a \$10 note inside. You do not know what is in box a, but you do know that money has been

---

[2] Due to Greg Littman. See Priest 1995: 61.

[3] See, for example, Sainsbury 1995, ch. 3 or the papers in Campbell and Sowden 1985.

[4] One possible reaction is to accept neither. For example, one may simply reject the principle of rational choice about to be enunciated. However, this is so central to most of game theory that its derogation is not an enticing one.

put inside by someone who knows exactly what you are going to do, a perfect predictor, p. That is, if you are going to choose one box, p knew this; and if you are going to choose both boxes, p knew this too. If p predicted that you would choose one box, $100 was put inside a; if p predicted that you would choose two boxes, nothing was put inside a. Should you choose one box or two?[5]

We may now show that you ought to choose one box, and that you ought to choose both boxes. Let α be 'you choose just box a' and β be 'you choose both boxes'. Then we have $C(\alpha, \beta)$ and $\sim(M\alpha \,\&\, M\beta)$. The two horns of the dilemma proceed as follows.

Horn 1: Let c be an abbreviation for the description 'whatever is now in box a', where this is to be understood as a rigid designator. If you make α true, then you get c: $M\alpha \rightarrow Gc$. If you make β true, you will get c plus the extra $10: $M\beta \rightarrow G(10 + c)$. But $10 + c > c$. By principle R, it follows that $OM\beta$.

Horn 2: If you make β true, then p knew that you were going to choose both boxes. Hence, there is $10 in a, and nothing in b; so you get $10. That is: $M\beta \rightarrow G10$. On the other hand, if you make α true, then p knew that you were going to choose one box. Hence, there is $100 in a, which is what you get. That is: $M\alpha \rightarrow G100$. But clearly, $100 > 10$. By principle R, it follows that $OM\alpha$.

A second example of a rational dilemma is provided by a suitably symmetrized version of the prisoners' dilemma. The following will do. You are in a room with two buttons, and you have to choose between pressing them. If you press button a, you will receive $10. If you press button b you will receive nothing (by your own efforts), but the person next door will receive $100. That person is in exactly the symmetric situation. Moreover, you have known the person in the other room for a long time, and you know that they are just like you: in choice situations you always both choose the same thing. Which button should you press?

Let α be 'you press button a'; let β be 'you press button b'. Then we have $C(\alpha, \beta)$, and $\sim(M\alpha \,\&\, M\beta)$. The dilemmatic argument proceeds as follows.

Horn 1: Let c be an abbreviation for the description 'whatever is obtained as a result of the action of the other person', where this is to be understood as a rigid designator. If you make a true, then you get $10 + c$: $M\alpha \rightarrow G(10 + c)$. If you make β true, you will get just c: $M\beta \rightarrow Gc$. But $10 + c > c$. By principle R, it follows that $OM\alpha$.

---

[5] It does not make much difference if p is not a perfect predictor, but merely a very good one. For then the conditionals in the argument take the form: if you do so and so, *it is very likely that* such and such. But much the same argument goes through with these.

Horn 2: If you make α true, then so will the person in the other room. Hence, you will get 10 plus nothing else: $M\alpha \to G10$. If you make β true, then so will the person in the other room. Hence, you will get 100: $M\beta \to G100$. Since $100 > 10$, by principle R, $OM\beta$.

4. *Objections*

The most plausible line of objection to the preceding arguments, it seems to me, concerns the nature of the conditionals employed. In particular, one may argue that the first horn of each dilemma is invalid as follows. I give the argument for the Newcomb problem; the argument in the prisoners' dilemma example is similar.

Let us suppose that, as a matter of fact, you do choose the one box, a, and consequently that $c = 100$. Consider the conditional $M\beta \to G(10 + c)$. If you were to make β true then you would not get $10 + c$, that is, 110, but 10, since there would be 0 in box a. (Recall that c is a rigid designator, and so does not change its value in different hypothetical situations.) Hence this conditional is false. Similarly, let us suppose that you do, as a matter of fact, choose both boxes, and consequently that $c = 0$. Consider the conditional $M\alpha \to Gc$. If you *were* to make α true, you would not get c, that is, 0, but 100, since that is what would be in box a. Consequently one or other of these conditionals is false.

This objection fails, since what it establishes is that certain subjunctive conditionals are false (note the 'were's). But the conditionals employed in the arguments are, in every case, indicative, not subjunctive. (Go back and check them!)

At this point, it might be argued that a correct formulation of principle *R requires* the use of subjunctive conditionals. But why should one suppose this? After all: the reasoning, including the formulation of principle *R*, seems quite in order as it is.

Here are two bad reasons why the conditionals in *R* must be subjunctives. The first: we are reasoning about cases at least one of which will not arise. Hence, the conditionals are counterfactuals, and reasoning about counterfactual situations requires subjunctive conditionals. Not so. Counterfactuals are often expressed by indicative conditionals, not subjunctives. The bus is due. I tell you: *if we don't leave now, we'll miss the bus*; and so we leave, and catch the bus, which is on time. The conditional I uttered was a true indicative conditional, though it is also a counterfactual. Even in cases where the antecedent is not just false, but its truth would require the past to be different from what it was, we can still use indicative conditionals. Let us suppose that you do not know who won the Grand Final yesterday. You do know that *if you read in a reputable paper today that the Broncos won the Final, then they won it yesterday*. This is a perfectly

legitimate indicative conditional, even though the Broncos did not, in fact, win yesterday.

The second, and more sophisticated, argument to the effect that condition R requires subjunctive conditionals goes as follows.[6] Again, I use the Newcomb example to illustrate. Suppose that, as a matter of fact, you make β true; then you do not make α true, $\sim M\alpha$. Then, by the properties of the material conditional, $M\alpha \rightarrow G100$. So if you know that you will make β true then you know that $M\alpha \rightarrow G100$. So you ought to make α true, since you know that $100 > 10$. And quite generally, whenever you have decided what to do, so that you know what you are going to do, you ought to do the opposite! So the principle, formulated with indicative conditionals, is incoherent.

The flaw in the argument is in identifying the indicative conditional with the material conditional (so that we can reason $\sim\gamma \vdash \gamma \rightarrow \delta$). Standard indicative conditionals are *not* material conditionals. There are just too many clear counter-examples to the identity for it to be credible. Of the many that could be given, here is just one.[7]

There is an electrical circuit with two switches in series, both off. If both switches are on (and only if both switches are on) a light in the circuit will go on. Let α be 'switch 1 is turned on'; let β be 'switch 2 is turned on'; let γ be 'the light goes on'. Then we have: if α and β then γ. If this conditional were material, and since $(\alpha \,\&\, \beta) \supset \gamma \vdash ((\alpha \,\&\, \sim\beta) \supset \gamma) \vee ((\beta \,\&\, \sim\alpha) \supset \gamma)$, it would follow that one or other of the following is true: 'if only switch 1 is turned on, the light will go on', 'if only switch 2 is turned on, the light will go on'. Both are clearly false.

5. *Conclusion*

There are, then, rational dilemmas. The fact that there are such things raises the question of what one should do if one finds oneself in a dilemma. What one should do, is, of course, the impossible. But one can't do that. One way or the other, one is going to be rationally damned. *Ex hypothesi*, rationality gives no guidance on the matter – or rather, it gives too much, which comes to the same thing. Hence, what one does do will have to be determined by other things. But who ever thought that there was a rational answer to everything?[8]

---

[6] The argument is a variation of a point sometimes made in discussions of backwards induction. See Priest 2000, n. 31.

[7] For others, see Priest 2001, ch. 1, where the claim that the indicative conditional is not material is mounted in more detail.

[8] Versions of this paper were given to audiences at the Universities of New England, Gent, La Trobe and Melbourne. I am grateful to the members of these for their helpful comments.

The University of Melbourne
Australia, 3010
g.priest@unimelb.edu.au

References

Campbell, R. and L. Sowden. 1985. *Paradoxes of Rationality and Cooperation;
Prisoner's Dilemma and Newcomb's Problem*. Vancouver: University of British
Columbia Press.

Priest, G. 1987. *In Contradiction*. Dordrecht: Kluwer Academic Publishers.

Priest, G. 1995. Gaps and gluts: reply to Parsons. *Canadian Journal of Philosophy* 25:
57–66.

Priest, G. 2000. The logic of backwards inductions. *Economics and Philosophy* 16:
267–85.

Priest, G. 2001. *Introduction to Non-Classical Logic*. Cambridge: Cambridge University Press.

Sainsbury, R. M. 1995. *Paradoxes*. 2nd edition. Cambridge: Cambridge University
Press.