Hoffman, M. 1975. Developmental synthesis of affect and cognition and its implications for altruistic motivation. *Developmental Psychology* 11: 607–622.

Onishi, K. and R. Baillargeon. 2005. Do 15 month old infants understand false beliefs? *Science* 308: 255–58.

Plotkin, H. C. 2004. *Evolutionary Thought in Psychology: A Brief History*. Malden, MA: Wiley-Blackwell.

Prinz, J. 2011. Is empathy necessary for morality? In *Empathy: Philosophical and Psychological Perspectives*, eds. P. Goldie and A. Coplan, 211–29 Oxford: Oxford University Press.

Simner, M. 1971. Newborn's response to the cry of another infant. *Developmental Psychology* 5: 136–50.

Smetana, J. and J. Braeges. 1990. The development of toddlers' moral and conventional judgements. *Merrill-Palmer Quarterly* 36: 329–46.

Thompson, R. 1981. Empathy and emotional understanding: the early development of empathy. In *Empathy and its Development*, eds. N. Eisenberg and J. Strayer, 119–45. Cambridge: Cambridge University Press.

Tomasello, M. 2014. *A Natural History of Human Thinking*. Harvard: Harvard University Press.

Zahn-Waxler, C. and M. Radke-Yarrow. 1990. The origins of empathic concern. *Motivation and Emotion* 14: 107–30.

Dicing with death

ARIF AHMED

Death in Damascus is decision theorists' usual name for unfortunate situations like this one.

> Death works from an appointment book that states time and place; a person dies if and only if the book correctly states in what city he will be at the stated time. The book is made up weeks in advance on the basis of highly reliable predictions of your actions. An appointment for tomorrow has been inscribed for you; you know that it is either for Aleppo or for Damascus. You must decide now whether to stay in Damascus overnight, or ride to Aleppo to arrive tomorrow morning.[1]

If your life is worth 10 units (1 unit = \$1M) then your position is as in Table 1. The column headings in Table 1 are the possible contents of Death's appointment book: either Death has you down for Aleppo tomorrow, or he has you down for Damascus tomorrow. Which one obtains is causally independent of what you do.

1 Gibbard and Harper 1978: 373 with trivial alterations. If you prefer something more realistic: imagine that smoking is harmful to you iff you possess a gene that makes you want to smoke.

Analysis Vol 74 | Number 4 | October 2014 | pp. 587–592 doi:10.1093/analys/anu084

Table 1. Death in Damascus

	Death in Aleppo	Death in Damascus
Ride to Aleppo	0	10
Stay in Damascus	10	0

The row headings are your options. These options represent once-and-for-all decisions: once you have committed to either row there is no going back.

The entries in the cells are the values (to you) of the four possible outcomes e.g. '10' in the top-right cell means that you value at 10 units the outcome that you go to Aleppo whilst Death expects you in Damascus (so in which you survive). Similarly '0' in the top-left cell means that you value at 0 units the outcome that you go to Aleppo and Death expects you there (because in that situation you die).

Everyone agrees that yours is an unfortunate situation. You are playing high-stakes hide-and-seek against someone who can predict where you will hide. Everyone you know who has played this game has lost. So there is every reason to think you will lose, too.[2]

At least there is until you chance upon the street vendor Rhinehart. Knowing your predicament, this man offers a third option. 'Round the back of my stall, I have a fair and truly indeterministic coin. I am about to toss it, and for an arbitrarily small fee Δ I can tell you how it lands. Why don't you let the toss decide? So if it lands heads you go to Aleppo; if it lands tails you stay in Damascus.'

You seem hesitant, so Rhinehart continues: 'I guarantee that Death's predictive powers don't cover this coin. Of course, he *has* correctly predicted

2 In this situation CDT (explained below) advises you to go wherever you now think Death is most likely not to be. This means that CDT creates a kind of instability: as you become more confident that you will (say) ride to Aleppo, you become more confident that Death is waiting for you there, and so more confident that you should have stayed in Damascus. But although it is true that CDT has this consequence (Gibbard and Harper 1978: 372–375), this is hardly a decisive objection to CDT.
Here is one of many possible replies: distinguish two types of regret. A-regret: regretting an act when you are halfway through implementing it. B-regret: regretting the act when you *have* implemented it and all the relevant facts are in. If you are confident *ex ante* that Death has you down for Damascus then CDT rationalizes riding to Aleppo. It is true in that case that you should be confident *ex ante* that you will A-regret riding to Aleppo. But since you are confident *ex ante* that Death has you down for Damascus, you should also be confident *ex ante* that you will *not* B-regret riding to Aleppo (but that you *would* B-regret staying in Damascus). But it is irrelevant that you will A-regret a decision if you know that you will not B-regret it. (I am considering whether to take a long flight. I know in advance that I will A-regret it, because when I am in the air I usually become convinced that something will go wrong. But I know in advance that I won't B-regret it, because I know in advance that nothing will go wrong. So I ought to take the flight.) So instability is harmless. And the argument in this paper is entirely independent of considerations about stability.

Table 2. Death in Damascus with the opportunity to randomize

	S_1: Death in Aleppo & H	S_2: Death in Aleppo & T	S_3: Death in Damascus & H	S_4: Death in Damascus & T
Ride to Aleppo	0	0	10	10
Stay in Damascus	10	10	0	0
Randomize	$-\Delta$	$10-\Delta$	$10-\Delta$	$-\Delta$

whether you accept my offer. But if you do, he, like everyone else, is no better than chance at telling where you will be tomorrow. I have had many clients in your present situation – and about half of them have cheated Death.'

Certainly this looks like a good offer. Would you rather be playing hide-and-seek against (a) an uncannily good predictor of your movements or (b) someone who can only randomly guess at them? Rhinehart is offering the chance to reduce Death from (a) to (b). Of course you should take the offer. And of course Death will have hoped that you won't.

Death was therefore pleased to learn that you, like most philosophers, follow Causal Decision Theory (CDT). According to this theory, your choices in a situation should take no account of their symptomatic bearing on states of the world that they do nothing to affect (for instance, predictions of those acts).

What does this means in the present case? Your possible options, the relevant possible states of the world and the value to you of each possible outcome are now as in Table 2.

The bottom row in Table 2 corresponds to your new option: pay the fee; see the coin; Aleppo if heads, Damascus if tails. And the columns represent four possible states, S_1–S_4, corresponding to the four possible results of two processes, namely Death's prediction and the toss of the coin. These processes are causally independent of one another, and they are both causally independent of what you now do. So in Table 2 (as in Table 1), which column obtains is causally independent of what you do.

So according to CDT, you should value each option by a weighted sum of the values it gets from each S_i, the weights being your current probability that that S_i is true. Writing $\Pr(S_i) = x_i$, this means that you should perform that act that maximizes *utility*, where the utility U of each option is[3]:

3 It would be tedious but straightforward to verify that (1)–(3) follow from almost all of the various forms of CDT that are now on the market. For one example, consider Lewis's theory (1981). According to Lewis, the utility of each option O is given by $\Sigma_j \Pr(K_j)\, V(O \wedge K_j)$, where $V(O \wedge K_j)$ measures the value that you would get if $(O \wedge K_j)$ were true, and the K_j are causal dependency hypotheses stating how everything that matters to you causally depends on what you do (1981: 313). In the present case, we can take the K_j to be S_1–S_4, since each S_i settles what the effect of each option would be, in so far as you care. For instance, S_1 settles that if you go straight to Aleppo then you will die, if you go straight to Damascus you will live, and if you randomize then you will die having first paid Δ to the vendor. It follows that Lewis's theory is committed to each of (1)–(3). Similar arguments

(1) U (Ride to Aleppo) $= 10x_3 + 10x_4$
(2) U (Stay in Damascus) $= 10x_1 + 10x_2$
(3) U (Randomize) $= -\Delta x_1 + (10-\Delta)(x_2 + x_3) - \Delta x_4$

Now the coin is independent of Death's appointment book. And you know that tossing it gives a 50% chance to each outcome. It follows that:

(4) $x_1 = x_2 = (x_1 + x_2)/2$
(5) $x_3 = x_4 = (x_3 + x_4)/2$

Given (4) and (5) we can simplify (1)–(3):

(6) U (Ride to Aleppo) $= 20x_3$
(7) U (Stay in Damascus) $= 20x_1$
(8) U (Randomize) $= (10-2\Delta)(x_1 + x_3)$

Straightforward algebra now shows that for *any* values of x_1, x_3, $\Delta > 0$, we have *either*

(9) U (Ride to Aleppo) > U (Randomize); *or*
(10) U (Stay in Damascus) > U (Randomize)

– and possibly both.[4] For instance, given the most natural probability distribution, $x_1 = x_2 = x_3 = x_4 = 0.25$, we have:

(11) U (Ride to Aleppo) $= 5$
(12) U (Stay in Damascus) $= 5$
(13) U (Randomize) $= 5-\Delta$

So *whatever* you think is in Death's appointment book, CDT rejects randomizing, however small the fee. It always gives the absurd advice

show that we get the same result from the decision theories of Skyrms (1980), Sobel (1989) and Joyce (1999).

The one version of CDT that does not straightforwardly deliver (1)–(3) is that of Gibbard and Harper (1978). But this is not because that theory assigns different U-scores to the options on Table 2, but only because it is not properly equipped to deal with cases involving genuine chance. But when amended as it should be to handle these cases (for the details of which see Lewis 1981: 329–335), it too gives the U-scores (1)–(3).

4 Filling in the details: suppose that U (Randomize) exceeds or equals both of U (Ride to Aleppo) and U (Stay in Damascus). Then by (6)-(8) we have:

(i) $(10-2\Delta)(x_1 + x_3) \geq 20x_1$
(ii) $(10-2\Delta)(x_1 + x_3) \geq 20x_3$

Adding these gives:

(iii) $(20 - 4\Delta)(x_1 + x_3) \geq 20(x_1 + x_3)$

But since $x_1 + x_3 > 0$, (iii) implies $-4\Delta \geq 0$, which is false since $\Delta > 0$. So the assumption must have been false: at least one of the two non-randomizing options must have a strictly greater (causal) utility than does randomizing.

that you turn down Rhinehart's offer.[5] So as Death foresaw, you turn it down and (as it happens) ride straight to Aleppo. I needn't say what happens next.[6]

5 The same point applies to a recent modification of CDT due to F. Arntzenius (2008: 292; see also Joyce's 'full information' requirement on CDT (2012: 127), to which all of the following equally applies). According to this *Deliberational Decision Theory* (DDT), the upshot of rational decision-making is an equilibrium in which you (a) have a probability distribution over which outcome you will realize such that (b) on this distribution, CDT reckons each option given non-zero probability to be at least as good as any other option. If Death is a perfect predictor of your choice in the original *Death in Damascus* case (Table 1), DDT does not unequivocally recommend either option but rather an equilibrium in which Pr (Ride to Aleppo) = Pr (Stay in Damascus) = 0.5.
Now in the modified case, where randomization is an option, we may suppose that Death is a perfect predictor of non-randomized acts but can only guess at the outcome of the toss. Let us abbreviate the three options A, D and R. So {A, D, R} is a partition of the event space. Then by the formula that $\Pr(S) = \Sigma_i \Pr(S|X_i) \Pr(X_i)$ for any partition $\{X_i\}$, we have:

(i) $\Pr(S_1) = \Pr(S_2) = 0.5\Pr(A) + 0.25\Pr(R)$
(ii) $\Pr(S_3) = \Pr(S_4) = 0.5\Pr(D) + 0.25\Pr(R)$

So by (a), the (causal) utilities for the three options in any DDT equilibrium are:

(iii) $U(A) = 10\Pr(D) + 5\Pr(R)$
(iv) $U(D) = 10\Pr(A) + 5\Pr(R)$
(v) $U(R) = (10\text{-}2\Delta)(0.5\Pr(R) + 0.5\ \Pr(D) + 0.5\ \Pr(R)) = 5\ \text{-}\ \Delta$

Suppose that there is an equilibrium in which $\Pr(R) > 0$. Then by (b) we have:

(vi) $U(R) \geq U(D)$
(vii) $U(R) \geq U(A)$

Adding (vi) and (vii) gives $2U(R) \geq U(D) + U(A)$; substituting (iii)–(v) then gives $10 - 2\Delta \geq 10\ (\Pr(A) + \Pr(D) + \Pr(R))$; but $\Pr(A) + \Pr(D) + \Pr(R) = 1$ since {A, D, R} is a partition. So there is an equilibrium in which $\Pr(R) > 0$ only if $10\ \text{-}\ 2\Delta \geq 10$; this is a contradiction since $\Delta > 0$. It follows that in any Arntzenius-type equilibrium you are certain that you will not randomize i.e. DDT rules out randomizing, just like CDT. (Of course, this also follows more directly from the fact that (9)–(11) are true on *any* probability distribution Pr.) In fact it is now easily seen that the unique DDT equilibrium has $\Pr(A) = \Pr(D) = 0.5$, just as in the original *Death in Damascus* case.

6 Let nobody object that this case is too unrealistic to count as a serious objection to CDT. (i) That objection cuts both ways. The case that motivated CDT in the first place was the Newcomb problem (Nozick 1969: 207–8), which is equally unrealistic. So if for this reason we should learn nothing from *Death in Damascus* or this variation upon it, then equally we should refuse to learn anything from the Newcomb problem itself. (ii) There are plausibly realistic versions of the Newcomb problem; but the back-stories that make *them* realistic can easily be modified to give realistic versions of the present case. For instance, one realistic reading of Newcomb's problem is to read it as a version of *Prisoners' Dilemma* played against an imperfect replica of one's self, for instance, against a person who is psychologically very much like you (Lewis 1979). But then, we can equally read the present case as a game of Matching Pennies played against such a replica. For instance: Holmes is chasing Moriarty; at present they are stuck in different carriages of the same train. Moriarty wins if he and Holmes get off at different stops; Holmes wins if they get off at the same stop. Holmes has proven himself very good at anticipating

Nor does the moral of this story need much elaboration. Platitude: in a game that you lose iff your causally isolated opponent correctly predicts your act, you are better off playing against a hopeless predictor than against a very good predictor. This is, I submit, intuitively compelling.[7] But the foregoing construction shows that CDT makes no room for it. That is reason to abandon not only CDT but also any variation on it that entails (1)–(3).

Faculty of Philosophy,
University of Cambridge, UK
ama24@cam.ac.uk

References

Arntzenius, F. 2008. No regrets, or: Edith Piaf Revamps Decision Theory. *Erkenntnis* 68: 277–97.

Gibbard, A. and W. Harper. 1978. Counterfactuals and two kinds of expected utility. In *Foundations and Applications of Decision Theory,* eds. C. Hooker, J. Leach and E. McClennen, 125–62. Dordrecht: Riedel. (Reprinted in *Decision, Probability and Utility,* eds. P. Gärdenfor and N.-E. Sahlin. Cambridge: CUP, 1988).

Hunter, D. and R. Richter. 1978. Counterfactuals and Newcomb's Paradox. *Synthese* 39: 249–61.

Joyce, J. 1999. *Foundations of Causal Decision Theory*. Cambridge: CUP.

Joyce, J. 2012. Regret and instability in causal decision theory. *Synthese* 187: 123–45.

Lewis, D. K. 1979. Prisoners' dilemma is a Newcomb problem. *Philosophy and Public Affairs* 8: 235–40.

Lewis, D. K. 1981. Causal decision theory. *Australasian Journal of Philosophy (AJP)* 59: 5–30. (Reprinted in his *Philosophical Papers Vol. II*. Oxford: OUP, 1986: 305–39).

Nozick, R. 1969. Newcomb's problem and two principles of choice. In *Essays in Honor of Carl G. Hempel,* eds. N. Rescher, 114–46. Dordrecht: D. Reidel. (Reprinted in *Rationality in Action: Contemporary Approaches,* ed. P. Moser, 207–34. Cambridge: CUP, 1990).

Skyrms, B. 1980. *Causal Necessity*. New Haven: Yale UP.

Sobel, J. H. 1989. Partition theorems for causal decision theories. *Philosophy Science* 56: 71–93.

Moriarty's behaviour in cases like this in the past. It is worth Moriarty's while to pay 1¢ to use a randomizing device to settle where he gets off, especially if there are many stops. But CDT rules that out.

7 That platitude is a mere notational variant of this one: in a game that you win iff your causally isolated partner correctly predicts your action, you are better off playing with a very good predictor than against a hopeless one. So understood, it is what underlies the thought that you should pay to play co-ordination games with a replica of yourself (Hunter and Richter 1978: 257–258). Lewis accepts that thought, and so presumably also the platitude behind it, with which he claims that CDT is consistent (1981: 335–337). What the present example shows is that it isn't.