

A Quantum-mechanical argument against Causal Decision Theory

The paper argues that on three out of five possible hypotheses about the Stern-Gerlach experiment we can construct novel and realistic decision problems on which (a) Causal Decision Theory and Evidential Decision Theory conflict (b) Causal Decision Theory and Quantum Mechanics conflict. I conclude that Causal Decision Theory is false.

Before turning to our main case let us briefly review the two theories. Evidential Decision Theory (EDT) recommends what is most auspicious. More formally, it recommends whatever option has the greatest V-score amongst those available. If $V(P) \in \mathbf{R}$ measures your news value for—i.e. how pleased you would be to learn the truth of—an arbitrary proposition P, and if $Cr(P) \in [0, 1]$ measures your confidence in P, then for any (proposition describing an available) option O, Cr and V jointly satisfy:

$$(1) V(O) = \sum_{S \in \mathbf{S}} V(S \wedge O) Cr(S | O)$$

— \mathbf{S} being *any* partition on the underlying space.¹

Causal Decision Theory (CDT) recommends what is most *efficacious* i.e. the option with the greatest U-score, which we define as:

$$(2) U(O) = \sum_{S \in \mathbf{S}^*} V(S \wedge O) Cr(O \rightarrow S)$$

$Cr(O \rightarrow S)$ measures your confidence that if you *were* to realize O then S *would* be true, on a reading of the subjunctive that makes it sensitive to the *effects* but not to the *causes* of what O describes.

Here, \mathbf{S}^* is any partition whose cells (sometimes called ‘states of nature’) capture everything you care about given what you do, i.e. if $S \in \mathbf{S}^*$ then $V(S \wedge O \wedge Y) = V(S \wedge O)$ for any available O, and Y such that $Cr(S \wedge O \wedge Y) > 0$.² A convenient choice for \mathbf{S}^* , if available, will be such that (i) its cells capture everything that matters to me given what I do and (ii) *which* state of nature or cell obtains is causally independent of what I do. For instance, if I am choosing between taking and not taking my umbrella and care only to stay dry, then the partition $\{S1 = I \text{ get wet}, S2 = I \text{ don't get wet}\}$ meets condition (i) but not condition (ii). But the partition $\{S1 = It \text{ rains}, S2 = It \text{ doesn't rain}\}$ meets both conditions. In that case the expression for an option’s U-score takes this particularly simple form:

$$(3) U(O) = \sum_{S \in \mathbf{S}^*} V(S \wedge O) Cr(S)$$

1. Bell’s Theorem

In order to appreciate the point it isn’t necessary to grasp any details of quantum mechanics but only the essentially statistical reasoning that creates the problem: for this purpose the following completely non-technical exposition, which follows Mermin (1981), is perfectly adequate.

¹ Jeffrey 2003: 99.

² See e.g. Gibbard and Harper 1978: 345f. For their treatment of the subjunctive see 344n2. For other definitions of U-score that differ in ways that make no difference here, see e.g. Skyrms 1984: 70; Lewis 1981: 313; Sobel 1989: 73; Joyce 1999: 161.

It is technically feasible to produce a device with the following features. It contains three components: a source S and two receivers A and B. The two receivers are placed on either side of the source and are so separated from one another that there is no possibility of causal commerce between them: at any rate, we are at the outset as sure that *they* are causally isolated as we are ever sure that any two systems are causally isolated. Each receiver contains a display and a switch with three settings labeled 1, 2 and 3. We can independently move each switch to any one of these three settings.

Once the switches have been set the source is turned on. It emits two signals, each receiver picking up one. The display of each receiver then shows one of two readings: let these be 'y' and 'n'. That represents one 'run' of the device. We record the run by noting down the setting of each receiver and the reading on its display. For instance, we might write '12yn' to indicate that A was set to 1, B was set to 2, the display on A was y and the display on B was n. Similarly, '33nn' indicates that both receivers were set to 3 and both displayed 'n'.

We perform repeated runs of the device with the receivers being set at random. Each possible setting of the receiver occurs with the same frequency as any other. Repeated runs reveal the following statistical facts:

- (1) *Whenever* the switches on A and B are on the *same* setting (i.e. both on 1, both on 2 or both on 3) the devices display the same reading i.e. either they both say 'y' or they both say 'n'. So we sometimes get runs like this: '11yy', '22yy'. But we *never* get runs like this: '22yn', '33ny'.
- (2) When the switches on A and B are on any particular different settings (e.g. A on 1, B on 3), the devices display the same reading about 25% of the time. So we get runs like '12yn' and '23ny' about three times as often as we get runs like '12yy' and '13nn'.

That completes what we need to know about the workings of this device. According to the quantum theory it is certainly feasible:

The two particles emerging from the [source] are spin 1/2 particles in the singlet state. The two receivers contain Stern-Gerlach magnets, and the three [positions of the switch on each receiver] determine whether the magnets are vertical or at 120° to the vertical in the plane perpendicular to the line of flight of the particles. When the switches have the same setting the magnets have the same orientation. One receiver [displays y or n] according to whether the measured spin is along or opposite to the field; the other uses the opposite... convention. Thus when the [displays give the same reading] the measured spin components are different.

It is a well-known elementary result that, when the orientations of the magnets differ by an angle θ , then the probability of spin measurements on each particle yielding opposite values is $\cos^2(\theta/2)$.

This probability is unity when $\theta = 0$ [case (1)] and $1/4$ when $\theta = \pm 120^\circ$ [case (2)].³

So far I've stated only the bare facts concerning the mechanics and performance of the device i.e. without any theoretical overlay. What follows is just one possible theoretical interpretation of the device.

Fact (1) records a maximally strong correlation between the readings on the displays of the two receivers when both are set in the same way. Given that there is no causal communication between the two receivers, it seems that the only explanation for this fact is that the two particles are emitted from the source in the same state or 'instruction set'. That is: let us write e.g. 'YYN' to describe e.g. the instruction set of a particle that would generate a reading of 'y' if the switch on the receiver were at setting 1 or 2 and 'n' if the switch were at setting 3 (etc.). If on a given run the source emits particles in the state NYN then that would explain our getting (say) the result '11nn' or '22yy' as well as the fact that we do *not* get (say) the result '11ny' or '33yn'.

Unfortunately the statistical fact (2) appears to be incompatible with this simple and (it seems) inescapable hypothesis. To see why, assume first that there is no correlation between the prior state of the particles and your decision to set the receivers to any particular pair of settings on any particular run. Let us write 'Fr (X)' for the frequency of some condition X and 'Fr (X|Y)' for the *relative* frequency of X given the condition Y. Let us write 'S_i' for the ith state of a given particle (so that $1 \leq i \leq 3$, and $S_i = Y$ or $S_i = N$). And let us write 'j;k' for the proposition that one of the receivers is set to j and the other to k in either order (so that $j, k = 1, 2, 3$). Then we may write down the 'no-correlation' assumption in the following form: for any S_1, S_2, S_3 and any j, k, we have:

$$(3) \text{Fr}(S_1 S_2 S_3 | j;k) = \text{Fr}(S_1 S_2 S_3)$$

Now fact (2) implies that when the receivers *are* set at different values, we get the same reading one-quarter of the time. So in particular we have:

$$(4) \text{Fr}(YYY | 1;2) + \text{Fr}(YYN | 1;2) + \text{Fr}(NNY | 1;2) + \text{Fr}(NNN | 1;2) = 0.25$$

$$(5) \text{Fr}(YYY | 1;3) + \text{Fr}(YNY | 1;3) + \text{Fr}(NYN | 1;3) + \text{Fr}(NNN | 1;3) = 0.25$$

$$(6) \text{Fr}(YYY | 2;3) + \text{Fr}(NYY | 2;3) + \text{Fr}(YNN | 2;3) + \text{Fr}(NNN | 2;3) = 0.25$$

From (3) we can simplify these to:

$$(7) \text{Fr}(YYY) + \text{Fr}(YYN) + \text{Fr}(NNY) + \text{Fr}(NNN) = 0.25$$

$$(8) \text{Fr}(YYY) + \text{Fr}(YNY) + \text{Fr}(NYN) + \text{Fr}(NNN) = 0.25$$

$$(9) \text{Fr}(YYY) + \text{Fr}(NYY) + \text{Fr}(YNN) + \text{Fr}(NNN) = 0.25$$

Adding these together we get:

³ Mermin 1981: 407-8

$$(10) \quad 2\text{Fr}(YYY) + 2\text{Fr}(NNN) + \sum_{S_1, S_2, S_3} \text{Fr}(S_1 S_2 S_3) = 0.75$$

But we know by the probability calculus—which certainly applies to frequencies—that:

$$(11) \quad \sum_{S_1, S_2, S_3} \text{Fr}(S_1 S_2 S_3) = 1$$

$$(12) \quad 2\text{Fr}(YYY) + 2\text{Fr}(NNN) \geq 0$$

But (10), (11) and (12) are inconsistent. Since they follow from (2) and (3), and since (2) has been observationally verified as convincingly as you like, it seems that we are left with two options: either (A) reject (3) and keep the hypothesis of a prior instruction set; or (B) reject the hypothesis of an instruction set. In fact each option itself involves further sub-options as follows.

2 Five responses

(A) We can insist that there is a prior common state of the particles (a prior instruction set) but deny (3) i.e. maintain that that state is correlated with one's possibly randomized choice of receiver setting. There are three ways to extend this line.

One might claim (A1) that one's present choice of setting of the receiver has a *retrocausal* effect on the prior state of the hidden variables, so that, for instance, switching the receivers to 1;2 has the effect of inhibiting (though without altogether excluding) the *prior* instruction sets YYY, YYN, NNY and NNN.⁴

Alternatively, one might claim (A2) that one's present choice is itself *caused*, either by the prior state of the particles itself, or by some still earlier state that was a common cause of both. Bell considered this line to be incompatible with free will⁵; whatever one thinks about that it certainly implicates the universe in a kind of conspiracy that nowadays is hard to credit.

Finally, one might take the view (A3) that the correlation between the setting on the receivers and the prior state of the particles is *acausal*, so that here we have a counterexample to Reichenbach's principle that if a coincidence occurs then there must be a common cause.⁶ This interpretation of events is perhaps less unpalatable than the other options on this branch; but as we'll see, it seems less preferable than at least one option on the other and more popular branch.

(B) We can maintain that the prior instruction set—if it exists at all—is *not* correlated with your decision to put the receivers in any particular settings. Hence, we must accept Bell's finding that there *is* no prior instruction set. But now we are left with trying to explain fact (1): the fact that the two particles, when tested by receivers that are at the same setting, will *always* give the same results. There are two possible lines.

(B1) One might maintain that each particle carries its own instruction set—in effect a disposition to produce a reading S_i when placed in a receiver switched to i —but that there is some non-local causal connection, that is,

⁴ Price 1996.

⁵ Bell 1977: 100.

⁶ Reichenbach 1984: 157.

action at a distance between the particles. Labelling the particles A and B, we can write S^A_i and S^B_j , $1 \leq i, j \leq 3$ to specify these instruction sets.

So in particular and in spite of a spacelike interval, screening devices and any other causal barrier that one might erect between the receivers, one would nonetheless be claiming that choosing to subject particle A to a receiver in setting 1 somehow influenced its arbitrarily distant and isolated 'twin' to acquire a specification $S^B_1 = S^A_1$. Like the retrocausal interpretation (A1) then, this view does commit us to the existence of non-relativistic causality i.e. to the faster-than-light transmission of causal influence.⁷

(B2) One might alternatively take the line that there is a *non-causal* correlation between the specifications of the particles, so that learning that the first receiver has displayed y when switched to setting 1 (say) can certainly tell us that the second receiver will give the same result if switched to that setting, but that there is no *causal* connection underlying this correlation, so that any reading on any receiver is in fact causally independent of any reading on the other receiver. This reading has the disadvantage of hypothesizing (causally) unexplained correlations between the states on the two receivers. On the other hand we might consider it preferable to the reading (A3), which also postulates unexplained correlations but which must in addition appeal to a pre-existing instruction set.

Not all of these five approaches carry the same interest for decision theory, at least not for the clash between Causal and Evidential Decision Theory. For present purposes the responses of interest are (A2), (A3) and (B2). What unites these views is that they all reject certain causal dependencies, though different ones in different cases. Thus (A2) grants that there is a common cause of the prior state of the hidden variables and the experimenter's choice of setting. But it denies any causal dependence of the former upon the latter, as also does (A3), which in fact denies *all* causal dependencies. Option (B2) on the other hand denies that there is any prior state to *be* causally dependent or independent. But what is of interest is that *it* denies any causal dependency between the disposition of one particle to provoke such and such display to a receiver in such and such setting, and the corresponding disposition of the *other* particle.

The argument will then be as follows. First I'll argue that anyone who takes any of the views (A2), (A3) or (B2) is committed to a disparity between EDT and CDT in certain decision cases that I'll shortly describe. This will take slightly different arguments for (A)-type and (B)-type theories (that is, for

⁷ But not however to superluminal *signaling*: it has been shown (Eberhard 1978) that there is no way in which somebody operating the first receiver could exploit these correlations to send any sort of signal to somebody operating the second. This fact opens the door to a kind of 'peaceful co-existence' between quantum nonlocality and relativity, if we take the latter to be claiming only that there is no superluminal *signaling*, not that there is no superluminal causality of any sort. On the other hand there is something unsatisfactory about taking the relativistic restriction against superluminal causality to be a principle only about signaling, for as Bell himself wrote: 'the "no signaling ..." notion rests on concepts that are desperately vague, or vaguely applicable. The assertion that "we cannot signal faster than light" immediately provokes the question: Who do we think we are? We who make "measurements," we who can manipulate "external fields", we who can "signal" at all, even if not faster than light? Do we include chemists, or only physicists, plants, or only animals, pocket calculators, or only mainframe computers?' (Bell 1990: 246)

adherents of (A2) or (A3) on the one hand, and for adherents of (B3) on the other hand).

Then, I'll argue that this disparity raises two problems for the *causal* theory. The first is that its recommendations in both types of case seem to involve a bet against the laws of nature. The second is a more general point, which is that what CDT recommends in these cases is something that seems to vary depending on which of various and (as far as we know) empirically indistinguishable theories is true. Thus CDT appears to be oversensitive: its recommendations turn on matters that ought to be irrelevant to rational decision. If this second objection applies it appears to rule out any theory of rational decision based upon anything stronger than a Humean i.e. statistical conception of causation (Hume 1949: I.iii).

3 A-type theories

Suppose first that you accept (A2) or (A3) (or maybe only their disjunction): you think that there *is* a prior instruction set and that its state is causally independent of your current decision to switch the receivers to any particular settings, either because you think that both have a common cause or because you think that there are *no* relevant causal relations in play.

Now consider the following arrangement. Your options are to set the receivers in any of these three possible ways: you can set A to 1 and B to 2; A to 1 and B to 3; or A to 2 and B to 3. So on every available option the receivers are in *different* settings. At the same time you must also make a bet: you can bet, either ('hom') that the two receivers will display the *same* reading, or ('het') that they will display *different* readings. In effect you are choosing i, j , for $1 \leq i < j \leq 3$, and betting either that $S_i = S_j$ or that $S_i \neq S_j$. Finally, the 'hom' bets have a payoff of \$2 and the 'het' bets have a payoff of \$1.

So there are $2 \times 3 = 6$ options; I'll abbreviate these by two numbers to reflect the settings, followed by 'hom' or 'het' depending on whether you bet 'same' or 'different'. Thus e.g. '12hom' denotes the option of switching receiver A to setting 1, receiver B to setting 2 and betting that they will give the same readings, and '23het' denotes the option of switching receiver A to setting 2, receiver B to setting 3 and betting that they will give *different* readings.

According to the (A)-type theories that we are now considering, what determines the causal relation between the setting of each receiver and the reading that it displays is the prior common state of the particles YYY, NYN etc. This common state therefore also determines the payoff of each option. For instance, if you take the option '13het' then if the particles are in state YYY you will win \$0, since $S_1 = S_3$ and you have bet that $S_1 \neq S_3$. We may therefore take the instruction sets to be the relevant states of nature. The relations between these, your options and your payoffs are then as summarized in the following table:

	YYY	YYN	YNY	YNN	NYN	NYN	NNY	NNN
12hom	2	2	0	0	0	0	2	2
13hom	2	0	2	0	0	2	0	2
23hom	2	0	0	2	2	0	0	2
12het	0	0	1	1	1	1	0	0

13het	0	1	0	1	1	0	1	0
23het	0	1	1	0	0	1	1	0

Table 1: (A)-type quantum case 1

Notice that you can never be certain of which column actually obtains even after the run. But this makes no difference to the feasibility of the game, because your *payoff* is always fixed and verifiable. For instance, if you take the option 13het and both receivers read ‘y’, then you don’t know whether the prior state is YYY or YNY. But you *do* know that your payoff is \$0.

Let us now consider which of the six options EDT and CDT recommends. First consider EDT. In this case the matter is simple: if she is sensible then the agent will reflect in her conditional credences the relative frequencies as recorded in fact (2). In particular she will think: given that the receivers are in different settings—and regardless of whether I bet ‘hom’ or ‘het’—, prior instruction sets in which the corresponding states are *different* are about three times as likely as prior instruction sets in which the corresponding states are the *same*. So for i, j such that $1 \leq i < j \leq 3$, we have:

$$(13) \quad \text{Cr}(S_i = S_j \mid ij\text{hom}) = \text{Cr}(S_i = S_j \mid ij\text{het}) = 0.25$$

$$(14) \quad \text{Cr}(S_i \neq S_j \mid ij\text{hom}) = \text{Cr}(S_i \neq S_j \mid ij\text{het}) = 0.75$$

It follows from (13), (14) and Table 1 that the same V-score applies to any ‘het’ option and the same to any ‘hom’ option; also that the former exceeds the latter. For instance:

$$(15) \quad V(12\text{hom}) = 2\text{Cr}(YYY \mid 12\text{hom}) + 2\text{Cr}(YYN \mid 12\text{hom}) + \text{Cr}(NNY \mid 12\text{hom}) + 2\text{Cr}(NNN \mid 12\text{hom}) \text{ by Table 1}$$

$$(16) \quad V(12\text{hom}) = 2\text{Cr}(S_1 = S_2 \mid 12\text{hom}) = 0.5 \text{ by (13), (15)}$$

$$(17) \quad V(12\text{het}) = \text{Cr}(YNY \mid 12\text{hom}) + \text{Cr}(YNN \mid 12\text{hom}) + \text{Cr}(NYY \mid 12\text{hom}) + \text{Cr}(NYN \mid 12\text{hom}) \text{ by Table 1}$$

$$(18) \quad V(12\text{het}) = \text{Cr}(S_1 \neq S_2 \mid 12\text{het}) = 0.75 \text{ by (14), (17)}$$

And the same reasoning clearly goes for each of the other options. So EDT reckons the V-score of any ‘hom’ option (the first three options in Table 1) to be 0.5 and that of any ‘het’ option (the last three options there) to be 0.75. Accordingly EDT is indifferent between any of the ‘het’ options and prefers any of them to any ‘hom’ option⁸:

$$(19) \quad \text{For any } i, j, k, l \text{ s.t. } (1, 1) \leq (i, k) < (j, l) \leq (3, 3): ij\text{het} \succ_{\text{EDT}} kl\text{hom}$$

Turn now to CDT. Its recommendations won’t depend on the *conditional* credences $\text{Cr}(YYY \mid 12\text{hom})$ etc. but upon one’s credences in the *counterfactuals* $\text{Cr}(12\text{hom} \rightarrow YYY)$ etc. But given the theoretical assumptions (A2) or (A3), we know that the prior state is causally independent of one’s current setting of the receivers. Assuming (as is surely plausible) that your choice of bet (‘hom’ or ‘het’) makes no difference to that prior state either, it

⁸ In (19) I’m adopting the notational conventions that $(a, b) < (c, d)$ iff $a < c$ and $b < d$, and $(a, b) \leq (c, d)$ iff $a \leq c$ and $b \leq d$.

follows that $\text{Cr}(12\text{hom} \rightarrow \text{YYY}) = \text{Cr}(\text{YYY})$ etc. More generally, for any $1 \leq i < j \leq 3$, and $S_1, S_2, S_3 \in \{Y, N\}$:

$$(20) \quad \text{Cr}(ij\text{hom} \rightarrow S_1S_2S_3) = \text{Cr}(ij\text{het} \rightarrow S_1S_2S_3) = \text{Cr}(S_1S_2S_3)$$

It follows from (20) that the U-scores for the three 'hom' options are as follows:

$$(21) \quad U(12\text{hom}) = 2(\text{Cr}(\text{YYY}) + \text{Cr}(\text{YYN}) + \text{Cr}(\text{NNY}) + \text{Cr}(\text{NNN}))$$

$$(22) \quad U(13\text{hom}) = 2(\text{Cr}(\text{YYY}) + \text{Cr}(\text{YNY}) + \text{Cr}(\text{NYN}) + \text{Cr}(\text{NNN}))$$

$$(23) \quad U(23\text{hom}) = 2(\text{Cr}(\text{YYY}) + \text{Cr}(\text{YNN}) + \text{Cr}(\text{NYY}) + \text{Cr}(\text{NNN}))$$

And the U-scores for the three 'het' options are as follows:

$$(24) \quad U(12\text{het}) = \text{Cr}(\text{YNY}) + \text{Cr}(\text{YNN}) + \text{Cr}(\text{NYY}) + \text{Cr}(\text{NYN})$$

$$(25) \quad U(13\text{het}) = \text{Cr}(\text{YYN}) + \text{Cr}(\text{YNN}) + \text{Cr}(\text{NYY}) + \text{Cr}(\text{NNY})$$

$$(26) \quad U(23\text{het}) = \text{Cr}(\text{YYN}) + \text{Cr}(\text{YNY}) + \text{Cr}(\text{NYN}) + \text{Cr}(\text{NNY})$$

Now *suppose* that each 'het' option that gets a higher U-score than its corresponding hom options. Then all of the following must be true:

$$(27) \quad U(12\text{het}) > U(12\text{hom})$$

$$(28) \quad U(13\text{het}) > U(13\text{hom})$$

$$(29) \quad U(23\text{het}) > U(23\text{hom})$$

Substituting (21)-(26) into (27)-(29) and adding the three resulting inequalities gives:

$$(30) \quad 2(\text{Cr}(\text{YYN}) + \text{Cr}(\text{YNY}) + \text{Cr}(\text{YNN}) + \text{Cr}(\text{NYY}) + \text{Cr}(\text{NYN}) + \text{Cr}(\text{NNY})) > 6(\text{Cr}(\text{YYY}) + \text{Cr}(\text{NNN})) + 2(\text{Cr}(\text{YYN}) + \text{Cr}(\text{YNY}) + \text{Cr}(\text{YNN}) + \text{Cr}(\text{NYY}) + \text{Cr}(\text{NYN}) + \text{Cr}(\text{NNY})); \text{ hence}$$

$$(31) \quad 0 > 6(\text{Cr}(\text{YYY}) + \text{Cr}(\text{NNN}))$$

But (31) is a contradiction since the credences on the right hand side are both *at least* zero⁹, and so the supposition that entails (27)-(29) must be false. There must be *some* 'hom' option that gets at least as high a U-score as its corresponding 'het' option. This is the only thing that is consistent with one's having any credences about the prior state at all:

$$(32) \quad \text{For some } i, j \text{ s.t. } 1 \leq i < j \leq 3: ij\text{hom} \succeq_{\text{CDT}} ij\text{het} \text{ in Table 1}$$

But taken together (19) and (32) imply that there must be some pair of options $ij\text{hom}$ and $ij\text{het}$ over whose relative ranking EDT and CDT disagree. In particular let this be the pair 12hom and 12het. Then:

$$(33) \quad V(12\text{het}) > V(12\text{hom})$$

$$(34) \quad U(12\text{hom}) \geq U(12\text{het})$$

⁹ I am here assuming that we are not invoking negative probabilities, as some quantum theorists have suggested (e.g. Muckenheim 1982).

Finally, consider the same decision situation as before but with payoffs that make irrelevant all of the options except for these two:

	YYY	YYN	YNY	YNN	NYY	NYN	NNY	NNN
12hom	2	2	0	0	0	0	2	2
13hom	0	0	0	0	0	0	0	0
23hom	0	0	0	0	0	0	0	0
12het	0	0	1	1	1	1	0	0
13het	0	0	0	0	0	0	0	0
23het	0	0	0	0	0	0	0	0

Table 2: (A)-type quantum case 2

Plainly nothing about the difference in payoffs under the *other* options makes a difference to the V-scores and the U-scores of 12hom and 12het, to which (33) and (34) still apply. It's also clear that *both* EDT *and* CDT take 12hom and 12het to be at least as good as any other option in this case.¹⁰ So in the case that Table 2 describes, EDT and CDT make different recommendations: EDT recommends only 12het, which gets a V-score 0.75, and CDT endorses 12hom, which is getting some unknown U-score that is no less than U (12het).

So we see that EDT and CDT give different recommendations in the case at Table 2 to anyone that accepts interpretations (A2) or (A3) of the EPR phenomena that I described in section 1. This situation, although lacking the *prima facie* realism of the medical Newcomb cases, is in fact relatively plausible. It is in fact technologically feasible today—something that you could not say *either* of the standard Newcomb case (which involves supernatural 'predictors') *or* of the medical cases (which involve correlations between physical state and choice that are unknown to medical science).

One immediate objection is that although the technical apparatus needed to arrange for a situation like Table 2 is in no way fantastical, still the existence of a disagreement between EDT and CDT does require some rather unusual beliefs on the part of the subject. In particular he must believe that there *is* a prior state ('hidden variables')¹¹; and this is something that many physicists have taken *not* to be a live option in face of Bell's Theorem.

Of course this objection needn't be fatal to the broader point that I'll use the example to make. As we'll see, all that I needed for that purpose was a case where EDT and CDT lead in different directions an agent whose beliefs are at least sane and coherent; the fact that these beliefs represent a minority position doesn't by itself make the case any more irrelevant than standard Newcomb cases. On the other hand the oddity of the perspective from which Table 2 forces this divergence inevitably diminishes the interest of the case. I turn therefore to the more plausible interpretation (B2), on which

¹⁰ For EDT this is clear from (16), (18) and the fact that every other option in Table 2 gets V-score 0. For CDT it follows from the fact that both 12hom and 12het weakly dominate all of the other four options.

¹¹ Of course, he must also believe that the prior state either (A2) causes or shares a common cause or (A3) is acausally correlated with one's present choice to set the receivers this or that way. But *this* is not implausible: given that one *is* already on the (A)-branch and so has swallowed hidden variables themselves, to balk at the idea that they lack retrocausal powers (these being the only alternative to (A2) and (A3)) is arguably straining at a gnat.

there are no hidden variables but the correlations between manifest states are acausal.

4 B-type theories

Let us then suppose what many physicists actually do think about this case, namely that there is a non-causal correlation between the results of measuring the spin of the particles along any particular pair of directions. (Recall that the correlation is this: if the receivers measure along directions that are separated by an angle θ , then the probability of getting a matching reading is $\cos^2(\theta/2)$.)

On this supposition there is no possibility of a situation quite like Tables 1 and 2, because there is no common prior state on which to 'bet'. But it *is* still feasible to offer and take bets on the displays on the two receivers, taken either individually or together. For instance: you might have the option, just prior to a particular run of the device, to switch receiver A to setting 1 and bet that its *display* will be 'y'. As a (B)-theorist, if we can call you that, you're *not* betting on the prior state of the hidden variables (in particular, you're not betting in this case that $S_1 = Y$). From your perspective, that's like betting on yesterday's weather in Narnia: there *is* no such state to bet *on*. But you *are* making a bet that is operationally bona fide: given your choice, the observable display on the receiver will invariably settle your monetary payoff.

For instance, '12het' still represents a feasible option: you are switching receiver A to setting 1, receiver B to setting 2, and you are betting that they will *display the same reading*. Even more simply, one can bet on the readings of either receiver taken individually: that is, one might bet e.g. that receiver B will display 'y' on the next run. These bets, and any others that depend for their payoffs only on verifiable events like displays on receivers, are ones that punters will certainly either win or lose. So even if we deny the existence of any instruction set that determines their outcomes in advance, both Evidential and Causal Decision Theories should tell us at what odds they represent good value, which out of many such bets to choose, etc.

It's clear enough that Evidential Decision Theory applies to such bets. The only credences that it needs agents to have are conditional credences on (say) the display on a receiver given that one takes this or that option. And *these* conditional credences are certainly available whether or not one accepts that there is any prior instruction set. In the simple scenarios that we consider here, they reflect available statistical records correlating options with outcomes, i.e. facts (1) and (2).

And *prima facie* there isn't any problem for Causal Decision Theory either. Consider for instance the case where one has three options O_1 , O_2 and O_3 , these being respectively the options of switching receiver A into setting 1, 2 or 3 whilst simultaneously betting *that the reading* on the receiver will be y. One is not here betting on any *prior* state (i.e. on the proposition $S_1 = Y$) but rather on a *subsequent* one that may or may not be causally dependent on the choice of setting.¹² The difference is that we cannot calculate the U-score of an option by partitioning on prior states of the world that (a) obtain causally

¹² Remember, (B2) is not denying that switching either receiver to this or that setting has any causal influence on *its own* reading. Rather what it denies is any superluminal, retroactive or common causality between anything going on in the region of one receiver, including its setting, and anything going on in the region of the other.

independently of the option chosen whilst (b) determining the chance of each option's producing this or that payoff. There is no such state of the world. So instead we are forced back on the direct calculation of U-scores by means of counterfactual credences themselves.

For instance, in the case at hand there are two possible readings on each receiver and so four possible combinations of each reading. Letting 'yn' correspond to 'y' on receiver A and 'n' on receiver B etc., the payoffs are as follows:

	yy	yn	ny	nn
O1	1	1	0	0
O2	1	1	0	0
O3	1	1	0	0

Table 3: illustrative quantum game without h.v.

The following expression therefore gives the U-score of, say, O1:

$$(35) \quad U(O1) = Cr(O1 \rightarrow yy) + Cr(O1 \rightarrow yn)$$

But in order to calculate the credences in (35) we cannot (as I said) partition over prior states of the world; instead we must directly evaluate these credences by means of formulas of this type:

$$(36) \quad Cr(O1 \rightarrow yy) = \int_{0 \leq x \leq 1} x \cdot Cr(Ch(yy | O1) = x) dx$$

—in which Cr expresses one's distribution function over the possible values for the conditional chance that a particular setting of receiver A gives to some particular combination of readings on receiver A and receiver B. Still, despite this change in the manner of calculating the U-score, it is easy to see that CDT will give some advice; and in this case, which in no way exploits the statistical peculiarity of the quantum world, there is no obvious reason why this advice should diverge from that of EDT.

But all of this changes when we turn to types of problems that exploit facts (1) and (2). The first of these is a family of cases D (i, z) for i = 1, 2, 3 and 0 ≤ z ≤ 1. Each one takes the following form: you may set both receivers to the *same* setting i, say setting 1. You then win \$(1 - z) if both receivers give the *same* reading. But you *lose* \$z if the readings are different. The alternative option Q ('quit') is to decline any bet. The payoffs for any particular D (i, z) are therefore as follows (remember that the headings to each column now describe the *readings* on the receivers, not any prior state):

	yy	yn	ny	nn
iihom	1-z	-z	-z	1-z
Q	0	0	0	0

Table 4: (B)-type quantum case 1: D (i, z)

What do EDT and CDT advise now?

Remember that in any D (i, z) we are, if we bet, switching the receivers to the same setting (i.e. both to 1, both to 2 or both to 3). Fact (1) therefore

assures us that they will always give the same reading, if we bet. So the relevant conditional credences are as follows:

$$(37) \quad \text{Cr}(yy \vee nn \mid \text{iihom}) = 1$$

$$(38) \quad \text{Cr}(yn \mid \text{iihom}) = \text{Cr}(ny \mid \text{iihom}) = 0$$

It follows that for any i , z the V-scores of the options in $D(i, z)$ are:

$$(39) \quad V(\text{iihom}) = 1 - z$$

$$(40) \quad V(Q) = 0$$

Hence for any $i = 1, 2, 3$ and z such that $0 \leq z \leq 1$ EDT will at least endorse playing iihom in $D(i, z)$; if in addition $z < 1$ it will definitely prefer iihom to Q . In other words it always endorses and sometimes requires that you should bet on the receivers giving the same reading if they are on the same setting.

What about CDT? For any i and z , the U-scores of the options in $D(i, z)$ are as follows:

$$(41) \quad U(\text{iihom}) = (1 - z)(\text{Cr}(\text{iihom} \rightarrow yy) + \text{Cr}(\text{iihom} \rightarrow nn)) - z(\text{Cr}(\text{iihom} \rightarrow yn) + \text{Cr}(\text{iihom} \rightarrow ny))$$

$$(42) \quad U(Q) = 0$$

Now consider the quantity $(\text{Cr}(\text{iihom} \rightarrow yn) + \text{Cr}(\text{iihom} \rightarrow ny))$ on the right-hand side of (41). It's easy to see that *if* this quantity > 0 then there is some strictly positive $z^* < 1$ such that $U(\text{iihom}) < U(Q)$ in $D(i, z)$ for any $z \geq z^*$.¹³ In other words we have a continuum of decision situations in which CDT and EDT diverge i.e. for any $z > z^*$, in $D(i, z)$ CDT will endorse quitting but EDT will endorse betting.

The situation would look like this. You have in effect the option of paying a fee of $\$z$ to take a bet that pays $\$1$ if you win and $\$0$ if you lose; and you will win if quantum mechanics is true, or at any rate if Fact (1) is something that can be relied upon. The evidentialist will therefore pay any fee short of $\$1$ to take this bet. But the causalist will decline the bet at any fee beyond some threshold $\$z^* < \1 .

So if both are offered these bets at a rate $\$z > \z^* the causalist will keep declining (and winning nothing) and the evidentialist will keep accepting (and winning $\$(1 - z)$). For instance, suppose we have $z^* = 0.8$. Then we can keep charging both parties 90¢ for a bet that pays $\$1$ iff both receivers give the same reading on the next run in which they are switched to the same setting. Then the evidentialist will always accept and the causalist will always decline, and the evidentialist will make 10¢ over the causalist every time. So here we have decision problem over which EDT and CDT disagree that does not depend on the assumption of hidden variables.

But it *does* depend on a different assumption, namely that for some i the factor $(\text{Cr}(\text{iihom} \rightarrow yn) + \text{Cr}(\text{iihom} \rightarrow ny))$ exceeds 0. If we drops this

¹³ Since $\text{Cr}(O1 \rightarrow yy) + \text{Cr}(O1 \rightarrow nn) + \text{Cr}(O1 \rightarrow yn) + \text{Cr}(O1 \rightarrow ny) = 1$, it is possible to write $U(\text{iihom})$ as $(1 - z)x + z(1 - x) = x - z$, where $1 - x = \text{Cr}(\text{iihom} \rightarrow yn) + \text{Cr}(\text{iihom} \rightarrow ny)$. So if $1 - x > 0$ then $x < 1$, hence there is some $z^* < 1$ s.t. $z^* > x \geq 0$. So if $z \geq z^*$ then $U(\text{iihom}) < 0 = U(Q)$.

assumption then the causalist will only decline the bet at $z \geq z^* = 1$, at which EDT will also endorse not betting (because the expected value of *i*ihom is now 0). So assuming that $0 \leq z \leq 1$ there is in that case no D (*i*, *z*) in that continuum of decision situations on which EDT and CDT diverge.

But as I'll now argue, if we *do* drop the assumption then there will inevitably be *another* quantum situation in which EDT and CDT disagree.

Suppose that we drop it. Since *Cr* is a probability function the only alternative is that for any $i = 1, 2, 3$:

$$(43) \quad Cr(iihom \rightarrow ny) = Cr(iihom \rightarrow yn) = 0$$

Now consider a decision situation just like the six-option case that we considered in section 3 in connection with the hidden-variable theories (A2) and (A3). You can choose any of three joint settings for each receiver: A on 1 and B on 2, A on 2 and B on 3, and A on 3 and B on 3. And for each setting you can bet either that the receivers will display the same reading on the next run or that they will display a different reading on the next run. As before I'll label the six resulting options 12hom, 13hom, 23hom, 12het, 13het and 23het.

The payoffs depend, determinately and in a decidable manner, on the readings on the receivers. These are similar to those in Table 1, except that now we have 'yy' etc. instead of 'YYY' etc. at the top of each column, to indicate that we are making a bet on the readings of the receivers themselves, without speculating about any prior state:

	yy	yn	ny	nn
12hom	2	0	0	2
13hom	2	0	0	2
23hom	2	0	0	2
12het	0	1	1	0
13het	0	1	1	0
23het	0	1	1	0

Table 5: (B)-type quantum case 2

What does EDT recommend in this situation? Here again the answer is quite straightforward if we suppose that your conditional credences reflect the statistical regularities (1) and (2). In particular then, the reasoning behind (16) and (18) applies in this case too and shows that any 'het' option (which gets a V-score of 0.75) is EDT-preferred to any 'hom' option (which gets a V-score of 0.5). So just as before and as you'd expect, even on this no-hidden-variables hypothesis EDT prefers any 'het' option to any 'hom' option. Whether or not there are 'hidden variables' is a purely theoretical question that makes no difference to the observed outcomes and so no difference to the practical advice that EDT gives.

When we turn to CDT, things are different. It would be nice to be able to represent the problem in terms of 'states of nature' that are causally independent of the agent's options and which together with the options determine the payoff. Table 5 is *not* such a representation, for there is no guarantee that e.g. the display on receiver A is causally independent of whether one sets that receiver to 1 or 2. Nor is it obvious that there *is* such a state.

This is in fact the crucial point of contrast between Table 1 and Table 5. In the case of Table 1 the prior instruction set is both causally independent of one's choice and determinative of one's payoff in conjunction with one's choice. But since there *is* no prior instruction set on B-type interpretations of the experiment, we cannot take any partition over its possible configurations as our set of 'causally independent states of nature'.

But it turns out that even though there is no prior instruction set, we can still generate a partition that plays exactly the same role as it. In intuitive terms, the argument is that the agent's hypothesized response to D (i, z) forces his credences to *mimic* those of an agent facing Table 1; in particular, what play the roles of configurations of the prior instruction set are *conditional chances of readings given settings*. The argument turns on four points.

(i) It is surely absurd to suppose that the choice of bet between 'het' and 'hom' makes any difference to the reading on either receiver once we are given their settings. We could in any case could impose this condition by brute force: i.e., by requiring that the agent chooses the kind of bet on any given run (i.e. between 'hom' and 'het') *after* the run is over but *before* she has had a chance to see the relevant readings. So we can write e.g.:

$$(44) \quad \text{Ch}(yy \mid 12\text{het}) = \text{Ch}(yy \mid 12\text{hom}) = \text{Ch}(yy \mid 12)$$

(ii) Since the readings on the receivers are by (B2) causally independent of one another, the chance of either reading on either receiver is independent of the setting on the other receiver, even given the settings on both receivers. So if we label the receivers 'A' and 'B', and if we write $1_A, y_B$ etc. for the settings and the readings on each receiver, then we have e.g.:

$$(45) \quad \text{Ch}(yy \mid 12) = \text{Ch}(y_A y_B \mid 1_A 2_B) = \text{Ch}(y_A \mid 1_A 2_B) \text{Ch}(y_B \mid 1_A 2_B)$$

$$(46) \quad \text{Ch}(yy \mid 22) = \text{Ch}(y_A y_B \mid 2_A 2_B) = \text{Ch}(y_A \mid 2_A 2_B) \text{Ch}(y_B \mid 2_A 2_B)$$

Note that, as (46) illustrates, this point of course applies to chances conditional on any pair of settings, including those that the present decision problem does not associate with any bet.

(iii) The reading on either receiver is causally independent of the setting on the other receiver given its own setting. (This follows from the assumption that there is no prior instruction set.) So we have e.g.:

$$(47) \quad \text{Ch}(y_A \mid 1_A 2_B) = \text{Ch}(y_A \mid 1_A)$$

$$(48) \quad \text{Ch}(y_B \mid 2_A 2_B) = \text{Ch}(y_B \mid 2_B)$$

Again and as (48) illustrates, this point applies to all pairs of settings, not only those that the present decision situation associates with bets.

(iv) The fourth simplification goes back to the family of decision problems D (i, z). Recall that if EDT and CDT agree over all of those cases then (43) must be true. (And if they do not, then we have already found a decision-theoretic case on which they disagree and to which all of the forthcoming arguments will apply.

Now it follows from (43) that the agent is certain of:

$$(49) \quad \text{Ch}(yy \vee nn \mid 11) = 1$$

$$(50) \quad \text{Ch}(yy \vee nn \mid 22) = 1$$

$$(51) \quad \text{Ch}(yy \vee nn \mid 33) = 1$$

Focusing on (49)—the argument from (50) and (51) is parallel—we see that:

$$(52) \quad \text{Ch}(yy \mid 11) + \text{Ch}(nn \mid 11) = 1$$

And so by (46) and (48) we have:

$$(53) \quad \text{Ch}(y_A \mid 1_A) \text{Ch}(y_B \mid 1_B) + \text{Ch}(n_A \mid 1_A) \text{Ch}(n_B \mid 1_B) = 1$$

And corresponding results follow from (51)-(53).

Now from (55) and its analogues we get:¹⁴

$$(54) \quad \text{Ch}(y_A \mid 1_A) = \text{Ch}(y_B \mid 1_B) \in \{0, 1\}$$

$$(55) \quad \text{Ch}(y_A \mid 2_A) = \text{Ch}(y_B \mid 2_B) \in \{0, 1\}$$

$$(56) \quad \text{Ch}(y_A \mid 3_A) = \text{Ch}(y_B \mid 3_B) \in \{0, 1\}$$

This gives eight possibilities for the values of these conditional chances depending on which ones take the value 1 and which take the value 0.

Now putting these four points together: we see that the conditional chance of each reading (yy, yn etc.) on each option (12hom, 13het etc.) is either 1 or 0; and this is determined by which of the eight possibilities just outlined obtains. For instance, suppose that the following situation obtains:

$$(57) \quad \text{Ch}(y_A \mid 1_A) = \text{Ch}(y_B \mid 1_B) = 1$$

$$(58) \quad \text{Ch}(y_A \mid 2_A) = \text{Ch}(y_B \mid 2_B) = 0$$

$$(59) \quad \text{Ch}(y_A \mid 3_A) = \text{Ch}(y_B \mid 3_B) = 1$$

Then by (44), (45) and (47) it follows that:

$$(60) \quad \text{Ch}(yy \mid 12het) = 0$$

More generally, any specification of the conditional chances in (54)-(56), together with a specification of the agent's option, determines the reading on the receivers and hence the agent's payoff. Finally, whichever of the eight possibilities in (54)-(56) obtains is causally independent of what the agent chooses: for if a conditional chance takes either value 0 or value 1 then nothing that you do could *make* any difference to that (of course this point doesn't hold for conditional chances that are strictly *between* 0 and 1).

What this means is that we can rewrite the decision problem in Table 5 in terms of states of nature that are causally independent of one's choice in that situation. To that end I'll use the following three numeral code: 'abc', where a, b, c $\in \{0, 1\}$, means:

¹⁴ To see this consider that (55) is of the form $xy + (1-x)(1-y) = 1$. So for $0 \leq x, y \leq 1$ the only solutions are $x = y = 0$ and $x = y = 1$.

$$(61) \quad \mathbf{abc} \equiv_{\text{def}} \text{Ch}(y_A | 1_A) = \text{Ch}(y_B | 1_B) = a \wedge \text{Ch}(y_A | 2_A) = \text{Ch}(y_B | 2_B) \\ = b \wedge \text{Ch}(y_A | 3_A) = \text{Ch}(y_B | 3_B) = c$$

So for instance, ‘101’ corresponds to the possible distribution stated at (57)-(59). The new representation of the problem then looks like this:

	111	110	101	100	011	010	001	000
12hom	2	2	0	0	0	0	2	2
13hom	2	0	2	0	0	2	0	2
23hom	2	0	0	2	2	0	0	2
12het	0	0	1	1	1	1	0	0
13het	0	1	0	1	1	0	1	0
23het	0	1	1	0	0	1	1	0

Table 6: (B)-type quantum case 2 with independent states of nature

And since the ‘states of nature’ that the top row represents—that is, the possible distributions of conditional chances of readings on setting—are independent of whatever option is chosen, the calculation of the U-score for each option is a straightforward matter. In particular, the U-scores for the three ‘hom’ options are:

$$(62) \quad U(12\text{hom}) = 2 (\text{Cr}(\mathbf{111}) + \text{Cr}(\mathbf{110}) + \text{Cr}(\mathbf{001}) + \text{Cr}(\mathbf{000}))$$

$$(63) \quad U(13\text{hom}) = 2 (\text{Cr}(\mathbf{111}) + \text{Cr}(\mathbf{101}) + \text{Cr}(\mathbf{010}) + \text{Cr}(\mathbf{000}))$$

$$(64) \quad U(23\text{hom}) = 2 (\text{Cr}(\mathbf{111}) + \text{Cr}(\mathbf{100}) + \text{Cr}(\mathbf{011}) + \text{Cr}(\mathbf{000}))$$

And the U-scores for the three ‘het’ options are:

$$(65) \quad U(12\text{het}) = \text{Cr}(\mathbf{101}) + \text{Cr}(\mathbf{100}) + \text{Cr}(\mathbf{011}) + \text{Cr}(\mathbf{010})$$

$$(66) \quad U(13\text{het}) = \text{Cr}(\mathbf{110}) + \text{Cr}(\mathbf{100}) + \text{Cr}(\mathbf{011}) + \text{Cr}(\mathbf{001})$$

$$(67) \quad U(23\text{het}) = \text{Cr}(\mathbf{110}) + \text{Cr}(\mathbf{101}) + \text{Cr}(\mathbf{010}) + \text{Cr}(\mathbf{001})$$

But these scores exactly parallel the U-scores of the corresponding options in the (A)-type quantum case 1, except with ‘1’ and ‘0’ in place of ‘Y’ and ‘N’ respectively. See Table 1 and equations (21)-(26). So from this point we can apply exactly parallel reasoning as that applied to the (A)-type case at the corresponding point in the argument, since nothing in *that* part of the argument (steps (27)-(32)) depended on any special feature of the hidden variables interpretation but only on the fact that Cr is a probability function.

Without explicitly repeating the reasoning to it, I therefore draw a conclusion that parallels that for the (A)-type case. There must be some ‘hom’ option that CDT takes to be at least as good as the corresponding ‘het’ option in Table 6. But since the options in table 6 just *are* the options in Table 5, this means that CDT must consider some ‘hom’ option to be at least as good as the corresponding ‘het’ option there too. Combining that with the entirely straightforward reasoning about EDT that immediately followed Table 5:

$$(68) \quad \text{For any } i < j: ij\text{het} \succ_{\text{EDT}} ij\text{hom in Table 5}$$

$$(69) \quad \text{For some } i < j: ij\text{hom} \succeq_{\text{CDT}} ij\text{het in Table 5}$$

Without loss of generality we can take $I = 1$ and $j = 2$ to be witnesses of (69), on which case the following decision situation represents a technically feasible quantum case in which EDT and CDT give conflicting advice to anyone who rejects both hidden variables and non-relativistic causation.

	yy	yn	ny	nn
12hom	2	0	0	2
13hom	0	0	0	0
23hom	0	0	0	0
12het	0	1	1	0
13het	0	0	0	0
23het	0	0	0	0

Table 7: (B)-type quantum case 3

In table 7 CDT endorses the first option 12hom, whereas EDT recommends only the fourth option 12het.

It may be worth briefly stepping back from the formal details to give an intuitive overview of the construction. The basic idea for (B)-type cases is that even in the absence of hidden variables, anyone who thinks that the receivers are causally independent must think that only *its own* setting is causally relevant to the reading on any receiver. If in addition this person thinks that when the receivers are in the same setting they always force the same reading (as he must do if he takes every bet in the family $D(i, z)$), then he is committed to saying that the causal relevances pertaining to each receiver are (a) perfectly synchronized; (b) completely deterministic. In short, any gambler who takes fact (1) seriously when the receivers are at the same setting must also be betting as if there was a prior instruction set when the receivers are at *different* settings.¹⁵

¹⁵ It's also worth contrasting the construction in this paper with two other attempts (the only ones known to me) to exploit violations of the Bell inequalities in order to make EDT and CDT disagree. Berkovitz's example (1995) assumes that the agent rejects all of the (A)- and (B)-hypotheses and instead believes in a prior instruction set that is *uncorrelated* with her setting of the receiver. It therefore depends on a theoretical assumption that is demonstrably false and so is no more realistic than the supernaturalistic Newcomb cases on which we had been seeking an improvement.

Cavalcanti's argument (2010), which invokes the CHSH arrangement (Clauser et al. 1969), appears to mischaracterize the causal theory. His case depends crucially on there being *two* agents, one at each wing of the experiment. But his calculation of the U-score of any option available to *one* of these agents treats both agents' choices as actions i.e. ignores their evidential bearing on anything other than their effects. (A formal symptom of this is the symmetric treatment of the terms 'A_R' and 'B_G' in his equation (16).) But this is a mistake: from the point of view of either experimenter the other agent's choice—which is not up to *her*—itself partly characterizes the 'state of nature', and her credence should reflect this. Cavalcanti's reasoning that the causalist must bet against quantum mechanics in these scenarios (2010: 585-6) is therefore invalid.

In any case Cavalcanti's argument concerns only the case in which the agent believes in a prior instruction set (i.e. the analogues of what I called (A)-type interpretations of the Stern-Gerlach experiment). He does mention (2010: 589) his own *belief* that CDT's advice in these cases would carry over to the case where the agent rejects any hidden variables (in particular to the case that I called (B2)); but he gives almost no argument that this is so. (There is a one-sentence argument to this effect at 2010: 589, which however the already-mentioned mischaracterization of CDT entirely vitiates.) It turns out that his suspicion is

It seems to me that this general pattern of reasoning may have fruitful application outside of decision theory, for instance to the vexed question of how fruitfully to characterize the doctrine of metaphysical realism. But I cannot pursue that here.

5. QM vs CDT

Of course the fact that a conflict between EDT and CDT can feasibly arise, at least on interpretations (A2), (A3) and (B3) of the experiment, does not by itself refute either theory. But it does make especially vivid just what is involved in preferring CDT to EDT. It is a realistic case where they genuinely clash; and it lacks all of the psychological clutter of ‘tickles’ and other forms of self-knowledge that so gummed up the works of previous attempts to come up with realistic cases where the theories gave different advice.¹⁶

And on reflection it prompts two obvious objections to CDT. The first is familiar: ‘Why ain’cha rich?’ CDT advises anyone who accepts interpretation (A2) or (A3) to take option 12hom in Table 2 whereas EDT will advise 12het. Similarly that CDT advises anyone who accepts (B2) to take option 12hom in the decision problem in Table 7¹⁷; whereas EDT will again insist on 12het. And everyone knows what will happen in either case. CDT i.e. 12hom will on average win \$2 in one out of every four runs. EDT i.e. 12het will on average win \$1 in three out of every four runs. So EDT is making \$1.50 for every \$1 that CDT is making. Everyone knows in advance that EDT will outperform CDT. How could you rationally recommend or follow a strategy that you *know* is going to underperform?

In terms of its form there is nothing new about this point, which dates back to early discussions of Newcomb’s problem.¹⁸ What is new is the context, which is naturalistic by the usual standards of these debates and, I believe, all the more vivid for all that.

Putting it a bit more precisely: it’s true that the example makes some demands on the reader’s beliefs. We are asking her to imagine that this highly unusual setup exists (though in fact it does) and that the statistics recording its performance are as described in (1) and (2) (though in fact they are). We are asking her to imagine payoff structures as described in tables 1-7 (though in fact these could easily be arranged). Much more seriously, we are asking the reader to accept one of three specific interpretations of the Stern-Gerlach experiment i.e. (A2), (A3) or (B2). And this is indeed something of a stretch: it is hard to believe, for instance and as (B2) would have it, that there really are lawlike (i.e. theoretically predicted) correlations between states that are causally independent and share no causal ancestor.¹⁹

correct. But it has taken some work to show this, (including the invention of a totally new family of problems $D(i, z)$).

¹⁶ E.g. the ‘medical’ Newcomb problems, on which see Price 1991 and Price 2012: 511-13.

¹⁷ At any rate this is so if the agent’s credences satisfy (43). If they do *not* satisfy (43) then there is some other situation $D(i, z)$ for $0 < z < 1$ in which EDT and CDT give conflicting advice to anyone that accepts (B2)—see Table 4. And it is in this scenario that we can then expect CDT to underperform relative to EDT, and the forthcoming remarks in the main text go through *mutatis mutandis* for it.

¹⁸ See e.g. Gibbard and Harper 1978: 369.

¹⁹ On the other hand this claim is not the contradiction that Maudlin appears to imply that it is when he writes that: ‘if a theory predicts a correlation, then that correlation cannot, according

I should like to put the point as strongly as this. Focus on (B)-type quantum case 3 as represented in Table 7 and suppose that all parties' credences make it a site of conflict between EDT and CDT. Let you and I be two financiers, and suppose that we take it in turns to choose an option from Table 7. On your turns, I pay you what you win; on your turn you pay me what I win. So if I follow EDT and you follow CDT then I will on average win \$3 from you on my goes, and lose \$1 to you on your goes, every eight runs. I hereby publicly challenge (or it would be public if anyone is still reading) any defender of CDT to play this game against me. Unfortunately I am certain that Lewis, Pearl et al. would if faced with this situation stop being causalists long before they stopped being solvent.

The second objection is not that CDT is giving bad advice in any identifiable case, but that what advice CDT is giving turns on theoretical questions that are (today, and perhaps in principle) impossible to settle by means of observation and experiment.

For instance, if you think that retrocausality is a live option then you may well take (A1) to be the—or at least: a possibly—correct description of what is happening in the Stern-Gerlach experiment; the same goes for action at a distance in connection with (B1). Of course nothing in the bare statistics forces either interpretation upon us; and yet the practical advice that CDT gives does depend on whether we adopt one of these interpretations or rather instead on of the non-causal interpretations (= (A2), (A3), (B2)). CDT prefers 12hom to 12het in e.g. Table 2 if we are given (A2) or (A3); but it reverses this preference on the hypothesis (A1). Similarly, on hypothesis (B2) CDT advises either not betting in some D (i, z) as in Table 4 or 12hom over 12het in Table 7; but again, it reverses these preferences given (B1). In short its recommendation depends not only upon the statistical facts that we can observe but also upon theoretical questions that they do not (and which maybe nothing ever could) settle.

But it should seem strange that the answer to a practical question ('Which bet?') turns on relatively abstruse theoretical matters. After all, nothing about the theoretical situation has any impact upon the facts that will actually settle your payoffs. We know in advance what these are. We know in advance that whether or not e.g. retrocausality is operating, the return to 12hom in Table 7 will in the long run exceed the return to 12het by 50%.

To make it more vivid: suppose that I am running Table 7-style books on two similar Stern-Gerlach devices, X and Y, and that you for some reason

to the theory, be accidental. A nomic correlation is indicative of a causal connection—immediate or mediate—between the events, and is accounted for either by a direct causal link between them, or by a common cause of both' (2002: 90).

But this argument involves a loaded understanding of 'accidental'. If 'accidental according to the theory' just means *not predicted by the theory* then of course the claim that no theory predicts accidental correlations is a mere tautology but hardly entails that that theory has any causal commitments. On the other hand if 'accidental according to the theory' means *has no causal explanation according to the theory* then certainly there are theories that predict 'accidental' correlations; but this, according to their advocates, reflects the *insight* that we should stop looking for causal explanations at this level (van Fraassen 1991: 372-4). Finally, if we simply define 'causality' in such a way as to be somehow involved in any nomic connection, then Laudisa's remark is apt. 'What we are doing... is nothing but saying that "connected events are connected"... using causal concepts in this case appears then to be a mere labeling devoid of any real physical and philosophical significance' (Laudisa 2001: 229).

think that action-at-a-distance is operative in X but not in Y, the devices being otherwise identical. Then CDT will advise different policies for both of them, even though you know in advance that they will generate the same payoffs to the same strategies. Worse still: suppose you forget which device is X and which is Y, and I offer to remind you for a fee. If you expect to play many times then CDT recommends that you pay up, even though you can be certain of the same return whether you play 12het on X and 12hom on Y or vice versa.

This complaint against CDT goes to the heart of what distinguishes it from the evidential theory. It makes a practical question of what to do depend on possibly irresolvable metaphysical matters that have no observable consequences. That in turn appears to implicate it in a complete misconception of what practical reasoning involves and why it should matter: to give non-trivially different practical advice in practically indistinguishable situations is to fail to understand that you are giving *practical* advice not theoretical speculation.

This aspect of CDT is not present in other cases that distinguish it from EDT. In standard Newcomb cases (Nozick 1970: 207-8) the causal structure of the situation is clear because stipulated: there simply *is* no retrocausality or action at a distance from your decision to the state of nature that it reveals, in this case the prediction²⁰; similarly in cases not involving dominance such as Egan's examples or 'Death in Damascus'.²¹ So although it is (in my own view) always true that the statistical facts are enough by themselves for practical purposes, it is only in the quantum cases here discussed that *they* are clear but the underlying causal structure is completely open. That is why they are well suited to reveal CDT's implausible sensitivity to variations in one's background theorizing about the operation of the device.

6. Objections

A defender of CDT might object (i) that CDT *does not* make the recommendations that I have claimed, given hypotheses (A2), (A3) or (B2); (ii) that it is unclear whether it does, because it is unclear what are supposed to count as causal connections here; (iii) that the cases being non-constructive pose no definite objection to CDT.

(i) *Counterfactual indefiniteness*. The (B2)-type cases in Tables 4-7 require that for CDT to give the verdicts that I am attributing to it e.g. at (69), there must be a definite credence in counterfactuals such as (12hom \rightarrow yy); for expressions denoting such quantities appear e.g. at (41). But—the objector says—(B2) is itself incompatible with this: according to it, there *is* no prior state of the particles that could make any such counterfactual true in the first place, and so no state of affairs the agent's confidence in which Cr (12hom \rightarrow yy) is measuring. So I cannot argue that CDT makes these recommendations after all.

²⁰ For an example of this explicit stipulation see Joyce 1999: 149. Of course there are some who deny that the stipulation is coherent on the grounds that my present act can only be symptomatic of its effects (Price 2012: 510). On that view it is hard to see that EDT and CDT ever diverge; and then we have a shortcut to my main conclusion that causal knowledge is unnecessary for practical reasoning.

²¹ Egan 2007; Gibbard and Harper 1978: 372-5.

The objection relies on the assumption that a counterfactual cannot be true unless there is in actuality some categorical fact (like the prior instruction set) to ground it: that there cannot in Dummett's terms be counterfactuals that are *barely* true.²² This is a very deeply rooted assumption. We should feel deep unease at the idea that two equally filled and identically constituted vessels (say, two otherwise indistinguishable bowls of water) should, when struck in the same way, give off different notes. If we came across a case that looked like this, it would be almost irresistible to think that what explains this difference in their propensities *is* some further difference in their *actual* constitution.²³

On the other hand it is not quite irresistible that we should think this in every case; and there are actual as well as possible philosophical positions that allow for counterfactuals that are barely true. An actual such position arises from the Lewisian semantics for counterfactuals, on which the truth-value of a counterfactual concerning an object's behaviour depends only on that object's behaviour at the relevant nearby possible worlds and not necessarily on any intrinsic feature of it. On that view it is entirely possible that two intrinsically identical objects should have different propensities i.e. be disposed to respond differently under the same counterfactual stimulation, and so there is nothing wrong with a distribution of credence that allows this. For instance in the case at hand, Cr ($12_{\text{hom}} \rightarrow yy$) is perfectly well defined as long as there is an appropriately measurable set of worlds in which setting A and B to '1' and '2' gives the reading 'yy' this or that chance of occurring. Nothing in this account demands that there be any categorical feature of the actual that makes the counterfactual true.²⁴

But in any case, even if we do accept the assumption that counterfactuals cannot be barely true, this makes things no better for CDT. If we reject hidden variables view then it now seems that we cannot make any claims about the counterfactual (hence causal) dependence or independence readings of the receivers upon their settings. And this means that far from agreeing with EDT in these cases, CDT actually gives no advice at all. So there is still a divergence between the two theories over these cases, only it is not a difference between a theory that advise (say) betting in Table 4 and one that advises not betting, but rather between a theory that advises betting and a theory that gives no advice. And this is just as damaging for CDT: what we have constructed (at least on assumption (B2)) is a family of cases in which which practical action is called for but whereof CDT is just silent.

That silence extends even to the simplest cases: if nothing makes the counterfactual $12 \rightarrow yy$ true then nothing makes $1_A \rightarrow y_A$ true either. But then CDT gives no advice even in the almost trivial situation where one must

²² 1976: 53.

²³ The example is from Evans (1980: 276-7).

²⁴ One possible such situation would be an atheistic version of Berkeleian phenomenalism. We usually think that what makes it true, that if I were in my office now then I'd see a desk in my office, is that there now is a desk in my office. But for Berkeley it is the other way around: it is counterfactuals about what I or somebody else would observe that make true the apparent categorical statements about 'physical' objects (1985: 90 (*Principles* s3)). For Berkeley himself the counterfactuals are themselves made true by God's categorical will; but for the atheist phenomenalist it would have to be barely true. It is for that phenomenalist simply a brute fact, not obtaining in virtue of anything that is already there or anyone's presently watching it, that if I were now in my office I should see my desk (Berlin 1999: 43ff.).

choose between switching receiver A to setting 1, thereby betting \$1 on y_A , and not doing so, as in the following table:

	y_A	n_A
1_A	1	-1
-1_A	0	0

Table 8: (B)-type quantum case 4

If CDT had anything to say about this case it would be that it's worth taking the bet if and only if your $\text{Cr}(1_A \rightarrow y_A) > 0.5$; but since, on the current proposal, the expression on the left hand side is meaningless, CDT has no advice to give about even this simplest of quantum decision problems. On the other hand, EDT gives here advice, and it is commonsensical: you should take such bets if and only if you'd expect to win them more often than to lose them.

Perhaps the causalist could reply that EDT gives correct advice in these quantum cases and others where the relevant causal or counterfactual statements do not make sense; but that in more everyday cases (which we *can* describe in terms of causality) we should follow the advice of CDT. But what could possibly motivate this eclecticism? Why wouldn't it be equally sensible, by causalist lights, to follow *maximin*, or *maximin regret*, or any other decision rule you please, in those cases where CDT is silent? If EDT is giving proper advice in quantum cases then that must be because the statistical facts (1) and (2) are decisive there. But if statistical facts alone are decisive in these cases then why are they not also decisive in other cases of divergence from EDT?

Specifically: consider a rival theory that advises you to follow CDT in cases where it makes sense to speak of causal dependence or independence etc. of states of the world upon your acts, but to follow Fictionalist CDT (FCDT) in the quantum cases, where FCDT asks us to pretend to accept the causal descriptions of these situations that would explain the regularities that we observe if only they made sense and were true; e.g. those that ultimately motivate (62)-(67) in connection with Table 6. FCDT then gives exactly the same results as those claimed for CDT in the quantum case. Now the eclectic view has no answer to the question: if we should prefer EDT to FCDT where they clash in quantum cases, then why should we not equally prefer EDT to CDT in classical i.e. non-quantum cases where *they* clash?

(ii) The varieties of causation. The second line of objection is that by presenting the five conditions (A1)-(A3), (B1) and (B2) as genuine alternatives I am ignoring the different notions of causation that might be of interest to physicists studying these phenomena: when we try to be more specific, we may find that one or more of these positions drops out. For instance, if we think that causality must involve the transfer of information then the wings of the experiment must be causally isolated because of the prohibition on superluminal signaling; so on this view we must rule out (B1). If we think of causality as involving correlations that no prior state screens off, then the receivers are causally related on any no-hidden-variables theory; on this view (B1) may be true but (B2) has to go.

It's true that I haven't in this version said anything about the different things that we might of causation as being. But that is only because my

purposes do not demand it. The idea behind the approach was supposed to be that there are *some* feasible theoretical assumptions in which EDT and CDT diverge, not that this *every* theoretical approach mandates that view of things.²⁵ To establish this it isn't necessary to defend any particular analysis of the causal relation but only to show that on some views of it there is no action at a spacelike distance.

Of course there is more to be said. The interesting question is really: which of these notions if any is the one that the causalist had in mind all along? What is it about the causal relation that makes it the one that rational decision-making should especially respect? But I don't think that there was any answer to this question. What was intuitively appealing about the causalist's appeal to modality was not any specific feature of the counterfactual or causal relation that some explications of this notion preserve but which others do not. It is rather the intuitive idea of *bringing about* that is supposed to be doing this work. And let us not enquire too closely, or at all, into the idea what it is about 'bringing about' that somehow works a magic that mere statistics can never achieve.

(iii) *Does it matter that the argument is non-constructive?* The discussion of the A-type interpretations in s3 was non-constructive in the sense that although it identifies a particular decision situation (Table 1) over which EDT and CDT are bound to disagree, it does not identify *which* option, of the ones that EDT rules out, is the one that CDT endorses. The discussion of the B-type interpretation (B2) in s4 was non-constructive in the further sense of not even identifying a specific problem over which EDT and CDT give conflicting advice: we know that they disagree *either* over some D (i, z) as described at Table 4 *or* over the B-type quantum case 2 at Table 5, but nothing in the argument tells us which. It was also non-constructive in the same sense as my discussion of the A-type interpretations i.e. even within table 5 itself there is nothing to tell us which of the 'hom' options that EDT rules out gets endorsed by CDT.

But this doesn't matter for the purposes of the two arguments against CDT that section 5 built upon these cases. All that those arguments required was that some such cases exist, also in the case of the first argument: that in those cases the statistical facts (1) and (2) favour EDT over CDT; and in the case of the second argument: that CDT will in such cases give differing advice depending on one's credence in metaphysical questions that remain undetermined by our actual, and perhaps all possible, observations of the Stern-Gerlach device. Constructive argumentation is not necessary for *these* purposes.

But in any case, it would certainly be feasible in principle to construct a locus of disagreement between the two decision theories, if once given an agent who accepts (say) interpretation (B2), on the supposition that the agent also takes the same attitude towards the relevant counterfactuals on different

²⁵ That stronger demand would certainly rule out at least some of the cases that are of interest to decision theory. E.g. the standard Newcomb Problem (Nozick 1970: 207-8) only generates divergence between EDT and CDT if we are willing to go along with the stipulation that the case involves no backwards causation, even though the phenomena of the problem are compatible with that interpretation if anything is. So there is nothing new about the idea of presenting an example against a background of specific theoretical assumptions.

runs of the device.²⁶ Then on the first three runs of the device offer him three successive decision problems of the form of Table 7. Problem 1 is just as in Table 7. Problem 2 is like problem 1 except that it permutes ‘12hom’ with ‘13hom’ and ‘12het’ with ‘13het’. And problem 3 is like problem 1 except that it permutes ‘12hom’ with ‘23hom’ and ‘12het’ with ‘23het’. If he takes the ‘hom’ option in one of these cases (or would do so for an arbitrarily small incentive) then we have found a disagreement with EDT. If he does *not* disagree with EDT on any of these cases, then let him face a sequence of decision problems (problem 3, problem 4...) where the *i*th problem is D (i^* , $1 - 2^{-i}$) as in Table 4, where $i^* = (1 + i \bmod 3)$. Then the argument of section 4 has been that we will eventually reach a problem D (i^* , $1 - 2^{-i}$), *i* being finite, in which the agent chooses to quit rather than to bet i^* hom, in contradiction to EDT. So if it matters (though it may not), we can, for any agent that follows CDT, at least in principle construct a quantum case in which his own choice violates EDT’s preferences over some specific and identifiable set of options.

(iv) *Mixed theoretical beliefs.* I have so far proceeded entirely on the assumption that an agent lends all of her credence to some one of the five theoretical options that I identified at section 2. Of course that is unrealistic: what is more likely is that a well-informed agent spreads her credence across various causal hypotheses concerning the working of the device, just as in any everyday case she spreads credence across various hypotheses concerning the effects of the actions that are available to her in that case. The question is whether this makes a difference to the overall decision-theoretic recommendations. Are causalists and evidentialists of this more realistic and ambivalent type bound to disagree over the quantum cases that I’ve been considering?

Yes there are, and in fact the examples that we have already considered will do perfectly well. First and in order to simplify matters, let us define C (for ‘causality’) to abbreviate those hypotheses (A1) and (B1) on which there is a causal influence from the setting of one receiver to the display on the other, either because (A1) the settings exert a retrocausal effect via the initial state of the particles, or (B1) because a direct causal influence somehow spans the spacelike interval between the setting on one wing and the reading on the other. So $\neg C$ abbreviates all of those other hypotheses (A2), (A3) and (B2) that deny any such form of influence:

$$(70) \quad C \equiv_{\text{def.}} A1 \vee B1$$

$$(71) \quad \neg C \equiv_{\text{def.}} A2 \vee A3 \vee B2$$

Next, consider some decision problem D (*i*, *z*) as at Table 4. For any such problem, the V-score of betting *i*hom is simply $(1 - z)$, and that of Q is simply zero. And this is true under any hypothesis about the causal structure of the device, since EDT makes recommendations that are independent of any metaphysical hypotheses about causation and instead depend only on the observed statistics—at least this is so if, as I here assume, the agent’s

²⁶ This assumption is not logically unquestionable but it is not really contentious either: if it were not the case that most people’s credences are relatively stable across time in the absence of new information, it would be very hard to know anyone’s beliefs about anything in the intervals between explicit avowals.

subjective conditional credences are themselves based upon these). So EDT will recommend *iihom* in Table 4 to any agent meeting that condition, including any agent whose credence is divided amongst the (A) and (B)-type hypotheses that I have outlined.

We cannot directly calculate what recommendation CDT makes to such an agent. However it is true even of *such* an agent that CDT *will* recommend quitting in D (i, z) for some $z < 1$ *unless* equation (43) holds. Recall that the only premises in the argument for (43) were (40) and (41), neither of which depended on the agent's specific credences in this or that particular causal hypothesis. Putting together this point with the insensitivity of EDT to these credences, we can see that the argument against CDT will hold even on the assumption of divided credence *unless* (43) holds. So we may take forward (43) from the foregoing argument.

Next, consider the following decision problem:

	C yy	C yn	C ny	C nn	¬C yy	¬C yn	¬C ny	¬C nn
12hom	α	0	0	α	α	0	0	α
13hom	α	0	0	α	α	0	0	α
23hom	α	0	0	α	α	0	0	α
12het	0	1	1	0	0	1	1	0
13het	0	1	1	0	0	1	1	0
23het	0	1	1	0	0	1	1	0

Table 9: Mixed Quantum Case α

In this problem there are two different types of states of nature: those in which the causal hypothesis C holds (C and $\neg C$ being as defined at (70)-(71)) and those in which the causal hypothesis fails. However the payoffs are completely fixed and verifiable, these depending *only* upon one's initial setting of the receivers and their readings. For instance, if one takes option 13hom and the both receivers give reading 'y' then one gets a payoff of α , whichever of the hypotheses C and $\neg C$ is true. (I am taking α to be some real number, $\alpha > 1$.) Which one of C and $\neg C$ is true is not in fact something on which the agent has any strong view, her credence being ex hypothesi divided between them.

What EDT recommends to this 'mixed' agent depends in the following manner on the precise value of α :

$$(72) \quad V(12\text{hom}) = \alpha \text{Cr}(yy \vee nn \mid 12\text{hom}) = 0.25\alpha$$

$$(73) \quad V(12\text{het}) = \text{Cr}(yn \vee ny \mid 12\text{hom}) = 0.75$$

—and similarly for the other 'hom' and 'het' options. So EDT recommends any 'het' option over every 'hom' option if and only if $\alpha < 3$; and in fact this recommendation is quite independent of the precise value of the agent's Cr(C).

What about CDT? Here things are only slightly more complicated. Comparing 12hom and 12het, the general expressions for the relevant utilities take the following forms:

$$(74) \quad U(12\text{hom}) = \alpha (\text{Cr}(12\text{hom} \rightarrow (C \wedge (yy \vee nn))) + \text{Cr}(12\text{hom} \rightarrow (\neg C \wedge (yy \vee nn)))$$

$$(75) \quad U(12\text{het}) = \text{Cr}(12\text{het} \rightarrow (C \wedge (yn \vee ny))) + \text{Cr}(12\text{het} \rightarrow \neg C \wedge (yn \vee ny))$$

To evaluate these, note first that since neither the choice of setting nor the choice of bet has any effect on which causal hypothesis is true, and in particular no effect upon which of C and $\neg C$ is true, the following identities must be true for any state $S \in \{yy \vee nn, yn \vee ny\}$ (here writing '12' indifferently for '12hom' and '12het':

$$(76) \quad \text{Cr}(12 \rightarrow (C \wedge S)) = \text{Cr}(C \wedge (12 \rightarrow S))$$

$$(77) \quad \text{Cr}(12 \rightarrow (\neg C \wedge S)) = \text{Cr}(\neg C \wedge (12 \rightarrow S))$$

Now the right hand sides of (76) and (77) resolve into:

$$(78) \quad \text{Cr}(C \wedge (12 \rightarrow S)) = \text{Cr}(12 \rightarrow S | C) \text{Cr}(C)$$

$$(79) \quad \text{Cr}(\neg C \wedge (12 \rightarrow S)) = \text{Cr}(12 \rightarrow S | \neg C) \text{Cr}(\neg C)$$

It is straightforward to calculate the conditional probabilities on the right of (78) and (79) for the two possible values of S . In particular, if the causal hypothesis is *true* then we should expect the settings to have a superluminal causal effect upon the readings that mirrors the statistics (1) and (2). So we have:

$$(80) \quad \text{Cr}(12 \rightarrow yy \vee nn | C) = 0.25$$

$$(81) \quad \text{Cr}(12 \rightarrow yn \vee ny | C) = 0.75$$

But if the causal hypothesis is *false* then no setting has any causal impact on the reading on the opposite wings; then by the argument at s4 we have:

$$(82) \quad \text{Cr}(12 \rightarrow (yy \vee nn) | \neg C) = \text{Cr}(\mathbf{111} \vee \mathbf{110} \vee \mathbf{001} \vee \mathbf{000} | \neg C)$$

—where $\mathbf{111}$, $\mathbf{110}$ etc. are as defined at (61).²⁷ Writing c for $\text{Cr}(C)$ and Cr_C , $\text{Cr}_{\neg C}$ for the marginal distributions $\text{Cr}(x | C)$ and $\text{Cr}(x | \neg C)$ respectively, we may now substitute into (74) and (75) to get:

$$(83) \quad U(12\text{hom}) = 0.25\alpha c + \alpha(1-c) \text{Cr}_C(\mathbf{111} \vee \mathbf{110} \vee \mathbf{001} \vee \mathbf{000})$$

$$(84) \quad U(12\text{het}) = 0.75c + (1-c) \text{Cr}_{\neg C}(\mathbf{101} \vee \mathbf{100} \vee \mathbf{011} \vee \mathbf{010})$$

By the same reasoning on the other four options we have:

$$(85) \quad U(13\text{hom}) = 0.25\alpha c + \alpha(1-c) \text{Cr}_C(\mathbf{111} \vee \mathbf{101} \vee \mathbf{010} \vee \mathbf{000})$$

$$(86) \quad U(13\text{het}) = 0.75c + (1-c) \text{Cr}_{\neg C}(\mathbf{110} \vee \mathbf{100} \vee \mathbf{011} \vee \mathbf{001})$$

$$(87) \quad U(23\text{hom}) = 0.25\alpha c + \alpha(1-c) \text{Cr}_C(\mathbf{111} \vee \mathbf{011} \vee \mathbf{100} \vee \mathbf{000})$$

²⁷ Note that on this definition (81) holds good on hypotheses (A2) and (A3) as well as on hypothesis (B2) because on the former hypotheses $\mathbf{111}$, $\mathbf{110}$ etc. are respectively equivalent to YYY , YYN etc.

$$(88) \quad U(23\text{het}) = 0.75c + (1-c) \text{Cr}_{\cdot C}(\mathbf{110} \vee \mathbf{010} \vee \mathbf{101} \vee \mathbf{001})$$

Now we know by the structurally identical reasoning of (21)-(34)—which applies just as well here because $\text{Cr}_{\cdot C}$ is a probability distribution—that (say) twice the marginal credence on the right of (83) equals or exceeds the corresponding quantity on the right of (84):

$$(89) \quad 2 \text{Cr}_{\cdot C}(\mathbf{111} \vee \mathbf{110} \vee \mathbf{001} \vee \mathbf{000}) \geq \text{Cr}_{\cdot C}(\mathbf{101} \vee \mathbf{100} \vee \mathbf{011} \vee \mathbf{010})$$

--at any rate, either this inequality holds or some corresponding one holds for the marginal credences in (85) and (86), or for those in (87) and (88). Suppose without loss of generality that (89) is true. If we now write $t =_{\text{def.}} \alpha - 2$, $p =_{\text{def.}} \text{Cr}_{\cdot C}(\mathbf{111} \vee \mathbf{110} \vee \mathbf{001} \vee \mathbf{000})$ and $q =_{\text{def.}} \text{Cr}_{\cdot C}(\mathbf{101} \vee \mathbf{100} \vee \mathbf{011} \vee \mathbf{010})$ then subtracting (84) from (83) gives:

$$(90) \quad U(12\text{hom}) - U(12\text{het}) = 0.25c(t-1) + p(2+t)(1-c) - q(1-c)$$

Since (89) tells us that $2p - q \geq 0$, it follows that:

$$(91) \quad U(12\text{hom}) - U(12\text{het}) > 0 \text{ if } 0.25c(t-1) + pt(1-c) > 0; \text{ hence:}$$

$$(92) \quad U(12\text{hom}) - U(12\text{het}) > 0 \text{ if } t > c / (c + 4p(1-c))$$

Elementary calculations tell us that if $c < 1$ and $p > 0$ then there is *always* some t strictly between 0 and 1 that satisfies the right hand side of (92).

But since $\alpha = t + 2$, this means that if $c < 1$ and $p > 0$ then we can *always* choose some payoff to the hom options α , strictly between 2 and 3, on which the causalist will prefer 12hom to 12het (or more generally, some 'hom' option to the corresponding 'het' option). But $p > 0$ is an innocuous assumption. And by (72) and (73), we know that $\alpha < 3$ guarantees that EDT always prefers any 'het' option to every 'hom' option. So if $c > 0$ i.e. if the agent gives *any credence at all* to the non-causal hypotheses (A2), (A3), or (B2)—surely a reasonable assumption—then EDT and CDT will diverge over Mixed Quantum Case α for some α .

So the objection fails: as long as the agent is not absolutely certain of superluminal or retroactive causation, it is possible to construct a quantum case in which EDT and CDT give divergent advice. And any such case will equally support both of the arguments against that section 5 based upon 'pure' quantum cases like those in Table 1, Table 4 and Table 5.

References

- Bell, J. S. 1977. Free variables and local causality. *Epistemological Letters*, Feb. 1977. Reprinted in his *Speakable and Unspeakable in Quantum Mechanics*. 2d ed. Cambridge: CUP: 100-4.
- . 1990. La nouvelle cuisine. In Sarlemijn, A and P. Kroes, eds, *Between Science and Technology*. Elsevier: 97-115.
- Berkeley G. 1985. *Philosophical Works*. Ed. M. Ayers. London: Everyman.
- Berkovitz, J. 1995. *Quantum Nonlocality: An Analysis of the Implications of Bell's Theorem and Quantum Correlations for Nonlocality*. Ph.D. thesis,

- University of Cambridge.
- Berlin, I. 1999. Empirical propositions and hypothetical statements. In his *Concepts and Categories: Philosophical Essays* (H. Hardy, ed.). London: Pimlico: 32-55.
- Cavalcanti, E. Causation, decision theory and Bell's theorem: a quantum analogue of the Newcomb problem. *BJPS* 61: 569-97.
- Clauser, J. F., M. A. Horne, A. Shimony and R. A. Holt. 1969. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* 23: 880-4.
- Dummett, M. A. E. 1976. What is a theory of meaning? (II) In Evans, G. and J. McDowell, ed., *Truth and Meaning*. Oxford: OUP. Reprinted in Dummett, M. A. E. 1993. *The Seas of Language*. Oxford: OUP: 34-93.
- Eberhard, P. H. 1978. Bell's theorem and the different concepts of nonlocality. *Nuovo Cimento* 46B: 392-419.
- Eells, E. 1982. *Rational Decision and Causality*. Cambridge: CUP.
- Egan, A. 2007. Some counterexamples to causal decision theory. *Phil. Rev.* 116: 93-114.
- Evans, G. 1980. Things without the mind. In van Straaten, ed. *Philosophical Subjects: Essays Presented to P. F. Strawson*. Oxford: Clarendon Press. Reprinted in Evans, G. 1985. *Collected Papers*. Oxford: OUP: 249-90.
- Gibbard, A. and W. Harper. 1978. Counterfactuals and two kinds of expected utility. In Hooker, C., J. Leach and E. McClennen, eds. *Foundations and Applications of Decision Theory*. Dordrecht: Riedel: 125-62. Reprinted in Gärdenfors, P. and N.-E. Sahlin, eds. *Decision, Probability and Utility* (1988). Cambridge: Cambridge University Press.
- Hume, D. 1949 [1739]. *Treatise of Human Nature*. Ed. with an analytical index by L. A. Selby-Bigge. Oxford: Clarendon Press.
- Laudisa, F. 2001. Non-locality and theories of causation. In Butterfield, J. and T. Placek (eds), *Non-Locality and Modality*. Dordrecht: Kluwer.
- Lewis, D. 1973. *Counterfactuals*. Oxford: Blackwell.
- Maudlin, T. 2002. *Quantum Non-Locality and Relativity*. Oxford: Blackwell.
- Mermin, N. 1981. Quantum mysteries for anyone. *J. Phil.* 78: 397-408.
- Muckenheim, W. 1982. A resolution of the EPR paradox. *Lett. al Nuovo Cimento* 35: 300-4.
- Nozick, R. 1970. Newcomb's problem and two principles of choice. In Rescher, N., ed. *Essays in honor of Carl G. Hempel*. Dordrecht: D. Reidel: 114-46. Reprinted in Moser, P. (ed.), *Rationality in Action: Contemporary Approaches* (1990). Cambridge: CUP.
- Price, H. 1991. Agency and probabilistic causality. *BJPS* 42: 157-176.
- . 1996. A neglected route to realism about quantum mechanics. *Mind* 103: 303-36.
- . 2012. Causation, chance and the rational significance of supernatural evidence. *Phil. Rev.* 121: 483-538.
- Reichenbach, H. 1984. *The Direction of Time*. New York: Dover.
- Van Fraassen, B. C. 1991. *Quantum Mechanics*. Oxford: OUP.

