

Causal Decision Theory and EPR correlations

Abstract. The paper argues that on three out of eight possible hypotheses about the EPR 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 the EPR statistics conflict. We infer that anyone who fully accepts any of these three hypotheses has strong reasons to reject causal Decision Theory. Finally, we extend the original construction to show that anyone who gives any of the three hypotheses *any* non-zero credence has strong reasons to reject Causal Decision Theory. However, we concede that no version of the Many Worlds Interpretation (Vaidman 2014) gives rise to the conflicts that we point out.

The paper argues (in ss. 1-7) that on three out of eight possible hypotheses about the EPR 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 the statistics for the EPR experiment conflict. We infer (in s. 8) that anyone who accepts any of these three hypotheses has strong reasons to reject causal Decision Theory. Finally, we argue (in s. 9) that by a simple extension of the original construction, anyone who gives any of these hypotheses *any* non-zero credence has strong reasons to reject Causal Decision Theory.

Before turning to our main case let us briefly review the two decision theories. Evidential Decision Theory (EDT) recommends doing what is most *auspicious*. More formally, it recommends whichever 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 an arbitrary proposition P, then for any (proposition describing an available) option O, Cr and V jointly satisfy:

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

— \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:

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

$Cr(O \rightarrow S)$ measures your confidence in the *counterfactual* proposition 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* or other correlated non-effects 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 $Cr(S \wedge O \wedge Y)$

¹ Jeffrey 2004: 99.

> 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 cannot be causally influenced by 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 $\{S_1 = I \text{ get wet}, S_2 = I \text{ don't get wet}\}$ meets condition (i) but not condition (ii). But the partition $\{S_1 = It \text{ rains}, S_2 = It \text{ doesn't rain}\}$ meets both conditions. In that case the expression for an option's U-score takes this particularly simple form:

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

It may seem odd that CDT is still applicable to a decision situation in which the state of nature is not causally influenced by what you do. But this is in fact a *consequence* of its concern with causality. Causally independent states of nature are those to which the causal contribution of my action is null: so it is a distinctive consequence of CDT, which all of its adherents accept, that in these cases one's credences in the states of nature make the same contribution to the evaluation of every act, this being reflected in (3). In fact in these cases and as (3) suggests, the causal theory is equivalent to Savage's original treatment.³

1. Bell's Theorem

In order to appreciate the point it isn't necessary to grasp any of the mathematical details of EPR beyond the essentially statistical reasoning that creates the problem. For this purpose the following completely non-technical exposition, which follows Mermin (1981), is perfectly adequate.

It is technically feasible to produce a device with the following features. It has 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 has 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.

After setting the switches we turn on the source. 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'.

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.

³ Savage 1972 ch. 2-5; see further Joyce 2002: 71-5. Thanks to a referee for raising this point.

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. Very many runs reveal the following statistical facts:

- (4) *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'.
- (5) 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, which is not a mere fiction but has actually been realized, for instance as follows:

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 $\theta = 0$ [as in (4)] and 1/4 when $\theta = \pm 120^\circ$ [as in (5)].⁴

So far we'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 it.

Fact (4) 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 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

⁴ Mermin 1981: 407-8

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’. To connect with the philosophy of physics literature, we add that the hypothesis of a prior instruction set for the particles is just the hypothesis that there are *local hidden variables* causing – perhaps probabilistically – the results at the receivers.

Unfortunately the statistical fact (5) looks 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 ≤ i ≤ 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 S₁, S₂, S₃ and any j, k, we have:

$$(6) \text{Fr} (S_1 S_2 S_3 | j;k) = \text{Fr} (S_1 S_2 S_3)$$

Now fact (5) 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:

$$(7) \text{Fr} (YYY | 1;2) + \text{Fr} (YYN | 1;2) + \text{Fr} (NNY | 1;2) + \text{Fr} (NNN | 1;2) = 0.25$$

$$(8) \text{Fr} (YYY | 1;3) + \text{Fr} (YNY | 1;3) + \text{Fr} (NYN | 1;3) + \text{Fr} (NNN | 1;3) = 0.25$$

$$(9) \text{Fr} (YYY | 2;3) + \text{Fr} (NYY | 2;3) + \text{Fr} (YNN | 2;3) + \text{Fr} (NNN | 2;3) = 0.25$$

From (6) we can simplify these to:

$$(10) \text{Fr} (YYY) + \text{Fr} (YYN) + \text{Fr} (NNY) + \text{Fr} (NNN) = 0.25$$

$$(11) \text{Fr} (YYY) + \text{Fr} (YNY) + \text{Fr} (NYN) + \text{Fr} (NNN) = 0.25$$

$$(12) \text{Fr} (YYY) + \text{Fr} (NYY) + \text{Fr} (YNN) + \text{Fr} (NNN) = 0.25$$

Adding these together we get:

$$(13) 2\text{Fr} (YYY) + 2\text{Fr} (NNN) + \sum_{S_1, S_2, S_3} \text{Fr} (S_1 S_2 S_3) = 0.75$$

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

$$(14) \sum_{S_1, S_2, S_3} \text{Fr} (S_1 S_2 S_3) = 1$$

$$(15) 2\text{Fr} (YYY) + 2\text{Fr} (NNN) \geq 0$$

But (13), (14) and (15) are jointly inconsistent. Since they follow from (5) and (6), and since (5) has been observationally verified as convincingly as you like, it seems that we are left with two options: either reject (6) and keep the hypothesis of a prior instruction set, i.e. retain local hidden variables; or reject

the hypothesis of an instruction set. This is the lesson of Bell's Theorem. In fact each option just mentioned itself involves further sub-options, as follows.

2. *Eight responses*

(A) We can insist that there is a prior common state of the particles (a prior instruction set) but deny (6), i.e. maintain that that state is correlated with one's possibly randomized choice of receiver setting.

In the quantum mechanics literature, assumption (6) is known by several names, each of which reflects a particular interpretative gloss. For example, Hofer-Szabó (2012) calls it the *no conspiracy* assumption, e.g. Bell (1977) calls it the *free will* assumption, and e.g. Evans, Price & Wharton (2013) call it the *no retrocausality* assumption. We will not enter into the justification of these names, but it will be helpful to us to group the alternative glosses into two sub-options.

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 $A = 1$ and $B = 2$ has the effect of inhibiting (though without altogether excluding) the *prior* instruction sets YYY , YYN , NNY and NNN .⁵

Sub-option (A2) covers two alternatives to (A1). The first is to claim 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 incompatible with free will.⁶ Whatever you think about that, it certainly implicates the universe in a kind of conspiracy that nowadays is hard to credit.

(A2) also includes the alternative view 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.⁷ This interpretation of events is perhaps less unpalatable than the other option on this branch; but as we'll see, it looks worse than at least one option on another, more popular branch.

A myth persists in the literature to the effect that Bell's Theorem rules out hidden variable completions of quantum mechanics. The matter is somewhat delicate, depending on what one means by 'local', but the options under (A) certainly are hidden variable responses to Bell's Theorem that rule out superluminal *signaling*, either between the experimenter and distant receiver, or between the receivers themselves.⁸

The alternative to (A) is to maintain that the prior instruction set—if there is one—is *not* correlated with your decision to put the receivers in any

⁵ Price 1996.

⁶ Bell 1977: 100.

⁷ Reichenbach 1984: 157.

⁸ See Lewis 2006 for a survey of these responses.

particular settings. In that case, we must accept Bell's finding that there *is* no prior instruction set. But now we are left with trying to explain fact (4): the fact that the two particles, when tested by receivers that are at the same setting, *always* give the same reading.

The taxonomy of ways that one may proceed here is determined by the other assumptions required to derive a Bell inequality—one of which, if we retain the no-conspiracy assumption, must therefore be rejected. Recall from Shimony (1990) that, apart from the no-conspiracy assumption (associated with (A), above), one must make two more assumptions: *outcome independence* and *parameter independence*.

(B) Outcome independence is the assumption that space-like separated measurement outcomes are probabilistically independent, when conditioned on the measurement settings for both devices and the value of the hidden variables (if any). This may be denied in two ways.

(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, action at a distance between the particles. Labeling 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 influences 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.⁹

Also under (B1) we put responses which deny outcome dependence by postulating a (non-local) causal connection between measurement *outcomes*—such responses may reject hidden variables altogether. An example of such a theory is the Ghirardi-Rimini-Weber theory of spontaneous dynamical collapse (see Ghirardi *et al.* 1986).

(B2) One might alternatively take the line that there is a *non-causal* correlation between the hidden variables associated with 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

⁹ But not however to superluminal *signaling*: as is widely recognized (e.g. Redhead 1987: 113-5), 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 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)

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 well prefer it to the second version of (A2), which also postulates unexplained correlations but which appeals in addition to a pre-existing instruction set.

(C) Parameter independence is the assumption that measurement outcomes are probabilistically independent of the choice of apparatus settings for space-like separated measurement devices, when conditioned on the hidden variables (if any). Theories that reject parameter independence *must* be hidden-variable theories, since bare quantum mechanics (in which nothing supplants or is added to the wave-function) obeys parameter independence. One example of such a theory is the de Broglie-Bohm ‘pilot wave’ theory, in which particles have determinate trajectories but are guided by the wave-function, which is construed as a physical entity.¹⁰

Parameter *dependence* allows for the possibility, given a particular specification of values for the hidden variables, of superluminal signaling, since the non-local correlations involve parameters (measurement settings) over which experimenters may have direct control. However, the *practical* possibility of such superluminal signaling for the experimenters is ruled out so long as the experimenters are *unaware* of the hidden variables’ true values, and the relative frequencies for the various values are as prescribed, via the Born rule, by the wave-function.¹¹ Later we will consider cases for the de Broglie-Bohm theory in which these frequencies are not given by the Born rule.

In any case, parameter dependence entails case-by-case superluminal signaling between measurement devices, whether or not it is observable, or manipulable, by experimenters. This has strange results. For, if two events, (a) the choice of a measurement setting at A, and (b) a measurement at B are space-like separated, then there will be frames of reference in which (a) occurs before (b) *and* other frames—equally good ones from the point of view of special relativity—in which (b) occurs before (a). Faced with a correlation between (a) and (b), as parameter dependence demands, there are two alternatives to consider.

(C1) One might wish to claim that the *human choice* of measurement setting (a) *always* causes (b) (perhaps for reasons similar to Bell’s in his acceptance of the no-conspiracy assumption). In this case, as in (A1), it seems that one must believe in an agent’s action to bring about an effect *retrocausally*; for, given *any* choice of inertial frame, one could construct an experiment in which the (b)-type event precedes the (a)-type event in that frame.

¹⁰ Of course, in the de Broglie-Bohm theory, this parameter dependence is made experimentally inaccessible by the epistemic uncertainty in the real locations of the corpuscles. This relies on the probabilities for the corpuscles’ locations taking their *equilibrium* values, as given by the Born rule (see Holland 1993, Ch. 11). However, there is the theoretical possibility of non-equilibrium states; in these, an experimenter can in principle take advantage of parameter dependence to signal non-locally (see Valentini 2002).

¹¹ Holland 1993: 471-6.

(C2) One might instead wish to allow that, at least sometimes—most plausibly in those cases in which the (b)-type event precedes the corresponding (a)-type event, in some hypothesized privileged inertial frame—the (b)-type event causes the (a)-type event, i.e. the human choice of measurement setting. In these cases we may treat the various outcomes at B as states of nature, as described above.

Also to be included in (C2) is any view that denies causation for at least some instances of the EPR experiment. Like (B2), the statistical correlation is treated as brute, i.e. non-causal.

(D) We have not yet discussed Everettian or ‘many worlds’ approaches to the EPR experiment. These are worth separating from (A)-(C), since the interpretation of probability for these theories, e.g. in the statement of the above probabilistic independence assumptions, is far from straightforward. It has been argued (Timpson and Brown 2002)—and we agree—that the postulation, in the Everett approach, that *every possible measurement outcome happens* allows the approach to account for the EPR experiment without invoking non-locality of any kind, even that brought about by retrocausality.

Within the Everettian response, we distinguish between (D1) the decision-theoretic project of Deutsch, Wallace and Greaves (Deutsch 1999; Wallace 2002, 2003, 2012; Greaves 2004) and (D2) the so-called ‘many minds’ response (Albert & Loewer 1988). We discuss these further at s. 7 below.

Not all of these eight 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), (B2) and (C2). 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. 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 at a receiver in such and such setting, and the corresponding disposition of the *other* particle.

The argument will then be as follows. First we argue that anyone who takes any of the views (A2), (B2) or (C2) is committed to a disparity between EDT and CDT in certain decision cases that we’ll shortly describe. This will take slightly different arguments for (A)-type, (B)-type and (C)-type theories: that is, first for adherents of (A2), second for adherents of (B2), and finally for adherents of (C2). We will discuss the situation for (D)-type theories separately.

Then, we argue that this disparity raises two problems for *Causal* Decision 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 seems to vary depending on which of various and (as far as we know) empirically indistinguishable theories is true. Thus CDT is 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. purely statistical conception of causality¹².

3. A-type theories

Suppose first that you accept (A2): you think that there *is* a prior instruction set and that its state is not caused by your current decision to switch the receivers to any particular settings, because: (i) you think that both have a common cause; or (ii) you think that it is the prior instruction set that has caused your current choice of measurement settings; or (iii) 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 all have a payoff of \$2 and the 'het' bets have a payoff of \$1.

So there are $2 \times 3 = 6$ options; we'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	NYY	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

¹² Hume 1949: I.iii.

Table 1: (A)-type EPR 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.

One problem that arises at this point is that according to (A2), the correlations that are relevant to one's payoffs are determined by a prior instruction set that is not in your causal control, and so in a sense fundamentally undetermined. And it might now seem unclear how a rational recommendation can be made at all, using any decision theory.

But the assumption that the correlations are non-causal does not in fact make either Evidential or Causal Decision Theory inapplicable. In the case of Evidential Decision Theory, the stochastic data are all that matter: anyone who accepts facts (4) and (5) will, from its perspective, have credences over the outcomes that are sufficiently precise to be action-guiding irrespective of one's confidence, or lack of confidence, in any causal structure underlying them. And as for *Causal* Decision Theory: the assumption that the prior instruction set is not causally sensitive to the settings on the receivers means that we may treat the former as causally *independent* of anything that you do. This in turn makes it possible to apply the representation (3) to the problem in Table 1 relative to a partition \mathbf{S}^* of the event space into possible prior instruction sets.

To see how these points apply in detail, let us now consider which of the six options EDT and CDT recommend. First consider EDT. In this case the matter is simple: if she is sensible then the agent's conditional credences will reflect the relative frequencies as recorded in fact (5). 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:

$$(16) \text{Cr}(S_i = S_j \mid ij\text{hom}) = \text{Cr}(S_i = S_j \mid ij\text{het}) = 0.25$$

$$(17) \text{Cr}(S_i \neq S_j \mid ij\text{hom}) = \text{Cr}(S_i \neq S_j \mid ij\text{het}) = 0.75$$

It follows from (1), (16), (17) 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:

$$(18) V(12\text{hom}) = 2\text{Cr}(YYY \mid 12\text{hom}) + 2\text{Cr}(YYN \mid 12\text{hom}) + \text{Cr}(NNY \mid 12\text{hom}) + 2\text{Cr}(NNN \mid 12\text{hom}) \text{ by Table 1; so:}$$

$$(19) V(12\text{hom}) = 2\text{Cr}(S_1 = S_2 \mid 12\text{hom}) = 0.5 \text{ by (16), (18)}$$

$$(20) V(12\text{het}) = \text{Cr}(YNY \mid 12\text{hom}) + \text{Cr}(YNN \mid 12\text{hom}) + \text{Cr}(NYY \mid 12\text{hom}) + \text{Cr}(NYN \mid 12\text{hom}) \text{ by Table 1; so:}$$

$$(21) V(12\text{het}) = \text{Cr}(S_1 \neq S_2 \mid 12\text{het}) = 0.75 \text{ by (17), (20)}$$

The same reasoning clearly goes for each of the other options. So EDT reckons the value 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' option¹³:

$$(22) \text{For any } i, j, k, l \text{ s.t. } (1, 1) \leq (i, k) < (j, l) \leq (3, 3): ij\text{het} \overset{\text{W}}{\boxplus}_{\text{EDT}} kl\text{hom}$$

Turn now to CDT. Its recommendations won't depend on the *conditional* credences $\text{Cr}(YYY | 12\text{hom})$ etc. but upon one's credences in the *counterfactuals* $\text{Cr}(12\text{hom} \rightarrow YYY)$ etc. But given the theoretical assumption (A2), we know that the prior state is not caused by 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 follows that $\text{Cr}(12\text{hom} \rightarrow YYY) = \text{Cr}(YYY)$ etc. More generally, for any $1 \leq i < j \leq 3$, and $S_1, S_2, S_3 \in \{Y, N\}$:

$$(23) \text{Cr}(ij\text{hom} \rightarrow S_1S_2S_3) = \text{Cr}(ij\text{het} \rightarrow S_1S_2S_3) = \text{Cr}(S_1S_2S_3)$$

So (3) applies, and it follows that the U-scores for the three 'hom' options are as follows:

$$(24) U(12\text{hom}) = 2(\text{Cr}(YYY) + \text{Cr}(YYN) + \text{Cr}(NNY) + \text{Cr}(NNN))$$

$$(25) U(13\text{hom}) = 2(\text{Cr}(YYY) + \text{Cr}(YNY) + \text{Cr}(NYN) + \text{Cr}(NNN))$$

$$(26) U(23\text{hom}) = 2(\text{Cr}(YYY) + \text{Cr}(YNN) + \text{Cr}(NYY) + \text{Cr}(NNN))$$

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

$$(27) U(12\text{het}) = \text{Cr}(YNY) + \text{Cr}(YNN) + \text{Cr}(NYY) + \text{Cr}(NYN)$$

$$(28) U(13\text{het}) = \text{Cr}(YYN) + \text{Cr}(YNN) + \text{Cr}(NYY) + \text{Cr}(NNY)$$

$$(29) U(23\text{het}) = \text{Cr}(YYN) + \text{Cr}(YNY) + \text{Cr}(NYN) + \text{Cr}(NNY)$$

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

$$(30) U(12\text{het}) > U(12\text{hom})$$

$$(31) U(13\text{het}) > U(13\text{hom})$$

$$(32) U(23\text{het}) > U(23\text{hom})$$

Substituting (24)-(29) into (30)-(32) and adding the three resulting inequalities gives:

¹³ In (22) and (35) we are 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 $b \leq d$. Also, if P and Q are options then $P \overset{\text{W}}{\boxplus}_{\text{EDT}} Q$ means that EDT strictly prefers P to Q . Similarly, $P \overset{\text{W}}{\boxplus}_{\text{EDT}} Q$ means that EDT reckons P at least as choiceworthy as Q . Similarly with 'CDT' replacing 'EDT'.

$$(33) 2 (\text{Cr}(\text{YYN}) + \text{Cr}(\text{YNY}) + \text{Cr}(\text{YNN}) + \text{Cr}(\text{NYY}) + \text{Cr}(\text{NYN}) + \text{Cr}(\text{NNY})) > 6 (\text{Cr}(\text{YYY}) + \text{Cr}(\text{NNN})) + 2 (\text{Cr}(\text{YYN}) + \text{Cr}(\text{YNY}) + \text{Cr}(\text{YNN}) + \text{Cr}(\text{NYY}) + \text{Cr}(\text{NYN}) + \text{Cr}(\text{NNY})); \text{ hence}$$

$$(34) 0 > 6 (\text{Cr}(\text{YYY}) + \text{Cr}(\text{NNN}))$$

But (34) is a contradiction since the credences on the right hand side are both *at least* zero¹⁴, and so the supposition that entails (30)-(32) 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 possibility that is consistent with one's having any credences about the prior state at all.

$$(35) \text{For some } i, j \text{ s.t. } 1 \leq i < j \leq 3: ij\text{hom} \not\prec_{\text{CDT}} ij\text{het} \text{ in Table 1}$$

But taken together (22) and (35) imply that there must be some pair of options *ijhom* and *ijhet* over whose relative ranking EDT and CDT disagree. In particular and without loss of generality, let this be the pair *12hom* and *12het*. Then:

$$(36) V(12\text{het}) > V(12\text{hom})$$

$$(37) U(12\text{hom}) \geq U(12\text{het})$$

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 EPR 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 (36) and (37) 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.¹⁵ So in the

¹⁴ We are here assuming that we cannot invoke negative probabilities, as some quantum theorists have suggested (e.g. Muckenheim 1982, Feynman 1987).

¹⁵ For EDT this is clear from (19), (21) and the fact that every other option in Table 2 gets V-score 0. For CDT it follows from the fact that *12hom* and *12het* both weakly dominate all of the other four options.

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 Table 2 to anyone who accepts interpretation (A2) of the EPR phenomena that we described in section 1. This situation, although lacking the *prima facie* realism of ‘medical Newcomb cases’ (currently the best known loci of disagreement between EDT and CDT), 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 a supernatural ‘predictor’, *or* of the medical cases, which involve correlations between physical state and choice that are unknown to medical science.¹⁶ In fact the cases discussed here are, to our knowledge, the *only* currently feasible ways to arrange for practical disagreement between EDT and CDT.

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 agent. In particular, she must believe that there *is* a prior instruction set (a ‘local hidden variable’)¹⁷; 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 we’ll use the example to make. As we’ll see, all that we need for that purpose is 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 to decision theory than is the standard Newcomb case. On the other hand, the oddity of the perspective from which Table 2 forces this divergence inevitably diminishes the interest of the case. We turn therefore to an argument that does not rely on there being hidden variables at all, but on which the correlations between *manifest* states are acausal.

In fact the argument at ss. 4-5 applies to *all* of the responses (A2), (B2) and (C2), relying as it does only on their common consequence that the settings at one receiver have no causal influence on the readings at the other; we shall exploit this fact at s. 9 part (iv). If that is right, it was not strictly necessary for our argument to treat (A2), (B2) and (C2) separately. Nevertheless, we think that doing so is instructive, because it reveals the distinctive difficulties for CDT that attend the three standard ways of avoiding a Bell inequality. (As we have said, Everettian approaches require quite separate treatment.)

¹⁶ For ‘medical’ Newcomb cases see e.g. Price 1991. Nozick 1969 presents the standard version of the Newcomb problem.

¹⁷ Of course, she must also believe that the prior state either causes or shares a common cause, or is acausally correlated, with one’s present choice to set the receivers this or that way. But *this* is not implausible: given that one *already* opts for (A) and so has swallowed hidden variables themselves, baulking at the idea that the experimenter lacks retrocausal powers (that being the only alternative to (A2)) is arguably straining at a gnat.

4. An argument not relying on hidden variables

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 those summarized in 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 before a particular run of the device, to switch receiver A to setting 1 and bet that its *display* will read 'y'. Here you are *not* betting on the prior state of the hidden variables (in particular, you're not betting in this case that $S_1 = Y$). But you *are* making an operationally *bona fide* bet. For given your choice, the observable display on the receivers will always settle your monetary payoff.

For instance, '12het' still represents a feasible option: you are switching receiver A to setting 1 and receiver B to setting 2, and you are betting that they will *display different readings*. 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 these bets 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 the readings given that one takes this or that option. And *these* conditional credences are certainly available whether or not one accepts any prior instruction set. In the simple scenarios that we consider here, they reflect available statistical records correlating options with outcomes, i.e. facts (4) and (5).

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.¹⁸ 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 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

¹⁸ Remember, (B2) is not denying that switching either receiver to this or that setting has any causal influence on *its own* reading (and neither are (A2) or (C2)). 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.

(3) does not apply, and instead we are forced back on the direct calculation of U-scores by means of (2), applied to counterfactual credences themselves.

For instance, in the case below 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	0	1	0
O₃	0	0	0	1

Table 3: illustrative game without h.v.

So by (2) and Table 3, the following expression gives the U-score of, say, O₁:

$$(38)U(O_1) = Cr(O_1 \rightarrow yy) + Cr(O_1 \rightarrow yn)$$

But in order to calculate the credences in (38) we cannot (as we said) partition over prior states of the world; instead we must directly evaluate these credences by means of formulas of this type:

$$(39)Cr(O_1 \rightarrow yy) = \int_{0 \leq x \leq 1} x \cdot Cr(Ch(yy | O_1) = x) dx$$

—in which Cr expresses one’s distribution function over the possible values for the *conditional chance*, just before one’s act, that a particular setting of receiver A gives to some particular combination of readings on receiver A and receiver B.¹⁹ So the right-hand side of (39) is Cr’s expectation of this conditional chance. Here, conditional chance is taken to reflect causal tendencies: the extent to which the conditional chance of Y on X, Ch(Y|X), exceeds the unconditional chance Ch(Y) of Y, reflects the extent to which the occurrence of X *causally promotes* the occurrence of Y.

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 Table 3, which in no way involves the statistical peculiarities of the EPR setup, there is no obvious reason why its advice should diverge from that of EDT.

But all of this changes when we turn to types of problem that exploit facts (4) and (5). 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

¹⁹ For the continuous case, read Cr (Ch (yy | O₁) = x) as F’ (x): the first derivative of the cumulative distribution function F (x) = Cr (Ch (yy | O₁) ≤ x).

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: EPR without h.v.: $D(i, z)$

What do EDT and CDT advise?

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

$$(40) \text{Cr}(yy \vee nn \mid \text{iihom}) = 1$$

$$(41) \text{Cr}(yn \mid \text{iihom}) = \text{Cr}(ny \mid \text{iihom}) = 0$$

It follows from (1), (40) and (41) that for any i, z the V-scores of the options in $D(i, z)$ are:

$$(42) V(\text{iihom}) = 1 - z$$

$$(43) V(Q) = 0$$

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:

$$(44) U(\text{iihom}) = (1-z)(\text{Cr}(\text{iihom} \rightarrow yy) + \text{Cr}(\text{iihom} \rightarrow nn)) - z(\text{Cr}(\text{iihom} \rightarrow yn) + \text{Cr}(\text{iihom} \rightarrow ny))$$

$$(45) U(Q) = 0$$

Now consider the quantity $(\text{Cr}(\text{iihom} \rightarrow yn) + \text{Cr}(\text{iihom} \rightarrow ny))$ on the right-hand side of (44). It's easy to see that *if* this quantity exceeds 0 then there is some strictly positive $z^* < 1$ such that $U(\text{iihom}) < U(Q)$ in $D(i, z)$ for any $z \geq z^*$.²⁰ In other words, we have a continuum of decision situations in which CDT and EDT diverge i.e. for any $z \geq z^*$, in $D(i, z)$ CDT will endorse quitting but EDT will endorse betting.

²⁰ Since $\text{Cr}(O_1 \rightarrow yy) + \text{Cr}(O_1 \rightarrow nn) + \text{Cr}(O_1 \rightarrow yn) + \text{Cr}(O_1 \rightarrow ny) = 1$, 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 $U(\text{iihom}) < 0 = U(Q)$.

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 fact (4) 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 repeatedly offered these bets at a rate \$z for $z > z^*$, 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 in a way 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 $(Cr(iihom \rightarrow yn) + Cr(iihom \rightarrow ny))$ exceeds 0. If we drop this assumption then the causalist will only decline the bet at $z \geq z^* = 1$, at which rate EDT will also endorse not betting, because the expected value of *iihom* 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 we now argue, if we *do* drop the assumption then there will inevitably be *other* EPR situations in which EDT and CDT disagree. We address this point first in connection with (B2).

5. B-type theories

Suppose then that we drop the assumption above. Since Cr is a probability function the only alternative is that for any i = 1, 2, 3 we have:

$$(46) Cr(iihom \rightarrow ny) = Cr(iihom \rightarrow yn) = 0$$

Now consider a decision situation just like the six-option case that we considered in section 3 in connection with the hidden-variable theories under (A2). 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 2 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 we'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 EPR case 1

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 (4) and (5). In particular then, the reasoning behind (19) and (21) applies in this case too and shows that any ‘het’ option, which gets a V-score of 0.75, is EDT-preferred to every ‘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 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 you set *that* receiver to 1 or 2. Nor is it obvious that there *is* such a partition.

This is in fact the crucial point of contrast between Table 1 and Table 5. In Table 1 the prior instruction set is both causally independent of your choice and determinative of your payoff in conjunction with your 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 your 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 impose this condition by brute force: i.e., by requiring that you choose the kind of bet on any given run (i.e. between ‘hom’ and ‘het’) *after* the run is over but *before* you have had a chance to see the relevant readings. So we can write e.g.:

$$(47) \text{Ch}(yy \mid 12) =_{\text{def.}} \text{Ch}(yy \mid 12\text{het}) = \text{Ch}(yy \mid 12\text{hom})$$

(ii) Since on (B2) the readings on the receivers are causally independent of one another, the chance of either reading on either receiver is independent of the reading 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.:

$$(48) \text{Ch}(yy \mid 12) = \text{Ch}(y_{AyB} \mid 1_A 2_B) = \text{Ch}(y_A \mid 1_A 2_B) \text{Ch}(y_B \mid 1_A 2_B)$$

$$(49) \text{Ch}(yy \mid 22) = \text{Ch}(y_{AyB} \mid 2_A 2_B) = \text{Ch}(y_A \mid 2_A 2_B) \text{Ch}(y_B \mid 2_A 2_B)$$

Note that, as (49) 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.:

$$(50) \text{Ch}(y_A \mid 1_A 2_B) = \text{Ch}(y_A \mid 1_A)$$

$$(51) \text{Ch}(y_A \mid 2_A 2_B) = \text{Ch}(y_A \mid 2_A)$$

$$(52) \text{Ch}(y_B \mid 2_A 2_B) = \text{Ch}(y_B \mid 2_B)$$

Again and as (51) and (52) illustrate, this point applies to all pairs of settings, not only those to that the present decision situation associates with bets.²¹

(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 (46) 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 (46) that you are certain of:

$$(53) \text{Ch}(yy \vee nn \mid 11) = 1$$

$$(54) \text{Ch}(yy \vee nn \mid 22) = 1$$

$$(55) \text{Ch}(yy \vee nn \mid 33) = 1$$

Focusing on (54)—the arguments from (53) and (55) are parallel—we see that:

$$(56) \text{Ch}(yy \mid 22) + \text{Ch}(nn \mid 22) = 1$$

²¹ Note that points (i)-(iii) suffice to derive a Bell inequality for the Ch function. This does *not* contradict the predictions of quantum mechanics so long as the chances given by Ch do not reflect long-run relative frequencies. It is permitted if, for example, Ch represents single-case chances that vary from case to case. And that is what Ch *should* represent if conditional chance matters to *Causal* Decision Theory: for CDT is supposed to be sensitive to the tendency of a setting to causally promote this or that outcome in the particular decision situation to which you are applying it. The situation here is similar to that in Newcomb's problem, where, even though there is a long-run correlation between one's choosing one box and this having been predicted, the latter is, on any occasion, conditionally *independent* of the former with respect to the appropriately causal chance function. This is consistent with the claim that chances control long run frequencies if, as in the Newcomb and as here, these one-off conditional chances vary from one occasion to the next.

And so by (49), (51), (52) and their analogues for 'nn' we have:

$$(57) \text{Ch}(y_A | 2_A) \text{Ch}(y_B | 2_B) + \text{Ch}(n_A | 2_A) \text{Ch}(n_B | 2_B) = 1$$

And corresponding results follow from (53) and (55).

Now from (57) and its analogues we get:²²

$$(58) \text{Ch}(y_A | 1_A) = \text{Ch}(y_B | 1_B) \in \{0, 1\}$$

$$(59) \text{Ch}(y_A | 2_A) = \text{Ch}(y_B | 2_B) \in \{0, 1\}$$

$$(60) \text{Ch}(y_A | 3_A) = \text{Ch}(y_B | 3_B) \in \{0, 1\}$$

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 (y , y_n 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:

$$(61) \text{Ch}(y_A | 1_A) = \text{Ch}(y_B | 1_B) = 0$$

$$(62) \text{Ch}(y_A | 2_A) = \text{Ch}(y_B | 2_B) = 1$$

$$(63) \text{Ch}(y_A | 3_A) = \text{Ch}(y_B | 3_B) = 1$$

Then by (47), (48) and (50) it follows that:

$$(64) \text{Ch}(yy | 12het) = 0$$

More generally, any specification of the conditional chances in (58)-(60), together with a specification of your choice, determines the reading on the receivers and hence also your payoff. Finally, whichever of the eight possibilities in (58)-(60) obtains is causally independent of your choice: for if chance conditional on what you do takes either value 0 or value 1 then nothing that you do could *make* any difference to that conditional chance itself.

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 your choice in that situation. To that end we'll use the following code: '**abc**', where $a, b, c \in \{0, 1\}$, means:

$$(65) \mathbf{abc} \equiv_{\text{def.}} \text{Ch}(y_A | 1_A) = \text{Ch}(y_B | 1_B) = a \wedge \text{Ch}(y_A | 2_A) = \text{Ch}(y_B | 2_B) = b \wedge \text{Ch}(y_A | 3_A) = \text{Ch}(y_B | 3_B) = c$$

So for instance, '**011**' corresponds to the possible distribution stated at (61)-(63). The new representation of the problem then looks like this:

²² To see this consider that (57) takes the form $xy + (1-x)(1-y) = 1$. So for $(0, 0)$ $0(x, y)$, $(1, 1)$ the only solutions are $x = y = 0$ and $x = y = 1$.

	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 EPR case 1 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 we can apply (3) to Table 6. The U-scores for the three ‘hom’ options are:

$$\begin{aligned} (66)U(12\text{hom}) &= 2 (\text{Cr}(\mathbf{111}) + \text{Cr}(\mathbf{110}) + \text{Cr}(\mathbf{001}) + \text{Cr}(\mathbf{000})) \\ (67)U(13\text{hom}) &= 2 (\text{Cr}(\mathbf{111}) + \text{Cr}(\mathbf{101}) + \text{Cr}(\mathbf{010}) + \text{Cr}(\mathbf{000})) \\ (68)U(23\text{hom}) &= 2 (\text{Cr}(\mathbf{111}) + \text{Cr}(\mathbf{100}) + \text{Cr}(\mathbf{011}) + \text{Cr}(\mathbf{000})) \end{aligned}$$

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

$$\begin{aligned} (69)U(12\text{het}) &= \text{Cr}(\mathbf{101}) + \text{Cr}(\mathbf{100}) + \text{Cr}(\mathbf{011}) + \text{Cr}(\mathbf{010}) \\ (70)U(13\text{het}) &= \text{Cr}(\mathbf{110}) + \text{Cr}(\mathbf{100}) + \text{Cr}(\mathbf{011}) + \text{Cr}(\mathbf{001}) \\ (71)U(23\text{het}) &= \text{Cr}(\mathbf{110}) + \text{Cr}(\mathbf{101}) + \text{Cr}(\mathbf{010}) + \text{Cr}(\mathbf{001}) \end{aligned}$$

But these scores exactly parallel the U-scores of the corresponding options in the (A)-type EPR case 1, except with ‘1’ and ‘0’ in place of ‘Y’ and ‘N’ respectively. See Table 1 and equations (24)-(29). So from this point we can apply exactly parallel reasoning to that applied to the (A)-type case at the corresponding point in the argument, since nothing in *that* part of the argument ((30)-(35)) 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, we 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, we have:

$$(72)\text{For any } i < j: ij\text{het} \not\boxtimes_{\text{EDT}} ij\text{hom in Table 5}$$

$$(73)\text{For some } i < j: ij\text{hom} \not\boxtimes_{\text{CDT}} ij\text{het in Table 5}$$

Without loss of generality we can take $i = 1$ and $j = 2$ to witness (73), in which case the following decision situation represents a technically feasible scenario 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 EPR case 2

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

It might be worth briefly stepping back from the formal details to give an intuitive overview of the construction. The basic idea is that 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 (4) 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.²³

6. C-type theories

Approaches to the EPR experiment that accept non-causal correlations between the actions of experimenters and the outcomes of measurements (perhaps because they are space-like separated), i.e. (C2)-type approaches, also offer the prospect of a divergence between EDT and CDT. One reason for this is that the foregoing argument of course applies as well to them as it does to (B2)-type theories: CDT and EDT will give conflicting advice over Table 4, or over Tables 5-7, to any agent who accepts (C2). But it is illuminating to discuss a further case over which EDT and CDT give different advice specifically in consequence of the (C2) version of parameter dependence.

First we need to outline a little more of the details of parameter dependence. Given any value X for the hidden variables, and letting s_A and s_B be variables ranging over the readings at each receiver, with settings i and j respectively, we have from Bayes's Theorem that:

$$(74) \quad \text{Cr}(s_A, s_B | i, j, X) = \text{Cr}(s_A | s_B, i, j, X) \text{Cr}(s_B | i, j, X)$$

Assuming outcome independence,²⁴ this becomes

²³ It's also worth contrasting the construction in this paper with two other attempts (the only ones known to us) 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 we 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 no good 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 correct. But it has taken some work to show this, including the invention of a totally new family of problems $D(i, z)$.

²⁴ Of course, it is consistent to deal with EPR-like phenomena by denying *both* outcome independence *and* parameter independence (although we know of no such approach in the literature). In that case, our argument above cannot be run; but the arguments covering (B2) in sections 4-5 will still stand.

$$(75) \quad \text{Cr}(s_A, s_B | i, j, X) = \text{Cr}(s_A | i, j, X) \text{Cr}(s_B | i, j, X)$$

But we must have that

$$(76) \quad \text{Cr}(s_A, s_B | i, j, X) \neq \text{Cr}(s_A | i, X) \text{Cr}(s_B | j, X)$$

on pain of otherwise deriving an experimentally violated Bell inequality (at least on the assumption, which we have made throughout, that one's credences match the observed frequencies). Due to symmetry between i and j , we can without loss of generality say that

$$(77) \quad \text{Cr}(s_B | i, j, X) \neq \text{Cr}(s_B | j, X)$$

But assuming (C2), it must also be true that

$$(78) \quad \text{Cr}(i \rightarrow s_B | j, X) = \text{Cr}(s_B | j, X)$$

since the agent does not *cause* the outcome at B by choosing a setting at A. We will rely on (77) and (78).

We invite our agent to play the following game, at cost $\$z$. If the agent agrees to play, then she sets her device at A to i . Meanwhile, it is arranged that, space-like separated from this setting event—space-like separated, if we like, even from the agent's choice whether to play the game—a measurement is made at B on setting j . Once the news comes in from B, if the outcome is y , then the agent receives some reward $\$w_y$; if the outcome is n , then the agent receives some other reward $\$w_n$. Before she decides to play, the agent knows both: (i) the value of the hidden variable X ; and (ii) that the setting at B is j . The choice facing the agent is represented by the following table:

	y_B	n_B
P	w_y	w_n
Q	0	0

Table 8: (C)-type EPR case

(Note that (C2) entails that we can treat the outcomes at B as states of nature.)

The V-score for the agent playing the game (= P) is:

$$(79) \quad V(P) = V(i, y_B, j, X) \text{Cr}(y_B | i, j, X) + V(i, n_B, j, X) \text{Cr}(n_B | i, j, X) = w_n - z + (w_y - w_n) \text{Cr}(y_B | i, j, X)$$

while, given (78), the U-score is:

$$(80) \quad U(P) = V(i, y_B, j, X) \text{Cr}(y_B | j, X) + V(i, n_B, j, X) \text{Cr}(n_B | j, X) = w_n - z + (w_y - w_n) \text{Cr}(y_B | j, X)$$

From (77), we know that these scores must be unequal. There are two possibilities. If $\text{Cr}(y_B | i, j, X) > \text{Cr}(y_B | j, X)$, then we set $w_y > w_n$. Then $V(P) > U(P)$. If $\text{Cr}(y_B | i, j, X) < \text{Cr}(y_B | j, X)$, then we set $w_y < w_n$. Once again $V(P) > U(P)$. If we now set the cost of the game $\$z$ such that:

$$(81) \quad w_n + (w_y - w_n) \text{Cr}(y_B | i, j, X) < 2z < w_n + (w_y - w_n) \text{Cr}(y_B | j, X)$$

(which is always possible) then EDT recommends that the agent plays the game, while CDT recommends that the agent declines.

One drawback of this argument against CDT is that it depends on the game-setter and the agent both having knowledge of the values of the hidden variables—or at least a probability distribution over such values other than that recommended by the Born rule. We merely point out in response to this that the most successful hidden variable completion of quantum mechanics, the de Broglie-Bohm theory, allows for the possibility of such ‘non-equilibrium’ probability distributions (see Valentini 2002).

7. *D-type theories*

Interpretations of quantum mechanics that fall under the broad heading of Everettian or ‘many worlds’ are characterized as postulating no collapse of the quantum state, nor any hidden variables.²⁵ In consequence, *all* possible outcomes of a given measurement are taken to occur, in some sense of ‘occur’. As a result, the assignment of non-extremal probabilities to measurement outcomes is a rather delicate matter, and so deserves a separate treatment.

(D1) The first version of the Everettian account we shall discuss is widely known as the Deutsch-Wallace account. This account may be characterized by what Wallace (2012: 228) calls the decision-theoretic strategy: non-extremal probabilities for measurement outcomes enter the theory through the rational preferences of agents. More specifically, there is a representation theorem: the (typically non-extremal) probabilities for measurement outcomes as prescribed by the Born rule are, under certain assumptions, the uniquely rational credences for an agent to assign to those outcomes just prior to measurement.

Before we get to these assumptions, we must first say something about how Everettian rational agents are to be located in the quantum formalism. Following Wallace 2002, rational agents are, like other macroscopic objects, relatively stable patterns in the states enjoyed by the underlying physical ontology (elementary particles or fields), which we assume is governed by quantum mechanics. As such, an experimenter, or rather her particles, upon making a measurement, becomes entangled with her measurement device (which, being a measurement device, has become entangled with the system being measured). This entanglement spreads uncontrollably to the surrounding environment, giving rise to an entangled superposition of approximately dynamically isolated, quasi-classical ‘branches’. Each branch exhibits its own relatively stable, experimenter-like pattern; in that sense we may say that a

²⁵ Everett 1957.

measurement causes a splitting of the experimenter into several co-existing post-measurement copies, each of whom experiences her own outcome.

This approximate branching into quasi-classical alternatives is taken to justify the application of decision theory in a context of classical branching (such as envisaged by Lewis 1976). The experimenter, prior to measurement, anticipates a splitting as just described, and must somehow weigh the outcomes expected to be experienced by her various post-measurement descendants, to decide now upon some course of action—such as whether or not to take a bet on the measurement's outcome. Note that to assume that this weighing of future outcomes makes sense does not require us to claim that the weights correspond to any degree of *uncertainty* on the experimenter's part, about e.g. which descendant she will turn out to be. (This is just as well, since it is doubtful whether this question has any sense; but see Saunders 1998 for a defence.) Following Greaves 2004, we may call the experimenter's weights on anticipated outcomes a 'caring measure': it is a measure of how much the experimenter cares about each of her post-measurement descendants.

The central claim of the Deutsch-Wallace approach is that there is one uniquely rational caring measure. The assumptions required to prove this result have been honed somewhat since the first paper by Deutsch 1999; the most recent treatment (which we rely on here) is summarized in Wallace 2012. The assumptions fall under three broad categories: axioms of richness, which put minimal conditions on the set of actions available to the agent; axioms of rationality, which require the agent's preferences to be transitive and consistent over time (particularly, between pre- and post-branching selves); and axioms of 'Everettian rationality'. These demand that: (i) the agent be indifferent to the fine-grained details of the physical state, given its corresponding macroscopic outcome; (ii) the agent be indifferent too to the branching brought about by the act of measurement; (iii) the agent's preferences supervene only on the post-measurement states; and (iv) the agent's preferences be robust under mild perturbations (where the metric to determine what counts as 'mild' is inherited from the inner product of the quantum formalism).

We will not engage here in a critical discussion of the assumptions required for the representation theorem; our interest is only in what the approach has to say about the comparative merits of EDT and CDT. On this matter, the Deutsch-Wallace approach seems to put us in rather a back-to-front position. We are looking for scenarios in which credences for various outcomes may be defined, and for which the recommendations from two theories of rational behaviour, EDT and CDT, come apart. But here we already have a theory of rationality: the one encapsulated in the Deutsch-Wallace assumptions. And we use that theory to *derive* credences for the experimenter. What can be said about the relative standings of EDT and CDT?

We doubt that any wedge can be driven between EDT and CDT in the Deutsch-Wallace approach. We do not expect the two theories to give conflicting advice, for the simple reason that the CDT proponent will be able to interpret any dependence between the experimenter's actions and the resulting outcomes, which the EDT proponent will represent with conditional probabilities, as causal. In performing any measurement the experimenter brings about her own splitting, so any probabilistic dependence of observed out-

comes on the experimenter's actions is explicable in terms of the experimenter's causing those outcomes.

This is true even in the case where, due to the caring measure's exhibiting outcome dependence, the non-local correlation between distant outcomes is thought to be brought about by acts of measurement at the local devices. The reason that it is *usually* (i.e. in non-Everettian approaches) tempting to say that the correlation here must be non-local is that a measurement could also be made, still space-like separated, on the distant device to find (as quantum mechanics predicts) correlated results. In an interpretation in which only *one* outcome occurs at each device, the distant device must already 'know' which outcome to give, given a measurement on the local device.

But in Everettian quantum mechanics, *every* possible outcome happens at each device, so there is no reason to demand that the distant device's outcome is already decided once the local measurement has been made. Instead, the rapid entanglement of the local device's state with its surrounding environment (including the experimenter), and the rapid entanglement of the distant device's state with *its* surrounding environment, can each happen separately and locally (that is, within the light-cone). It is only once the two expanding spheres of ever-entangling environment meet that the correlation between the two devices can be said to come about—it is only then that two measurement outcomes belong to the same quasi-classical branch. In that case, according to Everettian quantum mechanics, a local measurement brings about a *local* correlation between the two devices (see Timpson and Brown 2002). It is permitted for the CDT proponent to treat this correlation as caused—and perverse of him not to. Consequently, the credence assigned to any causal link between the experimenter's actions and measurement correlations ought to be the same as the corresponding conditional credences, and we have no disagreement between CDT and EDT.

(D2) The second and final version of the Everett interpretation we shall consider is the 'many minds' interpretation. According to this version (see Albert & Loewer 1988), we associate with each *prima facie* observer a continuum of minds—i.e., a continuum of separate consciousnesses. Given any *prima facie* observer, the proportion of its associated consciousness that sees a particular outcome of a measurement is given, by dint of a separate postulate, by the Born rule.

If we now trace the biography of a single such consciousness over time, then it becomes clear that life for such a consciousness is completely fatalistic. (We don't say 'deterministic', since that term is often used in a technical sense, to describe the supervenience of future states on past states; and here supervenience of the future on the past does not hold, since many consciousnesses share the same past but not the same future.) That is, by selecting any single mind one has thereby selected a maximally specific biography.

It is hard to see how a conflict between EDT and CDT can even get off the ground here, for the very notion of causation seems impossible to apply. In other cases of complete determination (e.g. as in Newtonian mechanics), we have *determinism*, so we can appeal to the laws to ground the counterfactual claims that may support an account of causation.

One proposal is to adopt an account of *probabilistic* causation in which the probabilities reflects self-locating ignorance on the part of individual minds. However, this tempting line leads to us to the same agreement between EDT and CDT on all matters that we saw for the Deutsch-Wallace approach. After all, the conditional probabilities appealed to will be exactly those that EDT will respect.

In sum: if we adopt the Everett interpretation, then we will find no conflict between EDT and CDT. Either the recommendations they make will, plausibly, be identical; or else we cannot talk of causation, and therefore not of CDT, at all.

Indeed the point goes beyond the two versions of the Many Worlds Interpretation that this section considers. Any plausible version of the Many Worlds approach must deal with the preferred basis problem, and it is almost universally accepted that the most promising solution lies with decoherence (see Vaidman 2014). But, as we argued in our discussion of the Deutsch-Wallace approach, it is likely that the CDT-proponent will seize on decoherence as the causal process responsible for correlations between outcomes; and this will result in agreement between the probabilities of the CDT-proponent's causal statements and the corresponding EDT-proponent's conditional probabilities. Therefore any version of the Many Worlds Interpretation that appeals to decoherence is immune to the difficulties that seem to arise for some version of every one of the other approaches that we have considered (i.e. (A), (B) and (C)-type approaches). Thus from the perspective of the Many World Interpretation as presented in Vaidman 2014, the conflict between CDT and EDT cannot be seen.

Let us now summarize the position with regards to the eight interpretations of the EPR experiment as described in s. 1. The following table identifies (i) the interpretations on which EDT and CDT are in disagreement over some practical question; (ii) the locus or loci of disagreement in each case.

Interpretation	Case where EDT and CDT disagree
Deny 'no conspiracy': (A1)	None
Deny 'no conspiracy': (A2)	Tables 1-2; Table 4 or Tables 5-7
Deny outcome independence: (B1)	None
Deny outcome independence: (B2)	Table 4 or Tables 5-7
Deny parameter independence: (C1)	None
Deny parameter independence: (C2)	Table 4 or Tables 5-7; Table 8
Everett – DWG: (D1)	None
Everett – Many minds: (D2)	None

Table 9: Summary of our findings

That is, on three of these eight interpretations it is possible to construct a case in which CDT and EDT give different advice. In particular, a causalist who accepts (A2), (B2) or (C2) is committed to some course of action that violates the evidentialist conception of rationality and indeed to betting against the EPR statistics themselves.

8. EPR vs CDT

Of course the fact that a conflict between EDT and CDT can feasibly arise, at least on interpretations (A2), (B2) and (C2), does not by itself refute either theory. But it does make especially vivid just what is involved in preferring CDT to EDT. The constructions in this paper are feasible cases where the theories 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 construct realistic cases where the theories gave different advice.²⁶

And on reflection they prompt two obvious objections to CDT. The first is familiar: ‘Why ain’cha rich?’ CDT advises anyone who accepts interpretation (A2) to take option 12hom in Table 2 whereas EDT will advise 12het. Similarly, CDT advises anyone who accepts (B2) or (C2) to take option 12hom in the decision problem in Table 7²⁷; 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 this; so 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.²⁸ What is new is the context, which is naturalistic by the usual standards of these debates and, we think, all the more vivid for all that.

We should like to put the point as strongly as this. Focus on (B)-type EPR case 2 as represented in Table 7 and suppose that all parties’ credences make it a site of conflict between EDT and CDT. Let you and one of us 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 my turns 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, in every eight runs. We publicly challenge any defender of CDT to play this game against us. Unfortunately we

²⁶ E.g. the ‘medical’ Newcomb problems, on which see Price 1991 and Price 2012: 511-13.

²⁷ At any rate this is so if the agent’s credences satisfy (46). If they do *not* satisfy (46) 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 (A2), (B2) or (C2)—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.

²⁸ See e.g. Gibbard and Harper 1978: 369.

are certain that these people would stop being causalists long before they stopped being solvent.

The second objection to CDT is not that it is giving bad advice in any identifiable case, but that *what* advice it 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 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 instead one of the non-causal interpretations (= (A2), (B2), (C2)). CDT prefers 12hom to 12het in e.g. Table 2 if we are given e.g. (A2); 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 prefers 12hom to 12het in Table 7; but again, it reverses these preferences under (B1). So its recommendation depends not only upon the statistical facts that we can observe but also upon theoretical questions about causal structure 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 12het in Table 7 will on average exceed the return to 12hom by 50%.

To make it more vivid: suppose that I am running Table 7-style books on two similar EPR devices, X and Y, and that you for some reason think that action-at-a-distance is operative in X but not in Y, the devices being otherwise identical. Then CDT will advise different approaches to X and to Y, even though you know in advance that they will generate the same expected returns 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 arbitrarily confident of the same long-run 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 may have no observable consequences. That in turns 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 supposed to be offering *practical* advice, not to be engaging in theoretical speculation.

This aspect of CDT is not a moving part of the other cases that distinguish it from EDT. In standard Newcomb cases (Nozick 1969: 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 prediction²⁹; similarly in cases not involving dominance such as ‘Death in Damascus’ and its variants.³⁰ So although it is (in our view) always true that the statistical facts are enough by themselves for practical purposes, it is only in the EPR 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.³¹

But just how serious a problem does CDT really face? After all, we have only shown that on *three out of eight* possible hypotheses about the workings of the device do EDT and CDT make different practical recommendations. So if you don’t fully grant any of (A2), (B2) or (C2), as many people do not, it seems that you can maintain CDT in the face of everything that we have said so far.

In fact this is not true. As we’ll see, all it takes for the clash to arise, and so all it takes for the criticisms here to bite, is that you are not *absolutely confident* that all of (A2), (B2), (C2) are *false*. For any agent who gives any of these non-causal hypotheses *any* positive credence at all, it’s possible to construct a case concerning which: (a) CDT and EDT give different advice to the agent; (b) EDT is known to have a better expected return than CDT; (c) CDT still gives practical advice that depends on your metaphysical views about the setup. So the argument here pressed against CDT does not only apply to that (presumably quite small) class of people who fully accept one of (A2), (B2) and (C2) but also to that (presumably much larger) class of people who don’t completely reject all of them.

The following section expands on this point before turning to three other objections.

9. Objections

A defender of CDT might object: (i) that—as just suggested—no reasonable person would place *all* of her credence in any of (A2), (B2) or (C2) (ii) that CDT *does not* make the recommendations that we have claimed, given hypothesis (B2); (iii) that it is unclear whether it does, because it is unclear what

²⁹ 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; but then it is unclear what is attractive about a distinctively *causal* formulation of their common content, at least on a non-Humean conception of causality.

³⁰ Gibbard and Harper 1978: 372-5; Egan 2007.

³¹ As Cavalcanti points out (2010: 569), CDT was itself developed as a response to the standard Newcomb problem, with which EDT *seems* unable to deal. If the examples in this paper show that CDT has to go, then we might wonder where this leaves EDT. Our own view, which we cannot defend at length here (but see [suppressed]) is that EDT is in fact a perfectly adequate normative decision theory—or at any rate it is more adequate than CDT. This commits us to a ‘one-boxing’ position on Newcomb’s problem. We bite that ‘bullet’: one-boxing, which EDT recommends, is superior in Newcomb’s problem for essentially the same, statistical, reason that the ‘het’ options are superior to the ‘hom’ options in Table 6. For a recent defence of this line on Newcomb’s problem see Price 2012. Thanks to a referee for raising this point.

are supposed to count as causal connections here; (iv) that in any case the examples, being non-constructive, pose no definite objection to CDT.

(i) *Mixed theoretical beliefs.*

We have so far proceeded entirely on the assumption that an agent lends all of her credence to some one of the eight theoretical options that we identified at section 2.

That is, we are asking her to accept one of three specific interpretations of the experiment i.e. (A2), (B2) or (C2). And this is something of a stretch: it is hard to be *certain* that there really are lawlike (i.e. theoretically predicted) correlations between states that are causally independent and share no causal ancestor.³²

What is more likely is that a well-informed agent spreads her credence across *all* of the causal hypotheses concerning the working of the device, just as in any everyday decision situation 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 EPR cases that we have been considering?

Yes they are, as we now argue. First and in order to simplify matters, let us define C (for 'causality') to abbreviate those hypotheses (A1), (B1), (C1), (D1) and (D2) on which the causal influence of the settings forces EDT and CDT to agree over all cases: in particular over Tables 4-7. So $\neg C$ abbreviates all of those other hypotheses (A2), (B2) and (C2) that deny any such form of influence:

$$(83) \quad C \equiv_{\text{def.}} A1 \vee B1 \vee C1 \vee D1 \vee D2; \text{ so:}$$

$$(84) \quad \neg C \equiv A2 \vee B2 \vee C2$$

Next, consider some decision problem D (i, z) as at Table 4. For any such problem, the V-score of betting iihom 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

³² On the other hand this claim is not the contradiction that Maudlin appears to imply it is when he writes: 'if a theory predicts a correlation, then that correlation cannot, according 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).

the observed statistics. So EDT will recommend *ihom* in Table 4 to any agent meeting that condition, including any agent whose credence is divided amongst the hypotheses that we 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 (46) holds. Recall that the only premises in the argument for (46) were (44) and (45), 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* (46) holds. So we may take forward (46) from the foregoing argument.

Next, consider the following decision problem, $\alpha > 1$:

	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 10: Mixed EPR case α

In this problem there are two different types of states of nature: those in which the causal hypothesis *C* holds 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 both receivers give reading 'y' then one gets a payoff of α , whichever of the hypotheses *C* and $\neg C$ is true. 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 :

$$(84)V(12\text{hom}) = \alpha \text{Cr}(yy \vee nn \mid 12\text{hom}) = 0.25\alpha$$

$$(85)V(12\text{het}) = \text{Cr}(yn \vee ny \mid 12\text{het}) = 0.75$$

—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 $\text{Cr}(C)$.

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:

$$(86)U(12\text{hom}) = \alpha (\text{Cr}(12\text{hom} \rightarrow (C \wedge (yy \vee nn))) + \text{Cr}(12\text{hom} \rightarrow (\neg C \wedge (yy \vee nn)))$$

$$(87)U(12\text{het}) = \text{Cr}(12\text{het} \rightarrow (C \wedge (yn \vee ny))) + \text{Cr}(12\text{het} \rightarrow \neg C \wedge (yn \vee ny))$$

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':

$$(88)\text{Cr}(12 \rightarrow (C \wedge S)) = \text{Cr}(C \wedge (12 \rightarrow S))$$

$$(89)\text{Cr}(12 \rightarrow (\neg C \wedge S)) = \text{Cr}(\neg C \wedge (12 \rightarrow S))$$

Now the right hand sides of (88) and (89) resolve into:

$$(90)\text{Cr}(C \wedge (12 \rightarrow S)) = \text{Cr}(12 \rightarrow S | C) \text{Cr}(C)$$

$$(91)\text{Cr}(\neg C \wedge (12 \rightarrow S)) = \text{Cr}(12 \rightarrow S | \neg C) \text{Cr}(\neg C)$$

It is straightforward to calculate the conditional probabilities on the right of (90) and (91) for the two possible values of S . In particular, if the causal hypothesis is *true* then we should expect the settings to have a causal effect upon the readings that mirrors the statistics (4) and (5). So we have:

$$(92)\text{Cr}(12 \rightarrow yy \vee nn | C) = 0.25$$

$$(93)\text{Cr}(12 \rightarrow yn \vee ny | C) = 0.75$$

But if the causal hypothesis is *false* then no setting has any causal impact on the reading on the opposite wings. So by the argument at s. 4 we have:

$$(94)\text{Cr}(12 \rightarrow (yy \vee nn) | \neg C) = \text{Cr}(\mathbf{111} \vee \mathbf{110} \vee \mathbf{001} \vee \mathbf{000} | \neg C)$$

—where $\mathbf{111}$, $\mathbf{110}$ etc. are as defined at (65).³³ 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 (86) and (87) to get:

$$(95)U(12\text{hom}) = 0.25\alpha c + \alpha(1-c) \text{Cr}_{\neg C}(\mathbf{111} \vee \mathbf{110} \vee \mathbf{001} \vee \mathbf{000})$$

$$(96)U(12\text{het}) = 0.75c + (1-c) \text{Cr}_{\neg C}(\mathbf{101} \vee \mathbf{100} \vee \mathbf{011} \vee \mathbf{010})$$

³³ Note that on this definition (94) holds good on hypothesis (A2) as well as on hypotheses (B2) and (C2) because on the former hypothesis $\mathbf{111}$, $\mathbf{110}$ etc. are respectively equivalent to \mathbf{YYY} , \mathbf{YYN} etc.

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

$$(97)U(13\text{hom}) = 0.25\alpha c + \alpha(1-c)Cr_{-c}(\mathbf{111} \vee \mathbf{101} \vee \mathbf{010} \vee \mathbf{000})$$

$$(98)U(13\text{het}) = 0.75c + (1-c)Cr_{-c}(\mathbf{110} \vee \mathbf{100} \vee \mathbf{011} \vee \mathbf{001})$$

$$(99)U(23\text{hom}) = 0.25\alpha c + \alpha(1-c)Cr_{-c}(\mathbf{111} \vee \mathbf{011} \vee \mathbf{100} \vee \mathbf{000})$$

$$(100)U(23\text{het}) = 0.75c + (1-c)Cr_{-c}(\mathbf{110} \vee \mathbf{010} \vee \mathbf{101} \vee \mathbf{001})$$

Now we know by the structurally identical reasoning of (24)-(34)—which applies just as well here because Cr_{-c} is a probability distribution—that (say) twice the marginal credence on the right of (95) equals or exceeds the corresponding quantity on the right of (96):

$$(101)2Cr_{-c}(\mathbf{111} \vee \mathbf{110} \vee \mathbf{001} \vee \mathbf{000}) \geq Cr_{-c}(\mathbf{101} \vee \mathbf{100} \vee \mathbf{011} \vee \mathbf{010})$$

--at any rate, either this inequality holds or some corresponding one holds for the marginal credences in (97) and (98), or for those in (99) and (100).

So suppose without loss of generality that (101) is true. If we now write $t =_{\text{def.}} \alpha - 2$, $p =_{\text{def.}} Cr_{-c}(\mathbf{111} \vee \mathbf{110} \vee \mathbf{001} \vee \mathbf{000})$ and $q =_{\text{def.}} Cr_{-c}(\mathbf{101} \vee \mathbf{100} \vee \mathbf{011} \vee \mathbf{010})$ then subtracting (96) from (95) gives:

$$(102)U(12\text{hom}) - U(12\text{het}) = 0.25c(t-1) + p(2+t)(1-c) - q(1-c)$$

Since (101) tells us that $2p - q \geq 0$, it follows that:

$$(103)U(12\text{hom}) - U(12\text{het}) > 0 \text{ if } 0.25c(t-1) + pt(1-c) > 0; \text{ hence:}$$

$$(104)U(12\text{hom}) - U(12\text{het}) > 0 \text{ if } t > c/(c + 4p(1-c))$$

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 (104).

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 that $p > 0$ is an innocuous assumption. And by (84) and (85), 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), (B2), or (C2) then EDT and CDT will diverge over Mixed EPR case α for some α .

So the objection fails: as long as the agent is not absolutely certain of retroactive causation, superluminal causation or the Everett interpretation, it is possible to construct an EPR case in which EDT and CDT give divergent advice. Any such case will equally support both of the arguments against CDT that section 8 based upon 'pure' EPR cases like those in Tables 1, 4, 5 and 8.

(ii) *Counterfactual indefiniteness.*

The (B2)-type cases in Tables 4-7 require that for CDT to give the verdicts that we are attributing to it, e.g. at (73), there must be a definite credence in counterfactuals such as (12hom \rightarrow yy); for expressions denoting such quanti-

ties appear throughout the argument, e.g. at (44). 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.

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.³⁴ This is a very natural 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* unnoticed difference in their *actual* constitution.³⁵

But 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 counterfactuals to be barely true. An actual such position arises from Lewis's semantics for counterfactuals (1973), on which the truth-value of a counterfactual concerning an object's behaviour depends only on that object's (or its counterparts') 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 (12hom \rightarrow yy) is perfectly well defined as long as there is an appropriately measurable set of worlds in which A and B are set to '1' and '2' and 'yy' has this or that chance of occurring. Nothing in this account demands any categorical truthmaker for the counterfactual.³⁶

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 then it now seems that we cannot make *any claim at all* 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 the

³⁴ Dummett 1976: 53.

³⁵ The example is from Evans (1980: 276-7).

³⁶ A possible such position 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 they would have to be barely true. It is for that phenomenalist simply a brute fact, not obtaining in virtue of anything that is actually already there, or in virtue of anyone's actual present willing, that if I were now in my office I should see my desk (Berlin 1999: 43ff.).

difference between a theory that advises (say) betting in Table 4 and one that advises not betting there, 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 practical action is called for but whereof CDT is silent.

That silence extends even to the simplest cases: if nothing makes the counterfactual $1_A \rightarrow y_A$ true then nothing makes $\neg 1_A \rightarrow y_A$ true either. But then CDT gives no advice even in the almost trivial situation where one must 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
$\neg 1_A$	0	0

Table 11: (B)-type EPR case 3

If CDT had anything to say about this case it would be that it's worth taking the bet if and only if your $Cr(1_A \rightarrow y_A) > 0.5$; but since, on the current proposal, the expression on the left hand side of this inequality is meaningless, CDT has no advice to give about even this simplest of decision problems. On the other hand, EDT gives advice here, and it is commonsensical: you should take such bets as Table 11 describes if and only if you'd expect to win them more often than to lose them (given risk-neutrality).

Perhaps the causalist could reply that EDT gives correct advice in EPR cases where the relevant counterfactuals make no sense; but in more everyday cases (which we *can* describe in terms of causality) we should follow the advice of CDT. But what could motivate this eclecticism? Why wouldn't it be equally sensible, by causalist lights, to follow *maximin*, or *minimax regret*, or any other decision rule you please, in those cases where CDT is silent?³⁷ If EDT *is* giving proper advice in EPR cases then that must be because the statistical facts (4) and (5) 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 EPR 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 (66)-(71) in connection with Table 6. FCDT then gives exactly the same results as those claimed for CDT in the EPR case. Now the eclectic view has no answer to the question: if we should prefer EDT to FCDT where they clash in EPR cases, then why should we not equally prefer EDT to CDT in classical i.e. non-EPR cases where *they* clash?

³⁷ For details of maximin see Resnik 1987: 26-7; for minimax *ibid.* : 28-32.

(iii) *The varieties of causation*

The third line of objection is that by presenting the eight interpretations as genuine alternatives we are 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 we haven't said anything about what causation is supposed to be. But that is only because our purposes do not demand it. The idea behind the approach was supposed to be that there are *some* feasible theoretical assumptions on which EDT and CDT diverge, not that *every* theoretical approach forces that view of things.³⁸ 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? We don't think that there ever was any answer to this question. What was intuitively appealing about CDT's appeal to causation was not any specific feature of the counterfactual or causal relation that some explications of this notion preserve but which others do not.³⁹ 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 what it is *about* 'bringing about' that somehow works a magic that mere statistics can never achieve.

(iv) *Does it matter that the argument is non-constructive?*

The discussion of (A2) in s. 3 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 ss. 4-5 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 Ta-

³⁸ 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 1969: 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 admit that interpretation if anything does. So there is nothing new about the idea of presenting an example against a background of specific theoretical assumptions.

³⁹ Papineau 2003: 179. As well as noting that no answer to this question is standardly forthcoming, Papineau discusses and criticizes an *evidentialist* justification of causal decision theory (*ibid.* ss. 12-15). For further criticisms of that approach see Ahmed 2014 ch. 8.

ble 4 or over the B-type EPR case 1 at Table 5, but nothing in the argument tells us which. It was also non-constructive in the same sense as our discussion of the A-type interpretations: even within Table 5 itself there is nothing to say 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 and other cases. All that those arguments required was (i) that some such cases exist; (ii) in the case of the first argument: that in those cases the statistical facts (4) and (5) favour EDT over CDT; and (iii) 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 by all possible, observations. 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 we are 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 runs of the device.⁴⁰ On the first three runs of the device we 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 the payoffs to ‘12hom’ and to ‘13hom’, and permutes the payoffs to ‘12het’ and to ‘13het’. Problem 3 is like problem 1 except that it permutes ‘12hom’ with ‘23hom’ and ‘12het’ with ‘23het’. If the agent 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 (46) must be false. In that case, let him face a sequence of decision problems (Problem 4, Problem 5...) where the Problem i 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, construct an EPR case in which his choice violates EDT’s preferences over some specific and identifiable set of options.

10. Conclusion

Suppose that your beliefs about the arrangement in s. 1 meet the following very light condition: that you have *some* non-zero credence in at least one of the hypotheses (A2), (B2) or (C2) on which the wings of the experiment are causally unrelated. In that case, the constructions in this paper represent realistic and novel loci of disagreement between the purely statistical (i.e. evidential) and the causal approach to decision theory. We also think that they represent clear counterexamples to the causal theory.

Our recommendation to all such persons, which presumably includes most of us, is to drop Causal Decision Theory, if they ever held it, and instead

⁴⁰ 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.

to base decisions about these games purely on the statistical evidence where this is available. Of course in the case considered here, those statistics, as embodied in (4) and (5), are not only available but as extensive, consistent and reliable as these things ever get.⁴¹

References

- Ahmed, A. 2014. *Evidence, Decision and Causality*. Cambridge: CUP.
- Albert, D. Z. and Loewer, B. 1988. Interpreting the many worlds interpretation. *Synthese* 77: 195-213.
- 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 (ed.), *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.
- Deutsch, D. 1999. Quantum theory of probability and decisions. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 455: 3129-3137.
- 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.
- 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.
- Evans, P. W., Price H., and Wharton, K. B. 2013. A new slant on the EPR-Bell experiment. *BJPS* 64: 297-324.

⁴¹ The authors thank Jeremy Butterfield, Huw Price, and two referees for this journal. AA wrote most of his contribution to this paper whilst on a Visiting Fellowship at the Research School of Social Sciences, ANU, Canberra, and wishes to thank that institution for its hospitality, and also the Faculty of Philosophy, University of Cambridge, for granting him leave for this period. AC wishes to thank the British Academy for its generous support. Some of the material in sections 3-5 and 9(i) appears in ch. 6 of A. Ahmed, *Evidence, Decision and Causality* (Cambridge University Press 2014) and appears here with the permission of Cambridge University Press.

- Everett, H. 1957. 'Relative state' formulation of quantum mechanics. *Reviews of Modern Physics*, 29: 454-462.
- Feynman, R. P. 1987. Negative probability. In Hiley, B., and Peat F. D. (ed.), *Quantum implications: Essays in honour of David Bohm*. New York: Routledge: 235-248.
- Ghirardi, G. C., Rimini, A., and Weber, T. 1986. Unified dynamics for microscopic and macroscopic systems. *Physical Review D* 34: 470.
- 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.
- Greaves, H. 2004. Understanding Deutsch's probability in a deterministic multiverse. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 35: 423-456.
- Hofer-Szabó, G. 2012. Separate common causal explanation and the Bell inequalities. *International Journal of Theoretical Physics* 51: 110-123.
- Holland, P. 1993. *The quantum theory of motion: an account of the de Broglie-Bohm causal interpretation of quantum mechanics*. Cambridge: CUP.
- Hume, D. 1949 [1739]. *Treatise of Human Nature*. Ed. with an analytical index by L. A. Selby-Bigge. Oxford: Clarendon Press.
- Jeffrey, R. C. 2004. *Subjective Probability: The Real Thing*. Cambridge: CUP.
- Joyce, J. 1999. *Foundations of Causal Decision Theory*. Cambridge: CUP.
- . 2002. Levi on Causal Decision Theory and the possibility of predicting one's own actions. *Phil. Stud.* 110: 69-102.
- Laudisa, F. 2001. Non-locality and theories of causation. In Butterfield, J. and T. Placek (ed.), *Non-Localilty and Modality*. Dordrecht: Kluwer.
- Lewis, D. K. 1973. *Counterfactuals*. Oxford: Blackwell.
- . 1976. Survival and Identity. In Rorty, A. O. (ed.), *The Identities of Persons*. University of California Press: 17-40.
- . 1981. Causal decision theory. *AJP* 59: 5-30. Reprinted in his *Philosophical Papers Vol. II*. Oxford: OUP 1986: 305-339.
- Lewis, P. J. 2006. Conspiracy theories of quantum mechanics. *BJPS* 57: 359-381.
- Maudlin, T. 2002. *Quantum Non-Localilty 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. 1969. 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.
- Papineau, D. 2001. Causation as a guide to life. In his *Roots of Reason*. Oxford: OUP 2003: 167-212.
- 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.

- Redhead, M. 1987. *Incompleteness, Nonlocality, and Realism*. Oxford: Clarendon Press.
- Reichenbach, H. 1984. *The Direction of Time*. New York: Dover.
- Resnik, M. 1987. *Choices: an Introduction to Decision Theory*. Minneapolis: University of Minnesota Press.
- Saunders, S. 1998. Time, Quantum Mechanics, and Probability. *Synthese* 114: 373-404.
- Savage, L. J. 1972. *The Foundations of Statistics*. 2d ed. New York: Dover.
- Shimony, A. 1990. An exposition of Bell's theorem. In Miller, A. I. (ed.), *Sixty-Two Years of Uncertainty*. New York: Springer: 33-43.
- Skyrms, B. 1984. *Pragmatics and Empiricism*. New Haven: Yale UP.
- Timpson, C. and Brown, H. R. 2002. Entanglement and Relativity. In Lupacchini, R. and V. Fano. (ed.), *Understanding Physical Knowledge*. University of Bologna: CLUEB: 147-166.
- Vaidman, L. 2014. Many Worlds Interpretations of Quantum Mechanics. In Zalta, E. N. (ed.), *Stanford Encyclopaedia of Philosophy*. Available online at: <http://plato.stanford.edu/entries/gm-manyworlds/> (accessed 10 June 2014).
- Van Fraassen, B. C. 1991. *Quantum Mechanics*. Oxford: OUP.
- Valentini, A. 2002. Signal-locality and subquantum information in deterministic hidden-variables theories. In Placek, T. and J. Butterfield, J. (ed.), *Non-locality and Modality*. Dordrecht: Kluwer: 81-103.
- Wallace, D. 2002. Everett and structure. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics* 34: 87-105.
- . 2003. Everettian rationality: defending Deutsch's approach to probability in the Everett interpretation. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics* 34: 415-439.
- . 2012. How to Prove the Born Rule. In Saunders, S., Barrett, J., Kent, A. and Wallace, D. (eds.), *Many Worlds? Everett, Quantum Theory, & Reality*. Oxford: OUP: 227-263.