Received: 19 November 2019 | Revised: 28 October 2022 | Accepted: 16 January 2023

DOI: 10.1111/phib.12298

ORIGINAL ARTICLE

ANALYTIC PHILOSOPHY WILEY

# Conditional causal decision theory reduces to evidential decision theory

**Mostafa Mohajeri**

Department of Management, Science and Technology, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran

**Correspondence**
Mostafa Mohajeri, Department of Management, Science and Technology, Amirkabir Unoversity of Technology, 3rd floor, Zakaria Razi Building, Saeed Alley, Enghelab Street, Tehran, Iran, 1591634311.
Email: mostafa.mohajeri@gmail.com

## 1 | INTRODUCTION

Advocates of Causal Decision Theory (**CDT**) argue that Evidential Decision Theory (**EDT**) is inadequate because it gives the wrong result in Newcomb problems. Egan (2007) provides a recipe for converting Newcomb problems to counterexamples to **CDT**, arguing that **CDT** is inadequate too. Egan's argument led to the formulation of several new decision theories designed to conform to the supposedly correct pre-theoretic judgments about the rationality of acts in Newcomb problems and Egan cases. Two major theories that proposed with this aim in mind are the Conditional Causal Decision Theory (**CCDT**) (as I call it), proposed by Edgington (2011), and the Benchmark Theory, proposed by Wedgwood (2013).

While the Benchmark Theory has been criticized, I assume, successfully by Briggs (2010) and Bassett (2015), **CCDT** has been widely taken uncritically as a version of **CDT** that conforms to the assumed pre-theoretic judgments about the rationality of acts in Newcomb problems and in Egan's cases: see Ahmed (2012, p. 386), Pittard (2016, p. 20) and Williamson (2019, p. 7).

This paper have benefited a lot from conversations and discussions with Martin Peterson and Amir Saemi and their helpful comments on its various drafts. I am immensely grateful to them for their help and support. I should tank an anonymous referee of this journal whose helpful comments and suggestions on an earlier draft of this paper helped me to improve the argument of section 5 of the paper. I would like to thank also Hamid Vahid, Dorothy Edgington, Seyed N. Mousavian, Mahmoud Morvarid, Mohsen Zamani, Sajed Tayebi, Sven Diekmann, Laleh Ghadakpour and Ali Hossein Khani for helpful discussion and valuable comments and feedbacks on some version of this paper. Previous versions of this paper has been presented at "P&E Departmental Workshops in autumn 2012/Winter 2013" of the section of Philosophy & Ethics of Technology at Eindhoven University of Technology (TU/e) in Sep. 2012, and at Biweekly Philosophy Seminars of School of Analytic Philosophy, Institute for Research in Fundamental Sciences (IPM), in Oct. 2012 and Apr. 2016.

  

After a brief introduction of **CCDT** (in section 2), I argue (in section 3) that, despite Edgington's promise, **CCDT** gives the wrong result in Newcomb problems. This, I assume, shows that **CCDT** is superior neither to the classic visions of **CDT** nor to **EDT**. Then (in section 4), I examine Edgington's treatment of Newcomb problems and explain why it fails to provide support for **CCDT**. Returning to **CCDT** (in section 5), I, then, argue, in general, that **CCDT**, though formulated in terms of causal notions, collapses back into **EDT**, i.e. it is systematically bound to deliver the same results as does **EDT** in any decision problem.

## 2 | CONDITIONAL CAUSAL DECISION THEORY

Let us begin our inquiry by explaining **CCDT**. Like other participants in the debate over causal and evidential decision theory, **CCDT** assigns expected values to acts and, in the end, commands the agent in a given decision situation to choose one of the acts with maximal expected value (provided that one exists, which is guaranteed where there are only finite number of acts available to the agent). The expected value which is assigned to each act is a weighted average of the values of its possible consequences.

Edgington introduces her favored theory in the context of discussing criticisms against a classic version of **CDT** in which "the weights are probabilities of counterfactual or subjunctive conditional propositions" (Edgington, 2011, p. 78). First, she argues that "we have [...] a reason to go for conditional probabilities" (p. 84), i.e. to go for "$p((A \rightarrow C)|A)$" rather than "$p(A \rightarrow C)$", where "$A \rightarrow C$" is to be read as counterfactual conditional (p. 83):

> Maybe conditional propositions are normally independent of their antecedents, in which case it does not matter whether you consider $p(A \rightarrow C)$ or $p((A \rightarrow C)|A)$. But where they do come apart, as in Egan's examples, it is the conditional probability that matters: the probability of the conditional *on the assumption that you do the action*. (p. 83).

And, then, observing that "on (almost) any understanding of the conditional proposition, $p((A \rightarrow C)|A) = p(C|A)$" (p. 83), she argues that decision theory should be formulated not in terms of counterfactual conditionals but directly in terms of causal notions:

> A properly causal decision theory should be up-front about causation (like causal theories of other things). We need conditional judgments like "If I do $x$, that will bring it about that $C$ (or cause, or produce as a result, or have as an outcome, that $C$)." More cautiously: "If I do $x$, that will contribute to the bringing about that $C$." [...] We need to assess the probability, on the assumption that I do $x$, that $C$ will be a causal consequence. (p. 84).

According to the theory that Edgington proposes, therefore, the weights are conditional probabilities of a certain kind of causal propositions—hence the name "Conditional Causal Decision Theory". More precisely, the value of a consequence $C$ of an act $A$ is weighted by $p((A \rightarrow C)|A)$, where "$A \rightarrow C$" is to be understood as "$A$ causally influences $C$". The expected value of an act $A$ with

**TABLE 1** Possible consequences of acts and values of the consequences in the smoking case.

| | Cancer | No cancer |
|---|---|---|
| $S$: Smoke | $C_{S,1}$: Short life with joy of smoking<br>$v(C_{S,1}) = 1$ | $C_{S,2}$: Long life with joy of smoking<br>$v(C_{S,2}) = 11$ |
| $Q$: Quit | $C_{Q,1}$: Short life without joy of smoking<br>$v(C_{Q,1}) = 0$ | $C_{Q,2}$: Long life without joy of smoking<br>$v(C_{Q,2}) = 10$ |

possible consequences of $C_1$, $C_2$, … $C_n$, therefore, is defined, according to **CCDT**, by the following formula[1]:

$$exv(A) = \sum_{i=1}^{n} p((A \rightarrow C_i)|A) \times v(C_i) \quad (1)$$

Now as an example, consider an ordinary case of smoking where an agent is deliberating whether or not to smoke. Let her believe that smoking is likely to cause cancer such that those who smoke are 60% probable to develop cancer while those who do not are only 20% probable to do so. Possible consequences of the acts and values of the consequences are shown in the following Table 1 (the figures are taken from a similar example in Edgington, 2011, p. 76):

Using formula (1), we have:

$$exv(S) = p((S \rightarrow C_{S,1})|S) \times v(C_{S,1}) + p((S \rightarrow C_{S,2})|S) \times v(C_{S,2}) \quad (2)$$

$$exv(Q) = p((Q \rightarrow C_{Q,1})|Q) \times v(C_{Q,1}) + p((Q \rightarrow C_{Q,2})|Q) \times v(C_{Q,2}) \quad (3)$$

To compute the expected value of smoking according to **CCDT**, so, we need to assess $p((S \rightarrow C_{S,1})|S)$ and $p((S \rightarrow C_{S,2})|S)$. To do so, note that according to the description of the case, the agent believes that smoking is 60% probable to cause cancer. Assuming that the agent smokes, whenever smoking causes cancer, $C_{S,1}$ will be the consequence and so $S$ causally influences $C_{S,1}$; otherwise, that is, when smoking does not cause cancer, $C_{S,2}$ will be the consequence and, as joy of smoking (which is a component of the consequence $C_{S,2}$) is an effect of smoking in those cases, we have "$S$ causally influences $C_{S,2}$". Therefore, $p((S \rightarrow C_{S,1})|S) = 0.6$ and $p((S \rightarrow C_{S,2})|S) = 0.4$. The expected value of smoking, therefore, is:

[1]Though the formula (1) is not recorded in Edgington's paper, there is, I believe, little room for doubt that Edgington commits herself to it based on the comparisons she draws between her favored theory and its rivals she criticizes and the general descriptions she gives of "the weights" she suggests to be used in the calculation of the expected values of acts. Considering the whole paper, including Edgington's discussion of specific examples where she is to apply her theory to specific cases, however, it may seem that there might be other ways of formalizing Edgington's favored theory. I argue, in section 4, that Edgington's calculations of the expected values of acts in the Fisher case do not accord with (1) and, then, briefly explore whether we can devise a general definition for the expected values of acts based on Edgington's calculations of the expected values of acts in the case. This brief exploration leads to the tentative conclusion that there is no simple and straightforward way to do so.
In the face of this conflict between Edgington's description of her favored theory in general terms and her way of calculating the expected values of acts in specific examples, I give more weight to her general remarks in part because I am relying on a response Edgington kindly wrote to an email of mine, in which she explains she thinks that her way of calculating the expected values of acts where she applies theory to specific examples (pp. 84–5) "is not a proper way of averaging the values of the different outcomes" and expresses her hope that "there is a 'proper' way of doing it that is consistent with the earlier claims of the paper".

$$exv(S) = 0.6 \times 1 + 0.4 \times 11 = 5 \quad (4)$$

Similarly, to compute the expected value of quitting, we need to assess $p((Q \rightarrow C_{Q,1})|Q)$ and $p((Q \rightarrow C_{Q,2})|Q)$. Now, note that according to the description of the case, the agent believes that those who do not smoke are 20% probable to get cancer. Assuming that the agent quits, whenever the agent gets cancer, $C_{Q,1}$ will be the consequence, and in all those cases we have $Q$ causally influences $C_{Q,1}$ because not enjoying the joy of smoking (which is a component of the consequence $C_{Q,1}$) is a result of quitting; otherwise, that is, when the agent does not get cancer, $C_{Q,2}$ will be the consequence and, for similar reasons we have in all these cases "$Q$ causally influences $C_{Q,2}$". Therefore, $p((Q \rightarrow C_{Q,1})|Q) = 0.2$ and $p((Q \rightarrow C_{Q,2})|Q) = 0.8$. The expected value of quitting, therefore, is:

$$exv(Q) = 0.2 \times 0 + 0.8 \times 10 = 8 \quad (5)$$

The expected value of quitting, therefore, exceeds that of smoking and, so, **CCDT** recommends quitting in this case.

Having explained **CCDT**, let us, now, examine what the theory commands in an example of Newcomb problem.

## 3 | APPLYING CCDT TO NEWCOMB PROBLEMS

A famous example of Newcomb problems, known as the Fisher case, which is sometimes categorized as a medical-Newcomb problem, is a variant of the ordinary case of smoking discussed above.

> *Fisher case*
> The agent believes that smoking is strongly correlated with lung cancer, but only because there is a common cause— a bad allele of a certain gene that tends to cause both smoking and cancer. Once the presence or absence of the gene is fixed, there is no additional correlation between smoking and cancer. She prefers smoking without cancer to not smoking without cancer, and smoking with cancer to not smoking with cancer. Should she smoke or not?[2]

It seems intuitively clear that in this situation the rational act for the agent is to smoke: though smoking is a sign that you have the bad gene, you either have the gene or you do not, and refraining from smoking is not going to reduce your chances of getting cancer; it only deprives you of the pleasure of smoking. This pre-theoretic judgment is widely shared by many decision theorists, including Edgington, many causalists[3] and some others[4]. Edgington (2011, p. 77) uses the Fisher case to make a case against **EDT** as it commands the agent to quit smoking in the Fisher case. I argue, in this section, that **CCDT** makes the same commands in the Fisher case.

To see **CCDT**'s recommendations in the Fisher case, we need to evaluate the expected values of acts using formula (1). As the possible consequences of smoking and quitting are not different in the Fisher case from the ordinary smoking case, using formula (1), therefore, we have:

[2]The description of the case is excerpted from Egan (2007, p. 94).

[3]See, for example, Gibbard and Harper (1978), Sobel (1978, 1986), Cartwright (1979), Skyrms (1980, 1982), Lewis (1981), Joyce (1999, 2002), Arntzenius (2008) and Cantwell (2010).

[4]See, for example, Jeffrey (1981, 1983/1990) and Wedgwood (2013).

$$exv(S) = p((S \rightarrow C_{S,1})|S) \times v(C_{S,1}) + p((S \rightarrow C_{S,2})|S) \times v(C_{S,2}), \tag{6}$$

and

$$exv(Q) = p((Q \rightarrow C_{Q,1})|Q) \times v(C_{Q,1}) + p((Q \rightarrow C_{Q,2})|Q) \times v(C_{Q,2}), \tag{7}$$

which are just like (2) and (3), while here $exv(S)$ and $exv(Q)$ are respectively the expected value of smoking and that of quitting in the Fisher case.

To facilitate the comparison between **CCDT**'s results in the Fisher case and the ordinary case of smoking, let us assume that values of different consequences for the agent and the coefficients of the correlation between smoking and cancer are the same in both cases. The values of consequences for the agent are again as is shown in Table 1, the agent's conditional probability of developing cancer given smoking is 0.6 and her conditional probability of developing cancer given quitting is 0.2. The only difference between the two cases, therefore, is the agent's belief about the causal background responsible for the correlation between smoking and cancer.

To evaluate the expected value of smoking, therefore, we just need to assess the values of $p((S \rightarrow C_{S,1})|S)$ and $p((S \rightarrow C_{S,2})|S)$. Let us, first, see how we can use the assumptions of the decision situation to assess $p((S \rightarrow C_{S,1})|S)$, which, to recall, is the probability of $S$ causally influencing $C_{S,1}$ conditioned on $S$: assuming that the agent smokes, it is 0.6 probable that the agent develops cancer and the consequence $C_{S,1}$, that is, "short life with joy of smoking", happens. Though in this case smoking does not cause cancer, it causes some other component of the consequence $C_{S,1}$, i.e. the agent's getting joy or relief out of smoking, and, so, whenever "short life with joy of smoking" happens in the Fisher case, we can say it is caused by smoking. The value of the conditional probability of smoking causally influencing $C_{S,1}$ conditioned on smoking, therefore, is 0.6, that is, $p((S \rightarrow C_{S,1})|S) = 0.6$.

Following the same logic, we can conclude also that $p((S \rightarrow C_{S,2})|S) = p(\neg Cancer|S) = 0.4$. Given the value of consequences $C_{S,1}$ and $C_{S,2}$ in Table 1 and using (6), so, the expected value of smoking is as follows:

$$exv(S) = 0.6 \times 1 + 0.4 \times 11 = 5 \tag{8}$$

To evaluate the expected value of quitting, the other act that is available to the agent in the Fisher case, we need to assess $p((Q \rightarrow C_{Q,1})|Q)$ and $p((Q \rightarrow C_{Q,2})|Q)$: Assuming that the agent quits smoking, it is 0.2 probable that the agent develops cancer and the consequence $C_{Q,1}$ "short life without joy of smoking" happens. Again, though, in this case, quitting does not cause cancer, it causally affects some other component of the consequence, i.e. the agent's not smoking and therefore not getting the relevant joy or relief, and, so, again whenever $C_{Q,1}$ happens, we can say it is causally influenced by quitting, and so $p((Q \rightarrow C_{Q,1})|Q) = 0.2$.

Following the same logic we can conclude also that $p((Q \rightarrow C_{Q,2})|Q) = p(\neg cancer|Q) = 0.8$. Given the value of consequences $C_{Q,1}$ and $C_{Q,2}$ in Table 1 and using (7), so, we can compute the expected value of quitting as follows:

$$exv(Q) = 0.2 \times 0 + 0.8 \times 10 = 8 \tag{9}$$

The expected value of quitting, so, exceeds that of smoking: $exv(Q) > exv(S)$. Contrary to what Edgington claims, **CCDT**, therefore, recommends quitting in the Fisher case. So, it gets the wrong result in Newcomb problems.

## 4 | WHY EDGINGTON'S TREATMENT OF THE FISHER CASE FAILS TO SUPPORT CCDT

As we saw in the last section, **CCDT** recommends quitting in the Fisher case. Why does, then Edgington claim that **CCDT** recommends smoking in the Fisher case? Let us see how Edgington (2011) analyses the case: the values she suggests to be assigned to the different consequences of acts are the same as those we assigned to the consequences (as shown in Table 1) in our computation of the expected values of acts above (p. 84). The same is true also for the magnitude of the correlation coefficients between smoking and cancer (p. 84). Edgington, however, suggests different weights to be used in the calculation of the expected values of the acts (which for each act, as mentioned, is a weighted average of the values of the possible consequences of the act in question). The following passage reveals her suggestion:

I suggest that the weights should be something like:

| | | **[the Fisher case]** |
|---|---|---|
| (*a*) | *p*(Short life as a result of smoking\|Smoke) | 0 |
| (*b*) | *p*(Not (short life as a result of smoking)\|Smoke) | 1 |
| (*c*) | *p*(Long life as a result of quitting\|Quit) | 0 |
| (*d*) | *p*(Not (long life as a result of quitting)\|Quit) | 1 |

(Edgington, 2011, p. 84) [5].

Given these assumptions, the expected values of the acts are to work out as follows:

$$\begin{aligned} exv(S) &= p(\text{Short life as a result of smoking}|\text{Smoke}) \times v(\text{Short life with smoking}) \\ &+ p(\text{Not (short life as a result of smoking)}|\text{Smoke}) \times v(\text{Long life with smoking}) \\ &= p(\text{Short life as a result of S}|\text{S}) \times v(C_{S,1}) + p(\text{Not (short life as a result of S)}|\text{S}) \\ &\times v(C_{S,2}) = 0 \times 1 + 1 \times 11 = 11 \end{aligned} \tag{10}$$

$$\begin{aligned} exv(Q) &= p(\text{Long life as a result of quitting}|\text{Quit}) \times v(\text{Long life without smoking}) \\ &+ p(\text{Not (long life as a result of quitting)}|\text{Quit}) \times v(\text{Short life without smoking}) \\ &= p(\text{Not (long life as a result of Q)}|\text{Q}) \times v(C_{Q,1}) + p(\text{Long life as a result of Q}|\text{Q}) \\ &\times v(C_{Q,2}) = 1 \times 0 + 0 \times 10 = 0 \end{aligned} \tag{11}$$

Computed in this way, as you see, $exv(S)$ exceeds $exv(Q)$. Based on this, Edgington, then, concludes that the theory implies that "it is rational to smoke" and, so, conforms to her assumed pre-theoretic judgment about the rationality of acts in the Fisher case (2011, p. 84).

The problem with Edgington's calculation of the expected values of acts in this case is simple: it does not match with the formalism of **CCDT**. Consider, for example, her calculation of the

[5] "An alternative, more explicit, and more symmetric way of setting out the possibilities in this example" is suggested in footnote 20 (p.85). The footnote, unlike this passage from the main text, however, does not specify how "the weights" should be set in the calculation of the expected values of acts and, so, does not by itself dictate a way of calculating the expected values of acts in the Fisher case different from the calculations of section 3.

expected value of smoking:[6] the expected value of smoking, as mentioned, is basically a weighted average of the values of its possible consequences. The possible consequences of smoking in our case are:

$C_{S,1}$: Short life with joy of smoking,
$C_{S,2}$: Long life with joy of smoking.

Analyzing the consequences to their components, $C_{S,1}$ and $C_{S,2}$ can respectively represented by "$(sl\&js)$" and "$(ll\&js)$" (where, naturally, "$sl$" is an abbreviation for "short life", "$js$" for "joy of smoking", and "$ll$" for "long life"). In Edgington's calculation of the expected value of smoking, the values of $(sl\&js)$ and $(ll\&js)$ are respectively weighted by $p((S \rightarrow sl)|S)$ and $p(\neg(S \rightarrow sl)|S)$, while, according to **CCDT**, they should be weighted respectively by $p((S \rightarrow (sl\&js))|S)$ and $p((S \rightarrow (ll\&js))|S)$ which neither in general nor even in the special condition of the Fisher case are respectively equal to those used in Edgington's calculation. $p((S \rightarrow sl)|S)$, for example, is not equal to $p((S \rightarrow (sl\&js))|S)$, except in a very special case where we have $p(sl|S) = 0$: we know that $p((S \rightarrow sl)|S) = 0$ (as it is assumed, in the Fisher case, that smoking does not causally influence short life) while $p((S \rightarrow (sl\&js))|S) = p(sl|S)$ (assuming that smoking causes joy of smoking and that the agent smokes, whenever the consequence $(sl\&js)$ happens, this consequence is causally affected by the agent's smoking).

As we saw, what needs to be taken into account when we evaluate the expected values of acts, according to **CCDT**, is the causal relations between the acts and the final consequences (/outcomes) of the decision problem, where the final consequences may be composed of many components. In her evaluation of the expected values of smoking and quitting in the Fisher case, reported above, Edgington, however, arbitrarily takes a component of a possible consequence of each act and evaluates the expected value of the act focusing only on the causal relation between the act and that very component of its consequence.

So far, I have argued that **CCDT** delivers the wrong result in our example of Newcomb problems and have shown that Edgington's calculations of the expected values of acts that seems to deliver the right results in the example in question do not accord with **CCDT**, as it is formulated in general terms by Edgington. In light of these observations, one might naturally wonder if we can take Edgington's way of calculation of the expected values of acts in the Fisher case as a base and revise the formalism of the theory to bring it in line with the calculations that deliver the right results in the Fisher case. Exploring this possibility in the rest of this section, I will argue that there is no simple and straightforward way to do so.

Let us recall, for example, formula (10) which Edgington suggested to be used in the calculation of the expected value of smoking in the Fisher case: using the abbreviations that I introduced above, (10) can be rewritten as:

$$exv(S) = p((S \rightarrow sl)|S) \times v(sl\&js) + p(\neg(S \rightarrow sl)|S) \times v(ll\&js), \quad (12)$$

which can be simplified as:

$$exv(S) = p((S \rightarrow sl)|S) \times v(sl\&js) + (1 - p((S \rightarrow sl)|S)) \times v(ll\&js). \quad (13)$$

To generalize this formula, we face an obvious problem: the formula essentially uses the accidental fact that there are only two possible consequences for smoking in the Fisher case.

[6]The situation of quitting is quite similar.

Another problem we face is that the result of Edgington's calculation of the expected value of smoking is sensitive to the choice of the consequence and the component of the consequence the causal effect of the act on which is being used in the calculation of the expected value. Given Edgington's assumptions about the values of the consequences in the Fisher case and using formula (13) the expected value of smoking is to work out as in (14) to equal 11:

$$exv(S) = p((S \to sl)|S) \times v(sl\&js) + (1 - p((S \to sl)|S)) \times v(ll\&js) = 0 \times 1 + 1 \times 11 = 11, \quad (14)$$

The evaluation of the expected value of smoking, however, varies if instead of $S \to sl$, one chooses to use $S \to js$ or $S \to ll$, the causal effect of smoking on some other component of a consequence of smoking in the Fisher case, in the formula and changes mutatis mutandis other parameters of the formula[7]. See (15) and (16):

$$exv_{S\to js}(S) = p((S \to js)|S) \times v(sl\&js) + (1 - p((S \to js)|S)) \times v(ll\&js) = 1 \times 1 + 0 \times 11 = 0. \quad (15)$$

$$exv_{S\to ll}(S) = p((S \to ll)|S) \times v(ll\&js) + (1 - p((S \to ll)|S)) \times v(sl\&js) = 0 \times 11 + 1 \times 1 = 1. \quad (16)$$

To generalize the formula even to cover cases where the acts have only two consequences, so, one need to specify in a principled way that the causal effect of the act to what component of what consequence is to be used in the calculation of the expected value of the act. Edgington does not do this, and I suspect that this cannot be done easily.

## 5 | CCDT COLLAPSES BACK INTO EDT

We have seen so far that, contrary to Edgington's claim, **CCDT** recommends quitting in the Fisher case, which is the same act that the classic **EDT** commands in this case. Here, I wish to argue that **CCDT**'s conformity to **EDT** in the Fisher case is not accidental or due to special features of the Fisher case or Newcomb problems but is, in fact, an instance of quite a general pattern. I argue, therefore, that **CCDT** is bound to give the same results in any decision problem as the ones the classic **EDT** does. To do so, it is enough to be argued that, for any act in a decision problem, the expected value that **CCDT** assigns to the act equals to the expected value that **EDT** assigns to it.

For an act $A$ in a decision problem $D$ where it is assumed that doing $A$ will be followed by exactly one of its possible consequences $C_1, C_2, \ldots C_n$, we have, on the one hand, the expected value defined according to **CCDT**:

$$exv_{CCDT}(A) = \sum_{i=1}^{n} (p((A \to C_i)|A) \times v(C_i)), \quad (17)$$

and, on the other, concluded from the definition of the expected values of acts according to **EDT**:

[7]Note that the same situation hold for the other act that is available to the agent in the Fisher case. Edgington's calculation of the expected value of quitting uses the conditional probability of the proposition "$S \to ll$". Assuming that "$Q$" stands for "quitting" and "*ns*" for "no smoking", formula (11) can be rewritten as:
(1) $exv_{Q\to ll}(Q) = p((Q \to ll)|Q) \times v(ll\&ns) + (1 - p((Q \to ll)|Q)) \times v(sl\&ns) = 0 \times 10 + 1 \times 0 = 0$,
The result of the calculation, however, varies if instead of "$Q \to ll$", one chooses to use, for example, "$Q \to sl$" in the formula and changes mutatis mutandis other parameters of the formula:
(2) $exv_{Q\to sl}(Q) = p((Q \to sl)|Q) \times v(sl\&ns) + (1 - p((Q \to sl)|Q)) \times v(ll\&ns) = 0 \times 0 + 1 \times 10 = 10$,
This change, in turn, may lead to the recommendation of quitting as a rational act in the Fisher case, as $exv_{Q\to sl}(Q) > exv_{S\to ll}(S)$.

$$exv_{EDT}(A) = \sum_{i=1}^{n}(p(C_i|A) \times v(C_i)). \tag{18}$$

To argue that for any act in a decision problem, the expected value that **CCDT** assigns to the act equals to the expected value that **EDT** assigns to it, I first consider a special group of decision problems, which we may call "adequately specified decision problems", and argue that, for any act $A$ in an adequately specified decision problem, $exv_{CCDT}(A) = exv_{EDT}(A)$.

*Adequately specified decision problems* are the decision problems in which the consequences of acts are discriminated and specified in such a fine-grained and detailed way that each possible consequence of each act contains at least a component that is supposed to be caused (or causally influenced) by the act. Consider, for example, the specification of the ordinary case of smoking in section 2 or that of the Fisher case in section 3: they both constitute decision problems that are *adequately specified* as the two possible consequences of each act in both cases contain at least a component that is supposed to be caused by the act. Both possible consequences of smoking in the Fisher case, for example, contain 'joy of smoking' that is caused by smoking.

To argue that, for any act $A$ in an adequately specified decision problem, $exv_{CCDT}(A) = exv_{EDT}(A)$, it is enough to be argued that for arbitrary act $A$ and consequence $C_i$ of $A$ in these decision problems, $p((A \to C_i)|A) = p(C_i|A)$. Given (17) and (18), the latter claim implies the former. Can it be argued for arbitrary act $A$ and consequence $C_i$ of $A$ in an adequately specified decision problem that $p((A \to C_i)|A) = p(C_i|A)$? I think so, and here is my argument:

Note that it is generally the case that for any $E_1$ and $E_2$, where $E_1$ and $E_2$ represent arbitrary events, $p(E_2|E_1) \geq p((E_1 \to E_2)|E_1)$. It is so because assuming that $E_1$ happens, whenever $E_1$ causes $E_2$ and, so, "$E_1 \to E_2$" is the case, $E_2$ also happens. It can be concluded, therefore, that for any $A$ and $C_i$, $p(C_i|A) \geq p((A \to C_i)|A)$.

Admittedly, it is not generally the case that for any $E_1$ and $E_2$, where $E_1$ and $E_2$ represent arbitrary events, $p((E_1 \to E_2)|E_1) \geq p(E_2|E_1)$. Assuming that $E_1$ happens, there might be cases where $E_2$ happens but is not caused by $E_1$. It can be argued, however, that in our special case where $A$ is an act and $C_i$ is a possible consequence of it in an adequately specified decision problem, we have $p((A \to C_i)|A) \geq p(C_i|A)$. This, given $p(C_j|A) \geq p((A \to C_i)|A)$, logically implies that $p((A \to C_i)|A) = p(C_i|A)$ for any act $A$ and consequence $C_i$ of it in adequately specified decision problems.

To see why it is the case that, for an act $A$ and consequence $C_i$ of it in an adequately specified decision problem, $p((A \to C_i)|A) \geq p(C_i|A)$, note that $C_i$ being a possible consequence of $A$ in an adequately specified decision problem has to be specified in part by a component that is supposed to be caused by $A$. It is, therefore, guaranteed, for any $C_i$ that is a consequence of $A$ in an adequately specified decision problem, that, assuming that the agent does $A$, whenever $C_i$ happens, it is caused by $A$ (or $A$ has contributed to the brining it about that $C_i$). It can be concluded, therefore, that for any act $A$ and consequence $C_i$ of it in adequately specified decision problems, $p((A \to C_i)|A) \geq p(C_i|A)$.

To show that for any act $A$ and consequence $C_i$ of it in adequately specified decision problems, $p((A \to C_i)|A) = p(C_i|A)$, the above argument relies on two assumptions about the causal relations in **CCDT**'s formulation. The assumptions can be restated as follows:

$$\text{If } A \to C_i \text{ and } A, \text{then } C_i \tag{19}$$

$$\text{If } A \to c_i, C_i = c_i \& \overline{c_i}, A \text{ and } C_i, \text{then } A \to C_i \tag{20}$$

Let us pause to have a few words about these assumptions.

Equation (19) can be read as "assuming that $A$ causes $C_i$, if $A$ happens, $C_i$ will happen". It is used to show that $p(C_i|A) \geq p((A \to C_i)|A)$. Note that (19) is talking about the situation where "$A \to C_i$" is true (not just the situation where "$A \to C_i$" is to some extent probable). To put it differently, (19) is talking about the situation where $A$ causally influences $C_i$ to the highest degree. Given this consideration, I take (19) to be a quite plausible assumption.

Equation (20) can be read as "assuming that $A$ causes $c_i$, where $c_i$ is a component of a compound event $C_i$ (which is represented by '$C_i = c_i \& \overline{c_i}$'), if $A$ and $C_i$ happen, $A$ can be said to have caused (or casually influenced) $C_i$". Various particular instances of (20) were assumed in sections 2 and 3 to evaluate expected values of acts in different examples. Recall, for example, our reasoning, in section 2, where we were seeking to evaluate the expected value of smoking in the ordinary case of smoking. We argued there that, where smoking does not cause cancer and the agent lives long, as it still causes *joy of smoking* which is a component of the compound consequence *long life with joy of smoking*, we might say "smoking causes (or causally influences) *long life with joy of smoking*". Aside from its particular instances, (20) is used above, in its general form, to show that, for any act $A$ and consequence $C_i$ of it in adequately specified decision problems, $p((A \to C_i)|A) \geq p(C_i|A)$.

It may seem at times odd to speak about an act's causal influence on compound events that may be as comprehensive as an entire state of the world. We normally ask, for example, whether smoking causes *cancer*, not whether smoking causes *short life with joy of smoking*. However, **CCDT**, being formulated in terms of the causal relations between acts and outcomes in decision situations, requires us to ask whether, or to what degree, acts causally influence different outcomes in dictions situations, where outcomes are typically compound events. To apply **CCDT** to general cases, so, we need to rely on, if not a well-developed theory of causation, at least, some general principles that relates causal relations between acts and compound events to causal relations between acts and components of compound events. While, to apply **CCDT** to specific examples, Edgington (see p. 84) seeks to estimate the relevant probabilities of acts causing compound consequences by looking to the causal influence of the acts to the components of the consequences, she does not address the issue in general terms and, in fact, seems to have been inclined to avoid going into these kinds of discussions:

> Finally: Although "causation" is in my title, you have not learned anything about causation in this talk. I do not apologize for that. You do not learn anything about causation in Grice's "The Causal Theory of Perception," Davidson's "Actions, Reasons and Causes," Goldman's "A Causal Theory of Knowing," or Kripke's causal theory of reference. This is part of a tradition of using causation in the explanation of other phenomena, which has become quite respectable in philosophy. (Edgington, 2011, pp. 86–87).

Whether or not one feels much sympathy with Edgington's conception of causal decision theory, as presented above, it seems indispensable for **CCDT** (perhaps unlike the other theories that are mentioned) to rely at least on some general guiding principles about causation to go further than our particular intuitions about causal relations in specific examples if the theory is to be applied to general cases. In this context, I take (20) to be a quite plausible assumption.

To sum up, relying on the above two minimal assumptions about the causal relations in terms of which **CCDT** is formulated, I have shown so far that $exv_{CCDT}(A) = exv_{EDT}(A)$, for any act $A$ in an adequately specified decision problem. I argue now that this result can in effect be generalized to acts in all decision problems. Key to this argument is to distinguish between a decision situation and representations of the decision situation based on which the expected values of acts are evaluated according to a decision theory. An adequately specified decision problem is, in fact, a special representation of a decision situation, for which a decision theory, in our case **CCDT**, may also accept many other representations as legitimate representations. In that case, i.e. where there are more than one legitimate representation for a decision situation according to a decision theory, it is naturally expected (if the theory is to be consistent) that the expected values of the acts remain the same, whatever legitimate representation is used in the evaluation of the expected values according to the theory. To generalize our result about adequately specified decision problems to all decision problems, therefore, it is enough to be argued that any decision situation can be represented by some adequately specified decision problem.

Can any decision situation be represented by an adequately specified decision problem? It seems plausible to assume that the alternative possible outcomes (/consequences) of acts in any decision situation can be discriminated and specified in a way that the specification of each possible consequence of each act includes a component that is supposed to be caused by the act if the acts are assumed to exert at least some influence on the world. There are, however, two issues to deal with if this assumption is to assure us that any decision situation can be represented by an adequately specified decision problem. First, we need to see whether the representations in question can be legitimate representations of the decision situation, i.e. a representation based on which the expected values of acts can be evaluated according to **CCDT**. Second, we need to see whether the assumption that the acts exert at least some influence on the world is restrictive or not. Now, let us first see how we can deal with the former:

Edgington provides no explanation about the specifications of legitimate representations of decision situations in **CCDT**. It can be concluded, however, from the definition of the expected values of acts in **CCDT** that the legitimate representations of a decision situation should specify the alternative possible outcomes (/consequences) of the acts. Moreover, it may seem plausible to require that the specification of the consequences should include all relevant features of the situation that the agent cares about. I see no reason at the outset to impose any additional requirement on legitimate representations of decision situations and, as specifications of the consequences of acts can well include all relevant features of the situation that the agent cares about and, also, a component that is supposed to be caused by the act (if the acts are assumed to exert at least some influence on the world), we seem to be able to conclude our desired result, that is, that any decision situation can be represented by an adequately specified decision problem.

What if one pursues to impose some other requirements on legitimate representations of decision situations, perhaps to block the argument that $exv_{CCDT}(A)$ should equal $exv_{EDT}(A)$ where one does not want it to do so? This strategy, I argue, cannot be successful in remedying **CCDT**, i.e. cannot block the argument that any decision situation can be represented by an adequately specified decision problem, as long as **CCDT** is to respect (21):

(21) A legitimate representation of a decision situation remains legitimate if its specification of the consequences of the acts is revised to become *more fine-grained*, that is, if some consequences of the acts in the first representation is replaced by partitions of the consequences specifying more details about newly identified consequences.

In other words, (21) requires that the expected values of acts evaluated based on a legitimate representation of a decision situation should remain the same if, instead of the representation in question, the expected values are evaluated based on a more fine-grained representation of the decision situation. Intuitively, it amounts to requiring that if a rational agent has evaluated an act taking to the account all the relevant aspects of the situation, her evaluation should not change when she is reminded of some further details about the consequences of the act she does not care about. This is a very plausible requirement, which, in fact, all well-known brands of decision theory satisfy.

Starting from any legitimate representation of a decision situation, i.e. any decision problem, if we revise the representation by making it more fine-grained so as to get a representation in which the specification of each possible consequence of each act includes a component that is supposed to be caused by the act (which we know we can, as discussed before, if the acts are assumed to exert at least some influence on the world), (21) guarantees that the revised representation that we get will be a legitimate representation of the decision situation and, therefore, an adequately specified decision problem, which represents the same decision situation. This will then let us to conclude that the expected values of the acts evaluated based on either of these two representations according to **CCDT** should be the same and equal to the expected values of the acts according to **EDT**.

Having discussed the first issue in the argument that any decision situation can be represented by an adequately specified decision problem, let us now turn to the second issue to see whether the assumption that the acts exert at least some influence on the world is restrictive or not.

Note, first of all, that the influence in question need not be assumed to be important, i.e. of the kind that the agent cares about, nor need it be expected to produce some sure result, i.e. a same result in all states of the world, and nor even need it be assumed to be a positive influence, i.e. lack of the influences of the alternative acts is also counted (like, for example, quitting in the Fisher case where we may say quitting causes lack of joy of smoking). This latter consideration lets us to count what may be described as "inaction" in many decision situations technically as an "act" that exerts influence on the world. Taking these considerations into account, I do not think the assumption that the acts exert at least some influence on the world is a restrictive assumption, at least not in the context of a causal decision theory for which it is natural to assume a causal order in the world and not in the way that may undermine the significance of our result.

Having said that, let me mention that, as far as the argument we have seen so far goes, it does not show that the expected value **CCDT** assigns to an act should equals the one **EDT** assigns to it if the act in question is not assumed to exert influence on the world. It remains to be seen, then, how the expected values of these "acts" can be evaluated according to **CCDT**. I will come back to this point later but let us first see if there are any other ways to argue that $exv_{CCDT}(A) = exv_{EDT}(A)$, for any act $A$ in any decision problem.

Another way to argue that $exv_{CCDT}(A) = exv_{EDT}(A)$, for any act $A$ in any decision problem, is to consider what would happen if, for an act $A$ and a consequence $C_i$ of it in a decision problem, $p((A \rightarrow C_i)|A)$ does not equal $p(C_i|A)$. As we saw above, it is generally the case that for any $E_1$ and $E_2$, $p(E_2|E_1) \geq p((E_1 \rightarrow E_2)|E_1)$, which implies that for any $A$ and all of its consequences $C_i$ in any decision problem, $p(C_i|A) \geq p((A \rightarrow C_i)|A)$. Assuming that the expected value of $A$ according to **EDT**, $exv_{EDT}(A)$, is a weighted average of the values of $A$'s consequences and so the relevant weights of the values of the consequences of $A$ sum up to one (that is, $\sum_i p(C_i|A) = 1$), therefore, we can argue, if for an act $A$ and a consequence $C_i$ of it, $p((A \rightarrow C_i)|A)$ does not equal $P(C_i|A)$, the expected value of $A$ according to the **CCDT**, $exv_{CCDT}(A)$, cannot be a weighted average of the values of $A$'s consequences. The argument goes like this: if for an act $A$ and a consequence $C_i$ of it, $p((A \rightarrow C_i)|A)$ does not equal $p(C_i|A)$, $p((A \rightarrow C_i)|A)$ has to be strictly lower than $p(C_i|A)$, and so, as in general for no consequence of $A$, $p((A \rightarrow C_i)|A)$ can be greater than $p(C_i|A)$ and $\sum_i p(C_i|A) = 1$, $\sum_i p((A \rightarrow C_i)|A)$ has to be less than one. For $exv_{CCDT}(A)$ to be a weighted average of the values of $A$'s consequences, therefore, $exv_{CCDT}(A)$ has to be equal to $exv_{EDT}(A)$.

Let us have a brief comparison between these two arguments and see what we can learn from this second argument about what we have discussed before: contrary to the first argument, in no part of this second argument the discussion is restricted to *adequately specified decision prob-*

*lems*. Moreover, among the two assumptions of the first argument about the causal relations in **CCDT**'s formulation, i.e. (19) and (20), this second argument only relies on (19), 'If $A \to C_i$ and $A$, then $C_i$'.

We may learn, from this second argument, something about **CCDT**'s special requirement for the legitimate representations of decision situations. We learned that for the value that **CCDT** assigns to an act $A$ in a decision situation as its expected value, evaluated based on a representation of the decision situation, to be a weighted average of the values of the act's consequences, we should have $p((A \to C_i)|A) = p(C_i|A)$ for any act $A$ and consequence $C_i$ that is identified as a possible consequence of $A$ in the representation of the decision situation. We can, therefore, conclude that for a representation of a decision situation to be a legitimate representation of the decision situation according to **CCDT**, the consequences of the acts should be discriminated and specified in the representation in question in a way that for any act $A_i$ and any possible consequence $C_{i,j}$ of the act we have $p\big((A_i \to C_{i,j})|A_i\big) = p\big(C_{i,j}|A_i\big)$. This, however, hardly provide any practical guide to how we can construct legitimate representations for decision situations. A plausible suggestion, then, may be to require that the consequences of acts should be discriminated and specified in legitimate representations of decision situations in such a fine-grained and detailed way that each possible consequence of each act contains at least a component that is supposed to be caused by the act, which, in effect, amounts to requiring that only *adequately specified decision problems* can be counted as decision problems. Imposing this requirement on legitimate representations of decision situations not only guarantees that the expected values of the acts, computed according to **CCDT** based on a legitimate representation of a decision situation, will be a weighted average of the values of the act's consequences, but it also guides us on how to construct a legitimate representation for a decision situation.

In light of these discussions we can finally say a few words about the expected value that **CCDT** should assign to "acts" that are not assumed to exert influence on the world. Note, first, that these acts include not just the act that is assumed to exert no influence on the world, but also the acts that are assumed to *probably* exert no influence on the world. These "acts", therefore, can be referred to as the acts that *may* exert no influence on the world. On the one hand, we learned from the second argument that *for any act in any decision problem* the expected value that **CCDT** assigns to the act is equal to the expected value that **EDT** assigns it. On the other hand, as *the decision situations* in which some "acts" may exert no influence on the world cannot be represented by *adequately specified decision problems*, it is not clear whether we can construct legitimate representations for these decision situation (assuming that there can be such decision situations). If we are given *a decision problem*, i.e. a representation of a decision situation, in which some "acts" are represented as having possible "consequences" that cannot possibly be influenced by them, however, we can deliver judgment based on what we have learned so far about **CCDT**: it can be said that the expected value of these acts cannot be evaluated based on the given "decision problem", that is, technically the case is not a decision problem according to **CCDT**.

**CCDT** collapses back into the classic **EDT**, therefore, in the sense that wherever the expected values of the acts can be evaluated by **CCDT**, the values that **CCDT** assigns to the acts are the same as the expected values that the classic **EDT** assigns to them. To put it another way, $exv_{CCDT}(A) = exv_{EDT}(A)$, for any act $A$ in any decision problem, i.e. in any representation of a decision situation that **CCDT** count as "a decision problem", not in any representation of a decision situation that **EDT** count as "a decision problem". In some cases the expected values of the acts can be evaluated according to **EDT**, but not according to **CCDT**.

Edington (2011, p. 83) observes a difference between conditional propositions (indicative or counterfactual conditionals), on the one hand, and causal propositions like "$A$ causes $C$", on the other: while "on (almost) any understanding of the conditional proposition ["$A \to C$"], $p((A \to C)|A) = p(C|A)$," if we read "$A \rightarrow C$" as a causal proposition, the equality does not hold in general. Edington's proposal that we should use the conditional probability of causal propositions in the calculation of the expected values of acts is partly based on this observation. It is clear that if "$A \rightarrow C$" were to be read as a conditional proposition (indicative or counterfactual) the resulting theory would easily collapse to **EDT**, and so would face the same problems that **EDT** does. I argued, however, that though it is not, in general, the case that $p((A \rightarrow C)|A) = p(C|A)$ where "$A \rightarrow C$" is to be interpreted as a causal proposition, the relevant equation is guaranteed to hold when $C$ is a consequence of an act $A$ in a decision problem (at least as long as we understand "decision problem" in the way **CCDT** requires us to do). So, like other brands of conditional decision theory that use conditional probabilities on the condition of the relevant acts being done in the calculation of the expected values of acts, the conditional causal decision theory also collapses back into the classic evidential decision theory.

## ORCID

*Mostafa Mohajeri* https://orcid.org/0000-0002-3509-8197

## REFERENCES


Ahmed, A. (2012). Push the button. *Philosophy of Science*, *79*(3), 386–395. https://doi.org/10.1086/666065

Arntzenius, F. (2008). No regrets, or: Edith Piaf revamps decision theory. *Erkenntnis*, *68*, 277–298.

Bassett, R. (2015). A critique of benchmark theory. *Synthese*, *192*(1), 241–267. https://doi.org/10.1007/s11229-014-0566-3

Briggs, R. (2010). Decision-theoretic paradoxes as voting paradoxes. *Philosophical Review*, *119*(1), 1–30.

Cantwell, J. (2010). On an alleged counter-example to causal decision theory. *Synthese*, *173*(2), 127–152.

Cartwright, N. (1979). Causal laws and effective strategies. *Noûs*, *4*, 419–437.

Edgington, D. (2011). Conditionals, causation, and decision. *Analytic Philosophy*, *52*(2), 75–87.

Egan, A. (2007). Some counterexamples to causal decision theory. *Philosophical Review*, *116*(1), 93–114.

Gibbard, A., & Harper, W. (1978). Counterfactuals and two kinds of expected utility. In C. A. Hooker, J. J. Leach, & E. F. McClennen (Eds.), *Foundations and applications of decision theory* (pp. 125–162). D. Reidel Publishing Company.

Jeffrey, R. C. (1981). The logic of decision defended. *Synthese*, *48*(3), 473–492.

Jeffrey, R. C. (1990). *The logic of decision*. University of Chicago Press.

Joyce, J. M. (1999). *The foundations of causal decision theory*. Cambridge University Press.

Joyce, J. M. (2002). Levi on causal decision theory and the possibility of predicting one's own actions. *Philosophical Studies*, *110*, 69–102.

Lewis, D. (1981). Causal decision theory. *Australasian Journal of Philosophy*, *59*(1), 5–30.

Pittard, J. (2016). Evil and God's toxin puzzle. *Noûs*, *52*(1), 88–108. https://doi.org/10.1111/nous.12154

Skyrms, B. (1980). The role of causal factors in rational decision. In *Causal necessity*. Yale University Press.

Skyrms, B. (1982). Causal decision theory. *Journal of Philosophy*, *79*, 695–711.

Sobel, J. H. (1978). *Probability, chance and choice: A theory of rational agency*. In At a workshop on Pragmatism and Conditionals at the University of Western Ontario.

Sobel, J. H. (1986). Notes on decision theory: Old wine in new bottles. *Australasian Journal of Philosophy*, *64*, 407–437.

Wedgwood, R. (2013). Gandalf's solution to the Newcomb problem. *Synthese*, *190*(14), 2643–2675.

Williamson, T. L. (2019). Causal decision theory is safe from psychopaths. *Erkenntnis*, *86*, 665–685. https://doi.org/10.1007/s10670-019-00125-2


**How to cite this article:** Mohajeri, M. (2023). Conditional causal decision theory reduces to evidential decision theory. *Analytic Philosophy*, 1–14. https://doi.org/10.1111/phib.12298