Overdetermination in Inuitive

Causal Decision Theory

Esteban Céspedes

Causal decision theory defines a rational action as the one that tends to cause the best outcomes. If we adopt counterfactual or probabilistic theories of causation, then we may face problems in overdetermination cases. Do such problems affect Causal decision theory? The aim of this work is to show that the concept of causation that has been fundamental in all versions of causal decision theory is not the most intuitive one. Since overdetermination poses problems for a counterfactual theory of causation, one can think that a version of causal decision theory based on counterfactual dependence may also be vulnerable to such scenarios. However, only when an intuitive, not analyzed notion of causation is presupposed as a ground for a more plausible version of causal decision theory, overdetermination turns problematic. The first interesting consequence of this is that there are more reasons to dismiss traditional theories of causation (and to accept others). The second is to confirm that traditional causal decision theory is not based on our intuitive concept of the causal relation.

1. Overdetermination and two theories of causation

Overdetermination cases present problems to a diverse range of causal notions. The particularity of overdetermination is the fact that there are two or more causes that are sufficient to produce a single event. Some of the classical examples of symmetric and asymmetric overdetermination are the following (Lewis 1973, 2000; Hall 2004).

Symmetric overdetermination

Suzy and Billy are throwing rocks at a bottle. Two stones, one thrown by Suzy and the other by Billy, hit the bottle at the same time and it shatters. Anyhow, either of both throws would have been enough to shatter the bottle without the other.

The next is the case of asymmetric overdetermination, also called *preemption*, and its difference from symmetric cases lies, normally though not necessarily, on the temporal order of the two causes, i.e. Billy and Suzy throwing rocks, and on the interruption of one by the other.

Asymmetric overdetermination

Suzy and Billy are throwing rocks at a bottle. One of Suzy's rocks hits the bottle first, breaking it. Billy's stone would have hit the bottle later, if Suzy's throw had not broken the bottle.

The ways in which Billy's stone might have hit the bottle can change depending on the time at which his throw is prevented to hit the bottle. Thus, there are cases of early and late preemption.

Early asymmetric overdetermination

Suzy throws a stone at a bottle, breaking it. Billy was prepared to throw his stone, but he did not throw it. He would have thrown his stone, if Suzy had given him a signal, meaning that she would not throw. In such case, his throw would have broken the bottle instead.

Late asymmetric overdetermination

Suzy and Billy are throwing rocks at a bottle. One of Suzy's rocks hits the bottle first, breaking it into many glass pieces. Billy's stone flies a bit later exactly over the spot where the bottle was.

It has been argued that such a distinction between early and late asymmetric overdetermination is not necessarily a fundamental one (Hall 2004: 236), but I am not going to get into that debate. There is another, non-temporal kind of asymmetry implied in overdetermination by trumping (Lewis 2000, Stone 2009).

Trumping asymmetric overdetermination

The officer and the sergeant simultaneously shout at the soldiers to advance. The soldiers advance, but only following the officer's order, which seems to be the cause of the fact that the soldiers start marching.

Let us now see why these are taken to be problems. A counterfactual theory of causation, specially, suffers from this and the above described overdetermination scenarios.

Counterfactual theory of causation

One event is caused by another if and only if the former depends causally on the latter. An event *e* depends causally on a distinct event *c*, if and only if both events occur and if *c* had not occurred, *e* would not have occurred.

Now, symmetric overdetermination presents a problem to this theory, because although it seems plausible to affirm that Suzy's throw was a cause of the bottle's breaking, it is not true that the breaking would not have happened without Suzy's throw. The different types of asymmetric overdetermination are problematic for similar reasons.

In facing difficulties with overdetermination in general, probabilistic theories of causation are not an exception.

Probabilistic theory of causation

An event *c* causes a distinct event *e*, if and only if the probability that *e* occurs, given the fact that *c* occurs, is higher than the probability that *e* occurs, given *c*'s absence. Expressed formally, $P(E|C) > P(E|-C)$.

Suppose a scenario of late asymmetric overdetermination, in which Suzy only throws if she sees that Billy is going to throw his rock as well. Suppose also that she always throws faster than him. Before the game starts, it would seem to be true that Billy's throw counts as a cause for the breaking of the bottle. For his throw raises the probability of the breaking. But it is actually Suzy's throw that breaks the bottle. It is thus a case of probabilistic influence without causation.

Symmetric overdetermination cases are also problematic. Suppose that both Suzy and Billy throw simultaneously and that, somehow, seeing other people throwing stones at the same time disconcentrates them. Such distraction lowers the probability of the shattering. Suzy's throws raise the probability of breaking the bottle when Billy is not round throwing rocks at the same time. Her throw is a cause of the breaking. But if Billy also throws, the probability that the bottle shatters, given her throw, will be very low. Again, overdetermination changes the causal scenario in a considerable way. While asymmetric overdetermination presents a case of probabity-raising without causation, symmetric ovedetermination illustrates causation without probability-raising.

Other accounts of causation, like regularity and dispositionalist theories, are also affected by overdetermination (see Lewis 1973, Maslen 2012 and Hitchcock 2013). Given that overdetermination affects somehow a general notion of causation, it might also affect theories in which the causal concept plays an important role. One particular theory that has such characteristic is causal decision theory. In general, such theory recommends performing actions that tend to cause valuable outcomes. I will now explore a case that might be problematic to this theory.

2. Decisional Overdetermination

Before getting into some details of causal decision theory, I will describe a decision scenario in which overdetermination is present, called *decisional overdetermination*. Here are the different types:

Decisional symmetric overdetermination

An agent, Suzy, whose only interest is that the bottle in front of her breaks, has two options. She can either throw a stone at the bottle (T) or omit the throw (-T). She knows that if she throws the stone, Billy's stone and her stone will break the bottle simultaneously. Should she throw the stone at the bottle?

Decisional asymmetric overdetermination

The agent can either throw or abstain of throwing. Billy's stone will break the bottle in a second instance anyway, if she does not throw. Should she throw?

Two types of decisional asymmetric overdetermination can also be distinguished depending on whether the interruption occurs *early* or *late*. *Decisional trumping* scenarios may also be considered. Suppose that the officer wants the bottle to be shattered. He can either shout at the soldiers to break the bottle, knowing that the sergeant will also shout at the same time, or stay silent. The sergeant will shout the order anyway. Should the officer shout to the soldiers? I am not going to analyze decisional trumping cases in depth here. I assume that, for any account of expected utility, the recommendation in decisional trumping cases will be the same as in decisional symmetric overdetermination and decisional late preemption. The reason for this assumption is not necessarily the simultaneity of the potential causes, but, more fundamentally, the causal independence between both, which is implied by simultaneity. Thus, the relevant distinction will be between decisional symmetric overdetermination and decisional asymmetric overdetermination. However, it might also be relevant to distinguish the latter in decisional early and late preemption.

Here is the outcome's matrix for all cases of decisional overdetermination, which contrast the values of the bottle breaking (B) with the values of the bottle remaining unbroken (-B). As usual, the rows represent the agent's possible actions, while the columns represent the different outcomes:

	B	**-B**
T	1	0
-T	1	0

Before evaluating decisional overdetermination from the point of view of causal decision theory, let us see whether the classical version of evidential decision theory recommends the rational thing to do. I assume that every account of rational decision is based on the maximization of some notion of expected utility. Thus, when defining some theory of decision I might leave that unmentioned and only provide the definition of expected utility that characterizes it.

Evidential Decision Theory

The most rational option is to perform the action that has the higher expected utility. This is defined as the sum of the products of the outcome's conditional probabilities, given that the action is performed, and the values of those outcomes.

$EU(A) = \Sigma_j P(O_j|A)V(AO_j)$

We can easily see that in a situation of symmetric decisional overdetermination, the theory recommends indifference. On the one side, the expected utility of throwing is near to one, if the probability that the bottle breaks, given the agent's throw, is also extremely high, which seems natural.

$EU(T) = \Sigma P_j(B_j|T)V(TB_j) = 1$

$EU(-T) = \Sigma P_j(B_j|-T)V(-TB_j) = 1$

On the other side, the expected utility of not throwing is also near to one. For the probability that the bottle breaks, given that the agent does not throw, is also very high; Billy will surely break the bottle with his stone.

In the case of asymmetric decisional overdetermination the expected utilities are also equal. The bottle is going to be shattered by Billy anyway and evidential decision theory recommends indifference again. It must be noticed that this verdict is the same in cases of decisional *early* preemption as in cases of decisional *late* preemption. The conditional probabilities will be equally high in both cases.

Let us now consider causal decision theory, starting with one of its first and most general versions.

> **Causal Decision Theory**
>
> Gibbard and Harper (1978) developed a notion of expected utility based on a solution made by Stalnaker (1972), according to which conditional probabilities are no longer crucial. Instead, probabilities of conditionals should be considered. The particular sort of conditionals are counterfactuals (□→). Thus, the expected utility of an act is defined as the sum of the products of the probability that the different outcomes would occur, if the action was performed, with the values of those outcomes:
>
> $EU(A) = \Sigma_j P(A \,\Box\!\!\rightarrow O_j)V(AO_j)$

Other accounts of causal decision theory were developed, prescribing the same results as Gibbard and Harper's version (Skyrms 1980, Lewis 1981). However, I am going to work with the one defined above. How does this theory confront decisional overdetermination? In the case of *symmetric decisional overdetermination*, causal decision theory recommends indifference as well. Both the expected utility of throwing and of not throwing approximate to one, because the bottle would break anyway, either by the agent's throw or her not throwing. Billy is throwing the stone at the same time the agent throws hers, which means that the bottle will surely break, even if the agent does not throw. *Asymmetric decisional overdetermination* does not make a difference in the expected values. The bottle would be broken by Billy's stone, if the agent decided not to throw. Thus, the counterfactual holds. Again, as well as in evidential decision theory, causal decision theory based on counterfactuals does not distinguish between early and late preemption. Indifference is recommended in both cases. Such result is already described by Lewis (1981) with the notion of *independence hypothesis*.

3. Intuitive Causal Decision Theory

It must be questioned whether the same results are obtained in a version of causal decision theory based on a simple, unanalyzed concept of causation. Instead of asking whether certain outcomes would take place, if a given action was performed, the agent should simply ask whether the action he is evaluating will *cause* the outcomes considered. Ignoring conditional probability, fundamental for evidential decision theory, or the probability of a conditional that could merely point out and perhaps inspire a plausible theory of causation, we have the option to take our most natural and general notion of the causal relation into account.

Intuitive Causal Decision Theory

Given an intuitive causal relation, the expected utility of an action is defined as the sum of the products of the probabilities that a certain outcome will be *caused* by that action, with these outcome's values. The following is the formal definition:

$EU(A) = \Sigma_j\, P(A\ \textit{causes}\ O_j)V(AO_j)$

In simple scenarios, this kind of expected utility delivers the same verdicts as causal decision theory and, even, as evidential decision theory. For instance, if I want to break a bottle (B) by smashing it with a hammer (H), the three kinds of probabilities about the breaking given my action will have, at a first moment, equal degrees, that is, $P(B|H) = P(H \Box\!\!\rightarrow B) = P(H\ \textit{causes}\ B)$ and $P(B|-H) = P(-H \Box\!\!\rightarrow B) = P(-H\ \textit{causes}\ B)$. It also tends to deliver the same verdict as causal decision theory in Newcomb's problem, but I will not discuss this here.

In other situations of simple everyday decision making involving so-called *causation by omission,* intuitive causal decision theory might give the right advice for the wrong reason. Should someone water his plant? The expected utility of watering the plant will be high, given the fact that such action will cause the plant to be healthy. The expected utility of not watering the plant will be much lower. But why is it so? It depends on whether we support a theory that accounts for negative causation or not. If negative causation is real, then one might say that to omit watering the plant would probably cause the plant's death. But if we believe that negative events do not cause anything and prefer to describe the plant's possible death as a product of other positive events like, for example, the complex stress produced in the roots by a very dry soil, then we would not say that to omit watering the plant will cause its death. In this latter interpretation, it is not the omission itself but the conjunct of its *consequences* that causes the undesirable outcome. That is, to water the plant will not be extremely preferable to omitting the watering. Interestingly, the expected utility of letting the plant without water will be different in both cases, depending on our assumptions about negative causation. If we argue that intuitive decision theory recommends watering the plant, the argument would lie on the wrong reasons, namely that the omission of watering the plant can itself be the cause of its death. I will avoid the debate on negative causation, because I think that it is a point where the intuition goes too far (see Beebee 2004 or Dowe 2009). However, negative causation is going to be crucial for an elaborated version of intuitive causal decision theory and will be considered in our further discussion below.

My main interest now is to see whether intuitive causal decision theory gives the right advice in cases of decisional overdetermination, and negative causation might play an important role here. First, in asymmetric decisional overdetermination the probability that the agent's throw *causes* the breaking is near to certainty and thus, the expected utility of throwing will approach to the value of the fact that the bottle is broken. It must now be asked whether it is likely that the agent's *not throwing* can cause the bottle to shatter. Exactly on this point and in contrast to evidential and traditional causal decision theory, this version of expected utility tends to distinguish between early and late decisional preemption scenarios. In the case of early asymmetric overdetermination, the omission of throwing must lead to the signal that communicates to Billy that he can throw. Billy's throw would then cause the breaking. But would we say that the absence of the agent's throw *causes* the bottle to break? Are we committed to say that absences and omissions can cause something? Again, this topic is not going to be discussed here. Anyway, the signal sending can help to grasp the decision scenario in such a way that one can avoid the debate. In a case of decisional *early* preemption, the agent's options are throwing a stone or sending a signal to Billy (and not throwing). We would not doubt in saying that the agent's signal causes Billy's throw and that Billy's throw causes the bottle to break. By the transitivity of causation, the breaking of the bottle would be caused in both cases by the agent's action. Therefore, the expected utility of throwing equals the expected utility of not throwing, making it indifferent to Suzy, whether she throws or not. This is the same verdict delivered by a traditional counterfactual-based causal decision theory.

In cases of decisional *late* preemption, the probabilities are different. For there is no signal that Suzy sends to Billy and his throw does not depend causally on what the agent does. But the fact that the bottle is not hit by the agent's stone could depend causally, according to our intuitive notion of causation, on the fact that she does not throw. Obviously, this is only plausible under the assumption that omissions and negative events can cause and be caused by others. On the one side, if omissions are considered to be possible causes, then it is possible to think that Suzy's omission of throwing causes the shattering by letting Billy's rock hit the bottle. Thus, intuitive causal decision theory recommends indifference in the scenario of decisional late preemption. On the other side, if omissions are not taken to have causal power or are not considered as possible causal relata, then the theory recommends to throw. This seems to be the wrong advice, since the interests of the agent are focused on the breaking of the bottle, no matter how or by whom it is broken.

The results at a symmetric decisional overdetermination scenario resemble the results of decisional late preemption. In this situation, the agent and Billy throw their stones simultaneously and with the same force. It is assumed that both stones are also going to hit the bottle at the same time. The probability that the agent's throw causes the bottle to break must be high, although it will not shatter the bottle alone. But what is the probability that her omission to throw her rock *causes* the bottle to break? Under a very general and natural notion of causation, it is a very low one, for we would not say that the agent's not throwing can *cause* the bottle to break. Her action is completely independent from Billy's throw, given that both events are simultaneous. Thus, the expected utility of not throwing approaches zero and, clearly, is much lower than the expected utility of throwing. It seems that causal decision theory recommends wrongly again. I am going to clarify later how this unexpected result can be rationalized. As was already mentioned, intuitive causal decision theory will recommend the same in decisional trumping as in decisional symmetric overdetermination.

There are cases where intuitive causal decision theory might give simply an irrational advice. Suppose that Suzy is deliberating about throwing the stone at a bottle or not in a scenario of decisional asymmetric overdetermination. This time she knows that she has a bad run at throwing rocks and it is unlikely that she will hit the bottle if she throws. But she also knows that it is very likely that Billy hits it if he throws. Suppose further that Billy will not throw if she does. Since it is less probable that the omission of the throw *causes* Billy's rock to break the bottle, the expected utility of omitting the throw will be lower than the expected utility of throwing.

EU(T) = P(T *causes* B)V(TB)
EU(-T) = P(-T *causes* B)V(-TB)
EU (T) > EU(-T)

However, if we take the definition of expected utility used by evidential decision theory, the inequality favors not throwing:

EU(T) = P(T|B)V(TB)
EU(-T) = P(-T|B)V(-TB)
EU (T) < EU(-T)

The probability that the bottle breaks given the agent's throw is lower than without her throw, because Billy, who is in a good run, will almost surely break it. This result should be the same in the framework of traditional causal decision theory. Thus, evidential decision theory and traditional causal decision theory recommend omitting the throw and letting Billy break the bottle, while intuitive causal decision theory recommends, maybe irrationally, to throw.

Let me go back and take an overview of our results. The following table shows the different results that the theories considered until now give in decisional overdetermination scenarios:

	Symmetric overdetermination	**Asymmetric overdetermination**
		Early - ***Late***
EVT	indifference	Indifference - indifference
CDT	indifference	indifference - indifference
ICDT	throw	indifference - throw (-NC)

Recapitulating and describing the table, we can say that decisional overdetermination cases do not present differences between traditional accounts of evidential (EDT) and causal decision theory (CDT). Nevertheless, there is a clear difference between intuitive causal decision theory (ICDT) and the other two mentioned theories, especially in scenarios of decisional symmetric overdetermination. Another distinction occurs in cases of decisional asymmetric overdetermination. While both evidential and traditional causal decision theories do not recommend differently in cases of early and late preemption, intuitive causal decision theory distinguishes those variations. For, on the one hand, the two traditional theories prescribe indifference, no matter whether it is a scenario of early or late preemption. On the other hand, intuitive causal decision theory prescribes indifference only in decisional early preemption, recommending throwing in cases that involve late asymmetric overdetermination, if negative causation (NC) is not assumed literally.

4. Contrastive Causation

It is not hard to find other theories that might be compatible in some cases with the results given by traditional accounts. Take, for instance, a definition of expected utility based on a *contrastive notion of causation* (see Hitchcock 1996, Schaffer 2005, Northcott 2008), where $EU(A) = \Sigma_j P(A$ *rather than* C *causes* $O_j)$ and C, a set of possible causes of the considered outcome different from the agent's action and its effects. For every action A, the consequences of its omission, i.e. of -A, are included in C. One can argue on whether the contrast class must be built around both the cause and the effect (Schaffer 2005) or just around the cause (Hitchcock 1996). For purposes of rational decision, only the side of the cause, i.e. of the action, is contrastive explicitly. The contrast class of the effect is already given by the outcomes' partition used in the definition of expected utility.

How do we interpret the contrast between different possible causes? The fact that, for instance, a hurricane rather than an earthquake caused the house to collapse simply means that the hurricane caused the collapse and the earthquake did not. Now, the fact that the earthquake did not cause the collapse either means that an earthquake occurred that was not strong enough (nor sufficient in any other sense) to cause the collapse or that there wasn't any earthquake at all. The specificity of the contrast class may be determined by the agent's interests.

The recommendation under such account will be, at least at decisional early preemption scenarios, the same as in traditional causal decision theory. For it seems to be very likely that the agent's throw rather than Billy's throw causes the bottle shattering. On the other hand, it is highly probable that her omission of throwing and the sending of the signal rather than throwing cause the breaking. Indifference is, thus, recommended. In decisional late preemption, this definition of expected utility may recommend to throw, depending on our assumptions about the intuitive concept of the causal relation.

At decisional symmetric overdetermination, it seems implausible to think that the agent's throw *rather* than Billy's throw can break the bottle, assuming that Billy will throw anyway. It is also very unlikely that the agent's omission of the throw and its consequences rather than throwing can cause the breaking of the bottle. That is, the theory would recommend indifference. But this is so because both options have very low expected values, which is a very different reason than the one that takes traditional accounts to the same results. This is the same result of a decision theory based on possible causation, discussed below. I am not going to discuss further details about how an approach of *contrastive causal decision theory* could be elaborated, and I will assume that the recommendations such a theory would give depend on parameters that are either crucial for traditional accounts of rational decision (e.g. the definition of the contrast class) or for intuitive causal decision theory (e.g. the definition of the causal relation).

5. Does intuitive causal decision theory get the right notion of causation?

As we have seen, a traditional version of causal decision theory is not actually based on causation, but on the counterfactual relation between options and possible outcomes. Hitchcock (2013) examines the vulnerability of traditional causal decision theory confronting overdetermination scenarios, concluding, as it was shown above, that such cases are far from being problematic.

Instead of simply saying that causal decision theory should not be considered as an account based on the notion of causation, Hitchcock's diagnosis is that the notion on which the theory is based is just a different one. A traditional account of causal decision theory is not based on a concept of *actual causation* supported by the counterfactual analysis. Instead, it uses the notion of the causal relation implied by the counterfactual conditional A □→ O, called *causal dependence*. The main difference that Hitchcock presents between actual causation and causal dependence is the fact that the former is a retrospective notion, while the latter is a prospective one. That is, when one evaluates whether *c* causes *e* using actual causation, one refers to past events. Both events *c* and *e* already occurred and if *c* had not occurred, *e* would not have occurred. But when we evaluate whether *c* causes *e* using causal dependence, we think of future counterfactual events. Both events *c* and *e* have not occurred yet and if *c* occurred, then *e* would occur.

I think that Hitchcock's distinction makes the case more understandable. But understandability does neither imply accuracy nor relevance. After all, every agent should assume that the actions he is deliberating about have not occurred yet. Otherwise he would just be wasting his time giving very extreme probabilities to the different possible outcomes. Another reason to rethink the relevance of the distinction between retrospective and prospective causation is to ask whether temporal precedence is a necessary assumption for the causal analysis or whether the notion of time must be included at all as a condition in the definition of the causal relation. The distinction between actual causation and causal dependence assumes temporal order, but I think that one of the virtues of the counterfactual analysis of causation is the way it handles with temporal asymmetry without presupposing it. If the conditional -C □→ -E is true, the conditional -E □→ -C might be false. The non-backtracking feature of these counterfactuals clarifies the precedence of the cause without postulating it. Why should we do this when we face decisional overdetermination scenarios, if these look harmless to our traditional account of causal decision theory?

Somehow, decisional overdetermination demands something from the theory. My diagnosis is, thus, double. Either traditional causal decision theory is not affected by decisional overdetermination, but we should strengthen our notion of causation to explain that, which shows that decisional overdetermination has some serious consequences after all, or traditional causal decision theory is indeed directly affected by such scenarios. In both cases, the causal notion involved in causal decision theory must be reconsidered.

6. Possible causation in expected utility

As I have already shown, while decisional overdetermination seems not to be problematic for traditional causal decision theory, it is only problematic for a causal theory of decision based on some intuitive concept of causality. I have not said much about that concept, nor is it necessary to say much, since it must be a mere product of our most general intuitions. But is not the fact that the effect would not have happened without the cause a feature of the causal relation that fits our most general understanding of it? That is actually another virtue of the counterfactual account of causation. Let us consider a notion of expected utility construed with probabilities of causal facts between actions and outcomes, as well as in our intuitive version of causal decision theory, but using the counterfactual analysis of causation. In order to do that we cannot just take the probability of the conditional -A □→ -O, because the counterfactual analysis of causation also presupposes that both events involved in the relation already occurred. Thus, we may have to face the following problematic definition, involving a would-cause counterfactual:

$$EU(A) = \Sigma_j P[A \,\&\, O_j \;\Box\!\!\rightarrow (-A \;\Box\!\!\rightarrow -O_j)]V(AO_j)$$

The agent must evaluate the probability of the fact that if he had performed some action and some outcome occurred, then that outcome would not have occurred without his action. The counterfactual considered turns out to be vacuously true, for, on the base of the importation principle (McGee 1985), the antecedent of the equivalent conditional (O&A&-A)□→-O is impossible. One should actually construct a definition of expected utility based on this latter conditional, but this is not going to be considered here, as the results will be the same. Since the validity of the importation principle can be put in doubt under specific assumptions (see Arló-Costa's work on epistemic conditionals [2001]), I am not going to take it here as a strong and uncontroversial assumption.

There are plenty of common cases where expressions of embedded conditionals of the form A □→ -A□→ -B do make sense. I think that one could make perfectly sense of that using traditional possible world semantics for counterfactuals (either Stalnaker's 1968 or Lewis's 1979). The relevant changes that such a semantic should confront in order to make sense of embedded counterfactuals are precisions regarding the notion of laws of nature and their violations. I am not going into these details now (see Dowe 2009). I will assume, however, that a counterfactual conditional is true, if and only if in the closest world (or in any of the closest worlds) where the antecedent is true, the consequent is also true.

For example, if I had a dry match, then, if it was not dry, it would not light. In the actual world, w_1, seeing a match box with a single wet match in it, I wonder what would happen if it was a dry match. Thinking about the constitution of the closest world (or worlds) where it is a dry match, I wonder whether the match would light, if it was not dry. In order to do this, I have to speculate, this time from w_2, about the closest world in which the (counterpart of or the same) match is not dry. The hasty response would be to think that such a world is the actual world. But that is not necessarily so. For the physical changes that distinguish the actual world from w_2 might be simpler than the changes that would take us from w_2 back to the actual world. It might be that by drying the match, the surroundings were also dried and to consider a world where the match is not dry would imply to change the entire environment. In such a case, the closest non-A-world is not the actual world. Hence, the iterated counterfactual of the form A □→ -A □→ B cannot be converted to a conditional with a contradictory antecedent. The occurrence of A, the fact that the match is dry, is assumed from w_1, but its negation is not. The occurrence of the fact that the match is not dry, is assumed from the perspective of w_2. This means that A & -A are not assumed in w_1 either. The closest world to w_2 in which the match is not dry might be different from the actual and since it is true that the match won't light there, the expression with the embedded counterfactual makes perfect sense:

(The match in the box is dry)□→(The match is not dry)□→(The match does not light)

That is, if the match in the box was not dry, it would not light. Let us clarify the evaluation of the expressions involved from the perspective of the corresponding possible worlds, assuming, just to make the case clearer, that for every world there is only one closest possible world:

w_1: -A	The match is not dry.
w_2: A	The match is dry.
w_3: -A & B	The match is not dry and it does not light.
w_1: A □→ -A □→ B	If the match was dry, then if it was not dry, it would not light.

Again, the closest possible world to the actual world, in which A is true, is the possible world w_2. Nevertheless, in w_2, the closest possible world where A is false is not the actual world, but w_3. In that possible world, B is also the case. Thus, the iterated counterfactual conditional is true in the actual world.

Let us now consider the recommendations for decisional overdetermination by a causal decision theory based on *possible causation* with embedded counterfactuals supported by a would-cause semantics (see Dowe 2009). In the case of *decisional symmetric overdetermination*, the expected utility of throwing is as low as the expected utility of omitting the throw. For, on the one side, given that the agent throws and the bottle shatters, the bottle would still have shattered, if she had not thrown. On the other side, given that the agent did not throw and the bottle shattered, it would have shattered, if she had thrown. The theory recommends indifference, which fits to the traditional account of causal decision theory, but differs from the intuitive approach. In decisional asymmetric overdetermination scenarios, the theory prescribes indifference again.

The account based on possible causation gives the same recommendations as traditional causal decision theory. However, the reasons for this are, as in contrastive causal decision theory, extremely different for each theory. According to traditional causal decision theory, the bottle would break, no matter what the agent did, i.e. the probability that the bottle shatters is for both options very high. But for a decision theory based on possible causation, none of the possible actions would cause the breaking. The bottle might break, when Suzy throws, but the throw cannot count as a cause. The bottle would still have shattered, if the throw had not occurred.

We have a new reason to think that counterfactuals do not give an intuitive notion of causation for a theory of rational decision, neither used for definitions of causal dependence, nor for actual causation. As it has been mentioned, this also shows that a general theory of causation that pretends to grasp our intuitive concepts of that relation cannot be based only on counterfactual dependence.

7. Ranking theory and decisional ovedetermination

There is a definition of expected utility that might, apparently, share some results with intuitive causal decision theory. I will consider such a definition with the ranking functional notion of causation at its basis (Spohn 2006, 2012). The ranking theoretic approach of causation proposes a clear solution to overdetermination cases that, instead of postulating fine-grained events (one of the most plausible solutions to that problem), suggests considering a *conceptual* fine-graining. Put in a very informal way, the variable C is a *direct cause* of variable E, if and only if both C and E occur, C precedes E and C is a *reason* for E, given the obtaining circumstances (Spohn 2012b: 354). Taking a negative ranking function κ that measures the degrees of disbelief for propositions, such that $\kappa(A) = 0$ expresses that A is believed, the belief in C is a reason for E if and only if $\kappa(\text{-}E|C) > \kappa(\text{-}E|\text{-}C)$, i.e. if the absence of E has a higher degree of disbelief, given C's occurrence, than given its absence.

Using a positive ranking function β, such that $\beta(A) = \kappa(\text{-}A)$, and a two-sided function τ, such that $\tau(A) = \beta(A) - \kappa(A)$, overdetermination cases can be described with clarity. In cases of symmetric overdetermination, the belief that the bottle breaks (B), given the fact that both Suzy (T_1) and Billy (T_2) threw their stones, is higher than the belief on the breaking, given only one throw. That is, for instance, $\tau(B|T_1 \& T_2) = 2 > \tau(B|T_1 \& \text{-}T_2) = 1 = \tau(B|\text{-}T_1 \& T_2) > \tau(B|\text{-}T_1 \& \text{-}T_2) = \text{-}1$. Clearly, it is highly disbelieved that the bottle breaks, given the fact that none of them throws a stone. Cases of early asymmetric overdetermination are clarified simply by appealing to the transitivity of the causal chain going from the actual cause to the effect, arguing that although the shattering does not depend counterfactually on Suzy's throw, it depends on some state of the stone before it hits the bottle and after the instant at which Suzy would have given the signal to communicate that she was not going to throw. Cases of late preemption are more problematic and find hardly an interpretation on this theory (Spohn 2012: 367).

To avoid getting into the debate about whether asymmetric overdetermination is really a problem for causation or not, let us go directly to our topic, namely decision theory, and see what this framework can do about the scenarios we are struggling with. I am going to consider a definition of expected utility based on the positive ranking function:

$EU(A) = \Sigma_j\, \beta(O_j|A)V(AO_j)$

A theory of decision that takes rational actions as the ones that maximize such notion of expected utility into account recommends the usual in decisional early preemption scenarios, namely indifference. For the fact that the agent's rock breaks the bottle does not have to be more believed than the fact that Billy's rock breaks it. Billy will not throw, if the agent does, so the degree of belief on the bottle shattering must not be too high. Let us say that $\beta(B|T) = \beta(B|\text{-}T) = 1$. I am not going to consider decisional late preemption for the already mentioned reason.

The scenario of decisional symmetric overdetermination is more interesting, since the account of expected utility based on ranking functions recommends the same as intuitive causal decision theory. For it is more believed that the bottle shatters (S), given the fact that both the agent and Billy throw their stones, than just given Billy's throw (B), i.e. $\beta(S|T \& B) > \beta(S|\text{-}T \& B)$. Therefore, the expected utility of throwing (T) is higher than the expected utility of omitting the throw. This case is also an example of one of the most important differences between ranking theory and probabilistic causation, considering that $P(S|T \& B) =$

P(S|-T & B) and assuming that both the agent and Billy are well-trained stone-throwers. As expected, ranking causal decision theory also recommends throwing in decisional trumping. It also delivers the right verdict in the bad run scenario described above, in which intuitive causal decision theory fails.

In this way, the ranking account of causation does not only analyze cases of redundant causation correctly; it can also ground a notion of expected utility that adjusts to intuitive causal decision theory. The fact that a decision theory based on ranking functions is more similar to a decision theory based on the intuitive concept of causation than to other accounts that give different recommendations in particular scenarios means that the ranking theoretic framework of causation fits better our general intuitions about the causal relation.

Other accounts of causation, different from the ranking functional approach, might also fulfill the requirements for intuitive causal decision theory. Some of these are the causal modeling approach (Pearl 2000; Spirtes, Glymour & Sheines 2000), the theory of causal processes (Salmon 1984, Dowe 2000) and the dispositionalist theory of causation (Harré & Madden 1975, Mumford & Anjum 2011). There are not enough reasons to think that these accounts of causation exclude each other in a fundamental and relevant way. I will leave aside this time the discussion of such differences and on how these theories may support intuitive causal decision theory.

8. Is there irrationality in the intuition?

Now that it has been explained how traditional accounts of decision making do not fit our intuitive notion of causation and that intuitive causal decision theory (and some specific accounts supporting it) might serve as an alternative, it must be put in doubt whether the intuitive account is really a practical guide for rational decision. Is it really rational to throw a stone at the bottle in a scenario of decisional symmetric overdetermination? Apparently, the recommendations given by the traditional accounts are right in saying that it does not matter, because Billy will also throw and the bottle will surely get broken, which is the agent's only interest. The decisional overdetermination scenarios described above do not consider the energy lost by throwing stones, which would be a reason to twist the agent's indifference. Suzy should not throw in such cases. Anyway, these situations, far from being problematic, do not worry us now.

There are two reasons that could explain why the prescriptions given by a theory based on an intuitive notion of causation tend to the option of throwing, rather than omitting the throw. The first one, which is clearly captured in ranking causal decision theory, is the fact that two throws ensure the shattering of the bottle more than just one. Even if we assume that it is very likely that nothing will go wrong, a rational agent should not dismiss the mere possibility that something can go wrong. If just one thing goes wrong when both Suzy and Billy are supposed to throw the stones, the bottle will still break. But if just one thing goes wrong after Suzy decides to omit the throw, leaving the responsibility to Billy, the bottle won't shatter. If Suzy does throw, then two things have to fail to expect that the bottle won't break.

The second reason is the very realistic and practical fact that agents weight their beliefs about causal facts according to the *confidence* they give to the possible causes considered. If a decisional scenario ends up in indifference, the agent might evaluate the potential causes of the desired outcome in terms of confidence. To understand the difference between this notion and the degree of belief under which an agent evaluates the act's possible causal relation with the outcomes, it is important to notice that the former only applies directly to the other possible causes, which are different from the evaluated action itself. We can hardly give probabilities to actions (Spohn 1977, Levi 1986); we just suppose what they would cause, if they were performed. But we can assign probabilities to alternative causes of the relevant outcomes. Confidence is expressed by such degrees of belief on the alternative causes and on the impossibility to attach probabilities to actions. Somehow, in decisional symmetric overdetermination scenarios, any rational agent will choose the actions that are related causally to his desired outcomes, in spite the fact that traditional versions of expected utility may initially promote indifference. If Billy and I are both great at smashing bottles, why should I put more confidence on his throw than on mine? Of course, this self-confident attitude must not always be assumed; it should be just considered as a strategic (and thus, rational) tie-breaking move in cases of initial indifference.

I have shown that decisional overdetermination is not problematic for traditional causal decision theory. Nevertheless, what seems to be threatening by such kind of scenarios is the notion of causation on which traditional definitions of expected utility are based. A version of expected utility based on an intuitive notion of causation differs from traditional accounts. This does not only confirm that the fundamental concepts considered in traditional accounts of expected utility, for example, counterfactual dependence, are far from a transparent definition of causation. It also suggests that a fundamental and general theory of causation should be searched with more independence from those concepts. Furthermore, a theory that tells us how to act rationally should not only consider the importance of an agent's causal influence on the relevant outcomes, but also be based on a notion of causation that clearly explains what is usually understood as a causal relation and what causation really is. An intuitive approach may be a crucial point in reconsiderations of the foundations of causal decision theory, although it should not be the last step.

Esteban Céspedes
J. W. Goethe-Universität Frankfurt a.M.
e.cespedes@stud.uni-frankfurt.de

Acknowledgements

I would like to thank André Fuhrmann for numerous illuminating discussions on this topic. This work was also supported by CONICYT and by the Konrad Adenauer Foundation.

References

Arló-Costa, H. 2001: 'Bayesian epistemology and epistemic conditionals: On the status of the export-import laws', *Journal of Philosophy* 98, 555-593.
Beebee, H. 2004: 'Causing and Nothingness', in Paul, Hall & Collins (ed.): *Causation and Counterfactuals*. MIT Press.
Dowe, P. 2000: *Physical Causation*. Cambridge University Press.
Dowe, P. 2009: 'Would-cause semantics', *Philosophy of Science* 76, 701-711.
Gibbard, A. & Harper. W. 1978: 'Counterfactuals and Two Kinds of Expected Utility', in Hooker, Leach & McClennen (ed.): *Foundations and Applications of Decision Theory*. Dordrecht: Reidel.
Hall, N. 2004: 'Two Concepts of Causation', in Hall, Paul, & Collins (eds.): *Causation and Counterfactuals*. Cambridge: MIT Press.
Harré, R. & Madden, E. H. (1975). *Causal Powers*. Oxford: Blackwell.
Hitchcock, C. 1996: 'The role of contrast in causal and explanatory claims', *Synthese* 107 , 395 - 419.
Hitchcock, C. 2013: 'What is the 'Cause' in Causal Decision Theory?', Erkenntnis.
Levi, I. 1986: *Hard Choices.* Cambridge University Press.
Lewis, D. 1973: 'Causation'. *Journal of Philosophy,* 70, 556-567.
Lewis, D. 1979: 'Counterfactual dependence and time's arrow', *Noûs* 13, 455-476.
Lewis, D. 1981: 'Causal Decision Theory', *Australasian Journal of Philosophy* 59, 5 – 30.
Lewis, D. 2000: 'Causation as influence', *Journal of Philosophy* 97, 182-197.
Maslen, C. 2012: 'Regularity Accounts of Causation and the Problem of Pre-emption: Dark Prospects Indeed', *Erkenntnis* 77, 419-434.
McGee , V. 1985: 'Counterexample to Modus Ponens', *Journal of Philosophy* 82, 462-471.
Mumford, S. & Anjum, R. L. 2011: *Getting Causes from Powers*. Oxford University Press.
Northcott, R. 2008: 'Causation and contrast classes', *Philosophical Studies* 139, 111 - 123.
Nozick, R. 1970: 'Newcomb's Problem and Two Principles of Choice', in Rescher (ed.): *Essays in Honor of Carl G. Hempel*. Dordrecht: Reidel.
Pearl, J. 2000: *Causality: Models, Reasoning, and Inference*. Cambridge University Press.
Salmon, W. 1984: *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.
Schaffer, J. 2005: 'Contrastive causation', *Philosophical Review* 114, 327-358.
Skyrms, B. 1980: *Causal Necessity*. Yale University Press.
Spirtes, Glymour & Sheines. 2000: Causation, Prediction, and Search. MIT Press.
Spohn, W. 1977: 'Where Luce and Krantz do really generalize Savage's decision model', *Erkenntnis* 11, 113 - 134.
Spohn, W. 2006: 'Causation: An alternative', *British Journal for the Philosophy of Science* 57, 93-119.
Spohn , W. 2012a: 'Reversing 30 years of discussion: why causal decision theorists should one-box', *Synthese* 187, 95-122.
Spohn, W. 2012b: *The Laws of Belief*. Oxford University Press.
Stalnaker, R. 1972: 'Letter to David Lewis', in Harper, Stalnaker, and Pearce. *Ifs: Conditionals, Belief, Decision, Chance, and Time.* Dordrecht: Reidel.
Stalnaker, R. 1968: 'A Theory of Conditionals', in Rescher. *Studies in Logical Theory.* Oxford: Blackwell.
Stone, J. 2009: 'Trumping the causal influence account of causation', *Philosophical Studies* 142, 153-160.