Universiteit Utrecht

Master Wijsbegeerte

15 ECTS

Causal One-boxing as a novel solution to Newcomb's paradox

Author:
Rosa Sterkenburg
5576296

First supervisor:
Dr. Janneke van Lith
Second supervisor:
Dr. Sander Beckers

July 10, 2018

Universiteit Utrecht

Contents

1 Introduction

> "Suppose a being in whose power to predict your choices you have enormous confidence. (One might tell a science-fiction story about a being from another planet, with advanced technology and science, who you know to be friendly, etc.) You know that this being has often correctly predicted your choices in the past (and has never, so far as you know, made an incorrect prediction about your choices), and furthermore you know that this being has often correctly predicted the choices of other people, many of whom are similar to you, in the particular situation to be described below. One might tell a longer story, but all this leads you to believe that almost certainly this being's prediction about your choice in the situation to be discussed will be correct." [Nozick, 1969, p. 114]

This is the first paragraph of Nozick's famous paper about Newcomb's paradox. Following his description of the predictor, there is a choice. You are presented with two boxes, a transparent and an opaque one. You either get to take both boxes, or just the opaque one. The transparent box will contain $1000,-, no matter what. The other box will contain $1.000.000,- if the being predicted that you would take only that box, and it will be empty if the being predicted you would take both boxes. What is the rational choice to make?

There are two arguments here, seemingly equally persuasive but giving contradictory recommendations. The first argument says that since you cannot influence the contents of the boxes as they were filled before you were presented with this choice, you should take both boxes. That way you will always get a thousand dollars more than when you would take only the opaque box. The other argument says that since you know the being's predictions are almost always accurate, taking just the opaque box will make it very probable that the box contains a million, while taking both boxes makes it very probable that the box is empty. Therefore, you should only take the opaque box.

In the literature these two possible strategies are called one-boxing and two-boxing, and both strategies are associated with a branch of formal decision theory. One-boxing is usually advocated by Evidential Decision Theory. EDT roughly states that a rational decision should be based on the available evidence. Two-boxing is associated with Causal Decision Theory. CDT states that rational decisions should be based on causal considerations, and one should choose the act that has the most desirable effects.

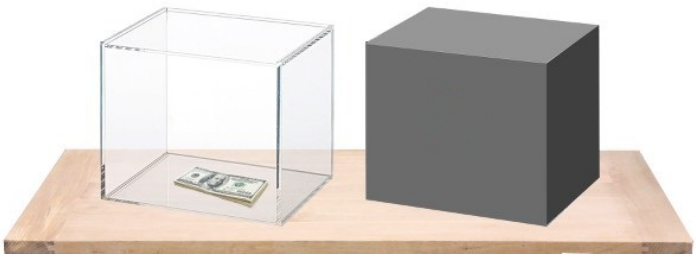

Figure 1: Schematic representation of Newcomb's paradox

In most normal situations EDT and CDT coincide: they recommend the same act as the most rational choice to make. However, in the case of Newcomb's paradox, they yield different answers. Since only one act can be the rational thing to do, this means a choice needs to be made between CDT and EDT.

In problems like this, usually intuition is the guideline used to choose between two opposing theories. In this case, however, people's intuitions seem to be equally divided between the two options. Some feel two-boxing is obviously the way to go, and see Newcomb's paradox as a clear counterexample to EDT, while others strongly feel that one-boxing is the best strategy and causal considerations are not all they're made out to be. Nozick writes:

> "To almost everyone it is perfectly clear and obvious what should be done. The difficulty is that these people seem to divide almost evenly on the problem, with large numbers thinking that the opposing half is just being silly." [Nozick, 1969, p.117]

However, there are also examples in the literature of philosophers whose intuitions are not so clear cut. In a paper about the short-comings of an important principle of CDT, for example, Bar-Hillel and Margalit write:

> "One is left with the uneasy feeling that choosing A_1 [one-boxing], though defensible on game-theoretic grounds, is somehow 'wrong' in a very fundamental way. That it is, in fact, tantamount to subscribing to backwards causality." [Bar-Hillel & Margalit, 1972, p. 299]

This shows that even though they argue for one-boxing, Bar-Hillel and Margalit think causal considerations cannot be ignored.

Another example of uneasiness can be found with Arntzenius. In a paper that advocates CDT and two-boxing, he writes:

> "In a Newcomb type case evidential decision theorists will, on average, end up richer than causal decision theorists. Moreover, it is not as if this is a surprise: evidential and causal decision theorists can foresee that this will happen." [Arntzenius, 2008, p.289]

He argues that the rational choice is to two-box, because causal considerations should always be counted as most important, but he feels forced to accept that the best way to get rich is to one-box.

These two examples illustrate that there is room for another view on Newcomb's paradox. One-boxing is the way to get rich, but causal considerations are too important to decision theory to just simply ignore. Newcomb's paradox prompts us to develop a causal theory of decision that advocates one-boxing.

This thesis discusses such a causal theory of one-boxing: EviCausalism. EviCausalism was introduced by Huw Price in 2012, and this thesis closely follows Price's work. The added value of this thesis is to provide a concrete application of Price's theory to Newcomb's paradox. While Price provides an extensive justification for his theory, his examples of how his theory should be applied are rather vague. This thesis describes not just the framework, but also specifies the concrete probabilities needed to tackle Newcomb's paradox.

In chapter 2 some preliminary concerns will be discussed. I will consider the role of (questionable) intuitions, since intuitions are an important part of most of the literature. Furthermore two questions that arise when defining Newcomb's paradox will be addressed. These are the problem of free will and questions about the infallibility of the predictor.

Chapter 3 will consider formal decision theory. The mathematics needed to formalize EDT and CDT are explained, and Lewisian CDT is discussed in detail. Lewisian CDT is chosen because it can be formalized in a way very similar to EDT, thereby clearly emphasizing the difference between the two theories.

In Chapter 4 Lewis's views on decision theory are compared with his ideas about chance and credence. There is considerable tension between these views, and this tension gives rise to the proposal of EviCausalism.

The final chapter considers possible criticism to EviCausalism. First we consider whether the classical counterexample to EDT, that of the smoking lesion, is also problematic for EviCausalism. Next we turn to backwards causation, and finally we discuss a potentially problematic discontinuity that is present in EviCausalism.

2 Preliminary Concerns

Before we can properly discuss causal one-boxing, there are some other issues that need to be addressed. The first two are questions that arise often when people are confronted with Newcomb's paradox for the first time. The first question is whether the predictor is infallible, or just very good. For a lot of people this seems to make a crucial difference. The other question is how this paradox relates to free will. Finally, an important issue about methodology needs to be addressed, namely the role of intuitions in arguments concerning Newcomb's paradox.

2.1 The infallible predictor

The first question that often arises when people are first confronted with Newcomb's problem is whether the predictor is certainly always right, or just right most of the time.

In the introduction we have seen that causal considerations lead to the conclusion that you should two-box in Newcomb's paradox. Another argument to arrive at that conclusion is the dominance argument. The dominance argument states that if an action is better than another action in all possible scenarios, that action is the rational one and should be acted upon. For Newcomb's paradox, the dominant action is two-boxing, since either there is a million in the opaque box, or there is not. In both cases you get $1000,- more if you two-box, where it is implicit that your choice does not cause the contents of the opaque box. Therefore, in both possible scenario's, two-boxing is better and therefore dominant.

A lot of people have the intuition that you should two-box in case of a fallible predictor, adhering to the dominance argument. In case of an infallible predictor, the dominance argument loses its appeal. It no longer feels like two-boxing will give you $1000,- more, regardless of what was predicted. If you know for certain that the predictor would be right, two-boxing will get you $999.000,- less, and the best choice suddenly becomes to one-box.

> "The dominance argument becomes less attractive if we suppose that the predictor is definitely right" [...] "Given an infallible predictor, many people prefer to one-box." [Ahmed, 2015, p. 263]

The preferred strategy for Newcomb's paradox would thus be to two-box in case of a fallible (but very good) predictor, and to one-box in case of an infallible predictor. Ahmed [2015] dubs this the discontinuous strategy.

The discontinuous strategy is motivated by what Ahmed calls the Certainty Principle. The certainty principle states that if you know for certain that A will result if you choose option 1 and B will result if you choose option 2, and A is preferable over B, then option 1 is rationally superior to option 2. The discontinuous strategy then arises from standard dominance reasoning if the predictor is infallible, and adherence to the certainty principle when he is not.

Now Ahmed continues to formulate the Newcomb problem in two different but equivalent ways. In one of these formulations the certainty principle applies, and therefore you play the discontinuous

strategy, and one-box in the case of an infallible predictor. In the other formulation there is no certainty, and therefore you always have to two-box because of standard dominance reasoning. The recommendation of the discontinuous strategy thus depends on the formulation of the problem, and therefore it cannot be a rational strategy.

To resolve this problem there are two options. Either you reject the certainty principle, and then the discontinuous strategy will lose all motivation. You will always use standard dominance reasoning and therefore endorse two-boxing regardless whether the predictor is fallible. The other option is to make the certainty principle apply in a wider range, so that in both formulations you'll play the same strategy. However, this will make your strategy indistinguishable from EDT.

Both these solutions remove the discontinuity that arose when the predictor was certainly correct. This means that it is irrelevant whether or not the predictor is infallible or just very good, since the rational strategy (whether you think this is one-boxing or two-boxing) remains the same in both cases. Therefore, we do not need to concern ourselves with questions about the fallibility of the predictor.

2.2 Newcomb's problem and free will

Another question that often arises is whether it is possible to have such a capable predictor as the one supposed in Newcomb's paradox and to have free will at the same time. The intuition is that if someone can almost always predict correctly what you will do when faced with the choice between one or two boxes, then your choice cannot be free. Presumably, if your choice can be predicted, it is already known in advance, and this robs you of the possibility to act otherwise. And if you could not have acted otherwise, we cannot speak of a free choice.

If this were the case, it would provide a trivial (but unsatisfactory) answer to Newcomb's paradox. If the predictor and free will are incompatible, there are two possible cases. Either we have free will and there cannot exist such a predictor, or we do not have free will but the predictor can exist. In the first case it is impossible to properly formulate Newcomb's paradox, since the predictor cannot exist, so the problem posed by Newcomb's paradox never arises. In the second case, we do not have free will and therefore are never faced with a true decision. This renders decision theory obsolete.

To show that Newcomb's paradox is not so easily defeated is not a simple task. It requires a firm view on what it is precisely to have free will, and how this would be effected by the existence of a very capable predictor. This means it requires an uncontroversial answer to the free will debate, i.e. the conflict between the notions of freedom and determinism.

A possible answer of that kind is compatibilism. Compatabilist theories of free will state that free will and determinism are compatible, or, in our case, that free will and the very capable predictor are compatible. While this theory is far from uncontroversial, for our purposes it is enough that it exists and has real advocates. It is thus possible that the free will debate is resolved in a way that still allows for Newcomb's paradox to arise.

2.3 The role of intuitions

The last issue that needs to be addressed before we can turn to causal one-boxing is one of methodology. In the above sections, intuitions were used multiple times in arguments, and they are also very commonplace in the literature. However, concerning Newcomb's paradox, not everyone has the same intuitions. Therefore, it is not obvious that intuitions can play a genuine role in an objective argument.

The role often fulfilled by intuitions in philosophical inquiry is that of a counterexample. When a definition is proposed, a hypothetical case is constructed that intuitively is a clear instance of the defined concept, but does not fit the proposed definition. Or a case is given that intuitively clearly is not an instance of the concept, but it does fit the proposed definition. The intuition is used as evidence to provide a refutation.

It seems important that intuitions that are used in this way must be felt strongly by everyone, and are beyond any doubt. An intuition that is disputed, that is felt by some but not by others, cannot count as convincing evidence.

This is precisely the problem when using intuitions with respect to Newcomb's paradox. Some feel that Newcomb's paradox is a clear counterexample to EDT, while others think EDT gives the right advise in Newcomb's paradox. The main problem is that we do not agree on what Newcomb's paradox teaches us about decision theory.

Bales says:

> "However, relying on controversial intuitions in strange cases strikes me as a dangerously overused move so I will instead champion a different approach." [Bales, 2015, p. 1500]

The different approach championed by Bales is showing that his argument can be used for all the different intuitions people have. This way he can arrive at a conclusion independently of which intuition is correct. Unfortunately, this is a strategy that can only be used in a limited amount of cases, and will not get us much further in finding a solution to Newcomb's paradox.

The aim of this thesis is to counter the intuitive problem with EDT, namely that it ignores the causality of the situation. This is only relevant to people who already have the intuition that one-boxing is the rational choice. For them the proposal for causal one-boxing might remove some of the uneasiness they feel about the causal aspect of Newcomb's paradox. It makes one-boxing more plausible for those who already think one-boxing is the way to go. Within this group intuitions are mostly shared, and can therefore be safely used in arguments or counterexamples.

However, this will not convince dedicated two-boxers, since their intuition regarding Newcomb's paradox fundamentally conflict with those of one-boxers. In his famous paper on causal decision theory Lewis states:

> "I will not enter into debate with [the evidentialists], since that debate is hopelessly deadlocked and I have nothing new to add to it. Rather, I address myself to those who join me in presupposing that Newcomb problems show the need for some sort of causal decision theory, and in asking what form that theory should take." [Lewis, 1981, p. 5]

Similarly, this thesis does not enter into debate with convinced two-boxers. It just tries to make the position of one-boxers more coherent by adding causal considerations to their theory.

3 Formal decision theory

With the preliminary concerns dealt with, we need to turn to EDT and CDT and their formal differences. While a formalization of EDT is quite straightforward, CDT has been formalized in many different ways. In this chapter the causal decision theory as developed by David Lewis [1981] will be discussed, because of its similarity in formalization to EDT.

3.1 Common Ground between EDT and CDT

To formalize EDT and CDT, we need a lot of common features. Both theories assume that a rational agent has a credence function and a value function. These functions are defined over the set of possible worlds, Ω. The credence function gives for every possible world $W \in \Omega$ the degree of belief of the agent that world W is the actual world. The value function gives for every possible world how desirable the agent thinks it is that W is the actual world.

Definition 3.1.

$$C(W) : \Omega \to [0, 1]$$
$$V(W) : \Omega \to \mathbb{R}$$

We go on to define a proposition as a set of worlds. We say a proposition holds at just those worlds that are its members. The credence of a proposition $X \subseteq \Omega$ is found by summing the credences of all the worlds in the proposition.

Definition 3.2.

$$C(X) = \sum_{W \in X} C(W)$$

Conditional credence is defined in the way familiar from probability theory:

Definition 3.3.

$$C(X|Y) = \frac{C(X \wedge Y)}{C(Y)}$$

The value of a proposition is found by a weighted average:

Definition 3.4.

$$V(X) = \sum_{W \in X} C(W|X) \cdot V(W) = \sum_{W \in X} \frac{C(W) \cdot V(W)}{C(X)}$$

The value of a proposition is the same as what we would normally call the expected value. The value of each world is multiplied by how likely it is that that world is the actual world.

Finally, we introduce the idea of partitions:

Definition 3.5. *A partition is a set of propositions of which exactly one holds at any world.*

A partition for example could consists of the two propositions "Brazil will win the soccer world cup in 2018" and "Brazil will not win the soccer world cup in 2018". There are no possible worlds where both are true, and there are no possible worlds where neither is true. Therefore exactly one of these propositions holds at any world, precisely per the definition of partition. Furthermore, the definition of a partition ensures that the conjunction of the partition equals Ω, the set of all possible worlds.

Let P be a partition and let Z be a variable ranging over that partition, i.e. a variable that can take the propositions in the partition as its values. Since the propositions in a partition together include all the possible worlds, and no possible world is in more than one proposition of a partition, we get the following rules of additivity:

$$C(X) = \sum_{Z \in P} C(X \wedge Z)$$

$$C(X)V(X) = \sum_{Z \in P} C(X \wedge Z) \cdot V(X \wedge Z)$$

From this last rule we can derive an alternative definition of expected value:

$$V(X) = \sum_{Z \in P} C(Z|X) \cdot V(X \wedge Z)$$

Even though a partition is needed in the calculation, the expected value is independent of the partition used. Any partition would give a proposition the same value.

3.2 Evidential Decision Theory

With all the groundwork done, EDT is now easily defined. The rule is simply to maximize (expected) value.

It is not hard to see that this leads to one-boxing in Newcomb's paradox. Take the value of a world to be the amount of money you get in that world. Furthermore, take your credence in a correct prediction to be 0.99. The partition to be used is {million in opaque box, no million in opaque box}.

$$\begin{aligned} V(\text{one-boxing}) &= C(\text{million in opaque box} \mid \text{one-boxing}) \cdot V(\text{million in opaque box}) \\ &\quad + C(\text{no million in opaque box} \mid \text{one-boxing}) \cdot V(\text{no million in opaque box}) \\ &= 0.99 \cdot 1,000,000 + 0.01 \cdot 0 = 990,000 \end{aligned}$$

$$
\begin{aligned}
V(\text{two-boxing}) &= C(\text{million in opaque box} \mid \text{two-boxing}) \cdot V(\text{million in opaque box}) \\
&\quad + C(\text{no million in opaque box} \mid \text{two-boxing}) \cdot V(\text{no million in opaque box}) \\
&= 0.01 \cdot 1,001,000 + 0.99 \cdot 1000 &&= 11,000
\end{aligned}
$$

This shows that by maximizing of the expected value, also called V-maximizing, EDT recommends one-boxing as the rational choice.

3.3 Lewisian Causal Decision Theory

Philosophers like Lewis [1981] and Skyrms [1982] argue for a causal decision theory that prescribes two-boxing as the rational strategy. V-maximizing as prescribed by EDT cannot be a correct decision theory, they say, because it leads to one-boxing. Furthermore, they agree that V-maximizing goes wrong by not taking the causal structure of the world into account. By using conditional credences, V-maximizing uses all the available evidence, whether this evidence is causally relevant or not. There are multiple takes on how to fix this, and therefore there are multiple versions of CDT. In this section we will discuss CDT as introduced by David Lewis in *Causal Decision Theory* [1981].

First we can identify a problem with standard dominance reasoning that might also be a problem for formal causal decision theories. The problem is that it is not always the same action that is dominant. Nozick formulates it as follows:

> "It may be that relative to one partition of the states of the world, one action A dominates another, whereas relative to another partition of the states of the world, it does not." [Nozick, 1969, p. 119]

This means we need to take special care when choosing our partition. If we cannot properly justify the partition used in our causal decision theory, the recommendations of the theory become meaningless, since a different partition would result in a different recommendation.

The solve this problem Lewis introduces dependency hypotheses. A dependency hypothesis is a maximally specific proposition about the causal structure of the world. More notably, it specifies how the things one cares about when deliberating causally depend on the possible actions.

For example say that I'm deliberating whether or not to make myself a nice cup of coffee before going to work. One possible dependency hypothesis could state that making a cup of coffee will cause me to miss the bus, and therefore be late. Another dependency hypothesis could say that making a cup of coffee will not cause me to be late (I won't miss the bus). In both hypotheses the coffee will cause me to be happier and have a more productive morning. There are many things I don't know, like whether the bus is running on schedule, or the precise amount of time it will take me to drink the coffee. Therefore, when deliberating, I am not completely sure which of the dependency hypotheses holds.

By definition one cannot causally influence which dependency hypothesis holds in the actual world. The dependency hypotheses specify what you can influence by your actions, and are maximally specific. If you could influence which dependency hypothesis is true, than apparently your hypotheses are not yet maximally specific since there is something more that you can influence, and you should augment your hypotheses. If they are maximally specific, everything you *can* influence is described in the hypotheses, and which hypothesis is true is thus something you cannot influence. This means that if in one dependency hypothesis your actions always have nice results, there is nothing you can do to make sure that that is the dependency hypothesis that holds.

However, if your actions can provide evidence for one dependency hypothesis or another, than V-maximizing could lead you to the nicest dependency hypothesis. However, since there is nothing you can do to bring about the nice dependencies, this is merely providing yourself with "good news", as Lewis puts it. He states:

> "Failures of V-maximizing appear only if, first, you are sensible enough to spread your credence over several dependency hypotheses, and second, your actions might be evidence for some dependency hypotheses and against others." [Lewis, 1981, p. 11]

If maximizing expected value is not the way to go, then how should we evaluate our options in the case that we have spread our credence over different dependency hypotheses? Lewis says:

> "You should consider the expected value of your options under the several hypotheses; you should weight these by the credences you attach to the hypotheses; and you should maximize the weighted average" [Lewis, 1981, p. 11]

Note that dependency hypotheses form a partition: in every possible world, there must be some causal structure, and only one causal structure can be actual in any given world. We thus have a partition based on the causal structure of the world, and not just some arbitrary partition. With this partition we can formalize the recipe above in the following way:

Definition 3.6. *Let H be the set of dependency hypotheses and let the variable K range over this set. We define expected utility of an option A by:*

$$U(A) = \sum_{K \in H} C(K) \cdot V(A \wedge K)$$

Now Lewis's CDT simply prescribes the rule of U-maximizing, instead of V-maximizing.

If we take the set of dependency hypotheses as our partition we can rewrite the formula for V as:

$$V(A) = \sum_{K \in H} C(K|A) \cdot V(A \wedge K)$$

We see the formula for expected value is almost identical for the one for expected utility. The only difference is the use of $C(K)$ instead of $C(K|A)$. Lewis says about this:

> "It is essential to define utility as we did using unconditional credences $C(K)$ of dependency hypotheses, not their conditional credence $C(K|A)$. If the two differ, any difference expresses exactly that news-bearing aspect of the options that we meant to suppress." [Lewis, 1981, p. 12]

Lewisian CDT has thus two major features. First, it explicitly takes the causal structure of the world into account by using a partition of dependency hypotheses. Second, it makes sure evidential considerations that are not causal do not get to play a role by using unconditional probabilities rather than conditional ones.

4 Price's Evicausalism

In *Causation, Chance and the Rational Significance of Supernatural Evidence* Price [2012] discusses a tension in Lewis's views about Newcomb-like problems. He then uses this tension as an argument to set aside Lewisian CDT, and to introduce his own concept of causation and thereby his own version of CDT.

4.1 The Principal Principle

In the previous section we have seen Lewisian CDT, but to understand the tension in Lewis's views we need another one of Lewis's ideas, that of the principal principle.

The principal principle links objective chance to credence. It tells us that rational credence ought to follow objective chance in all normal situations. The only exception is when we have inadmissible evidence.

> "Let C be any reasonable initial credence function. Let t be any time. Let x be any real number in the unit interval. Let X be the proposition that the chance, at time t, of A's holding equals x. Let E be any proposition compatible with X that is admissible at time t. Then $C(A|X \wedge E) = x$ " [Lewis, 1980, p. 272]

For example the proposition X could be that the chance of a fair coin landing heads right now equals 0.5. In this case t = right now, A = the landing on heads and $x = 0.5$. Let E be any evidence we have. If E is admissible, the principal principle tells we should have credence x in A, so we should have credence 0.5 that a fair coin lands on heads.

Inadmissible evidence is any information about the outcome of a chance process. This can be fairly innocent. For example, it could be that the coin has already been tossed, and someone tells you that the outcome was heads. Inadmissible evidence can also be strange information about the outcome of a future chance process, like a text from the future, or a vision you trust, or any other story that provides information about the future and that you have good reason to believe.

The principal principle only tells us what our credences should be when there is no inadmissible evidence, and gives no recommendations for when we do have inadmissible evidence. However, elsewhere Lewis states:

> "Since E is inadmissible, the Principal Principle does not apply. The fatal move [...] is no better than the obvious blunder: C(the coin will fall heads | it is fair and will fall heads in 99 of the next 100 tosses) = 1/2" [Lewis, 1994, p. 485]

And in *A Subjectivist's Guide to Objective Chance* Lewis gives the following example:

> "Next question. As before, except that now it is afternoon and you have evidence that became available after the coin was tossed at noon. [...] the witness has told you that it fell heads in nine out of ten tosses of which the noon toss was one. You remain as sure as ever that the chance of heads, just before noon, was 50%. To what degree should you believe that the coin tossed at noon fell heads? Answer. Not 50%, but something not far short of 100%." [Lewis, 1980, p. 270]

Together these suggest that in case of inadmissible evidence, rational credence should go with the evidence. This means that if you have access to a reliable crystal ball, and this tells you the outcome of a future coin toss is heads, your credence in that coin coming up heads should be 1.

Of course, inadmissible evidence in the form of reliable information about the future is very rare, or perhaps does not exist. However, the only point here is that when an agent has good reason to believe he has inadmissible evidence, rational credence should go with the evidence and not with the objective chance.

4.2 Tension in Lewis's views

Now we have all the concepts available to show the tension in Lewis's views. To do this we follow Price [2012] in introducing two Newcomb-like games, that should have the same rational strategy, but get different recommendations from Lewis.

Suppose you are offered to place a free bet with the following pay-out table:

	Heads	Tails
Bet heads	\$100,-	\$0,-
Bet tails	\$0,-	\$50,-

It is a good bet either way, but the rational choice would be to bet Heads. However, suppose you are told that the coin is fair, but for some reason, in all the cases that you actually place a bet the chance of the coin coming up tails is 99%. You are told this by your best friend who has access to a reliable crystal ball. That means, this is precisely the kind of information about the outcome of chance process that Lewis dubs inadmissible information. Now, with this extra information, what is the rational choice?

If we want to maximize profit, intuitively we look at the expected value of a bet. For heads this is $C(heads) \cdot 100$ and for tails it is $C(tails) \cdot 50$. The main question is thus what the credences should be. According to the Lewis we should take the inadmissible evidence into account, so the credences become $C(heads) = 0.01$ and $C(tails) = 0.99$. With these credences the rational choice is to bet tails.

We could also use Lewis's CDT with dependency hypotheses to show this same result. The result

of the coin toss does not depend on which bet we place. Our possible actions therefore do not have very shocking results. No matter what we do, the result will be the same. We do however have two dependency hypotheses, because there are two possible results.

K_{heads}: The causal structure of the world is such that the coin will land heads
K_{tails}: The causal structure of the world is such that the coin will land tails

Remember: $U(A) = \sum_{K \in H} C(K) \cdot V(A \wedge K)$

$$\begin{aligned} \textit{U(Bet heads)} &= \mathrm{C}(K_{heads}) \cdot \mathrm{V}(\text{Bet heads} \wedge K_{heads}) + \mathrm{C}(K_{tails}) \cdot \mathrm{V}(\text{Bet heads} \wedge K_{tails}) \\ \textit{U(Bet tails)} &= \mathrm{C}(K_{heads}) \cdot \mathrm{V}(\text{Bet tails} \wedge K_{heads}) + \mathrm{C}(K_{tails}) \cdot \mathrm{V}(\text{Bet tails} \wedge K_{tails}) \end{aligned}$$

Again, the credences we use have to take the inadmissible evidence into account, so we get:

$$\begin{aligned} \textit{U(Bet heads)} &= 0.01 \cdot 100 + 0.99 \cdot 0 = 1 \\ \textit{U(Bet tails)} &= 0.01 \cdot 0 + 0.99 \cdot 50 = 49{,}50 \end{aligned}$$

So we see that also in the complete formalism of CDT the conclusion is that the rational choice in this case is to bet tails.

Next we add an extra option to the game; you now also have the possibility not to bet at all. Your new pay-off table becomes:

	Heads	Tails
Bet heads	\$100,-	\$0,-
Bet tails	\$0,-	\$50,-
No bet	\$0,-	\$0,-

It should be clear that this extra option does not change the rational policy. We certainly want to place a bet, it's a free chance to win money, so we will never go for the option of not betting. Since it will never be chosen, the extra option should not be able to change rational policy.

However, the extra option of not betting does have the effect that the inadmissible information only becomes available when you place a bet. Remember, we were told that in all the cases *that you actually place a bet* the chance of the coin coming up tails is 99%.

We can formulate the same dependency hypotheses as in the previous case. However, this time the credence we have in these hypotheses must be 0.5, since we know the coin is fair and the information about its particular behaviour in the cases that you bet does not get to play a role.

U(Bet heads)	$= 0.5 \cdot 100$	$+ 0.5 \cdot 0$	$= 50$
U(Bet tails)	$= 0.5 \cdot 0$	$+ 0.5 \cdot 50$	$= 25$

We see that adding in a clearly flawed option leads CDT to give a different recommendation, which does not bode well for CDT. The obvious remedy would be to use the credences conditional on placing a bet, since that way the inadmissible information is available for use. However, when defining CDT, Lewis has stressed that if the conditional and the unconditional credences differ, this is precisely because of the news-bearing aspect we have to suppress. Therefore, when calculating utility, Lewis says we should always use the objective, unconditional credences.

Cases like these thus show a tension between Lewis's policy on inadmissible evidence on one hand, and his insistence on using unconditional credences on the other. Lewis is aware of these cases and calls them much more problematic for decision theory than Newcomb's paradox. Furthermore he notices that these cases only occur in exceptional situations when someone has knowledge about the future, and that it might not matter so much if CDT is incapable of dealing with them. Price summarizes:

> "We thus have a class of Newcomb-like problems in which Lewis's policy on inadmissible evidence concurs with EDT; and in which CDT escapes defeat only by withdrawing from the field." [Price, 2012, p. 18]

4.3 EviCausalism

As we have seen, the tension in Lewis's views is between his views on inadmissible evidence and his focus on unconditional probabilities. They give different recommendations in cases that are so similar they should have the same rational policy. There are two simple solutions; either abandon the policy on inadmissible evidence, or move to conditional credences when calculating utility.

Lewis is willing to do neither. Price, however, advocates the move to using conditional credences. As we have seen the only difference between EDT and CDT was in the use of unconditional credences, so in a way we are back at EDT.

Fortunately, Price does not simply propose to go back to EDT. He proposes a new understanding of causation to justify the use of unconditional credences, thereby unifying causal consideration with the calculations of EDT. He calls his decision theory EviCausalism, because it has both evidential and causal features. EviCausalism is not based on a definition of causation itself, but of the related concept of causal dependency.

Definition 4.1. *"B is causally dependent on A just in case an expert agent would take* $P(B|A) \neq P(B)$*, in a calculation of the V-utility of bringing it about that A (in circumstances in which the agent is not indifferent to whether B)" [Price, 2012, p. 21]*

First, lets illustrate this definition with an example. Is whether or not it's going to rain today causally dependent on you bringing an umbrella? To determine this we need to compare the chances P(It will rain today) and P(It will rain today | You brought an umbrella). The expert

mentioned in the definition is a theoretical being who knows all the relevant facts and how they relate to each other. Now, would this expert take these chances to be different? Hopefully the reader agrees that he would not, and therefore we can conclude that whether it rains today or not is not causally dependent on whether you bring your umbrella.

We could also look at this example the other way round, and ask whether you bring your umbrella is causally dependent on the rain that is coming this afternoon. The relevant chances are now P(You bring your umbrella) and P(You bring your umbrella | It's raining this afternoon). In this case an expert would take these chances to be different, and therefore in this case we can speak of causal dependence.

Definition 4.1 ties the concepts of decision making and causation closely together. The reason we are considering these particular chances is to use them in a calculation of V-utility, after all. At first sight this might seem strange, and even circular. To say that things are causally dependent because we take them to be relevant when making a decision is the wrong way round. We think things should be relevant when making a decision just because they are causally relevant.

Price, however, considers this close tie between decision making and causation not as weakness but as a virtue. He states that with a completely objective understanding of causation, it is unclear why causal considerations should be relevant to decision making. He calls this the missing link. By connecting causation explicitly to decision making one circumvents this problem.

Another remarkable aspect of Price's definition of causal dependency is its subjectivity. It seems strange that something as basic as causation should depend on agents to be defined. Price puts the objection as follows:

> "Readers may baulk here at the subjectivism of this proposal. "Were there no causal dependencies before there were agents?," " [Price, 2012, p. 25]

To answer to this objection Prices once again draws a parallel with Lewis:

> "Indeed there were, just as according to Lewis there were chances before there were any creatures with credences. Nevertheless, Lewis holds that we cannot properly characterize chance unless we do so in terms of credence - unless we say, in effect, that information about chance *is* information about credence. The EviCausalist says the same about causal dependence." [Price, 2012, p. 25]

Price does not mean to claim objective causation does not exist, merely that causation also has a subjective side that is crucial to understanding causation. For a human agent causation cannot be understood without taking into account the effects it has for deliberation, just as chance cannot be understood without also considering how chances affect credences.

EviCausalism applied to Newcomb's paradox

In Newcomb's paradox, the usual assumption is that the contents of the opaque box do not depend causally on your choice since the box was filled *before* you made the choice. Underlying this

assumption is the usual temporal direction of causation. Price's definition of causal dependency does not mention any temporal direction, and allows for the possibility that the contents of the opaque box do in fact depend causally on the choice made.

If we apply Prices definition to Newcomb's paradox, the relevant chances to determine causal dependency are P(Million in the opaque box | One-boxing) and P(Million in the opaque box). We know from the description of the paradox that the first of these is really high, and usually this is quantified as 0.99. The unconditional probability P(Million in the opaque box) is, unfortunately, never mentioned.

However, we can use the fact that {one-boxing, two-boxing} is a partition to calculate P(Million in the opaque box).

$$\begin{aligned} P(\text{million in the opaque box}) &= P(\text{million in the opaque box} \mid \text{one-boxing}) \cdot P(\text{one-boxing}) \\ &\quad + P(\text{million in the opaque box} \mid \text{two-boxing}) \cdot P(\text{two-boxing}) \\ &= 0.99 \cdot P(\text{one-boxing}) + 0.01 \cdot P(\text{two-boxing}) \end{aligned}$$

It is reasonable to assume that while someone is still deliberating, P(one-boxing) and P(two-boxing) are both neither 0 or 1. And if P(one-boxing) $\neq$ 1, P(Million in the opaque box) will be smaller than 0.99, hence P(Million in the opaque box) $\neq$ P(Million in the opaque box | One-boxing).

This means that according to the definition of EviCausalism, the contents of the opaque box *are* causally dependent on your action. That means if we calculate expected value in the usual way, our use of C(one-box | million in the opaque box) is not merely a conditional credence to account for an evidential connection. It also reflects a genuine causal connection. Therefore, the old rule of V-maximizing will still give the advise to one-box, but no longer because it ignores causal considerations. By looking at causation in a new way, Price has found a way to defend one-boxing on causal grounds.

Causal dependency as an expert

One final aspect of Price's definition deserves extra clarification; that of the expert. When introducing his own definition, Price writes:

> "The resolution turns on the proposal that causal dependence should be regarded as an analyst expert about the conditional credences required by an evidential decision maker." [Price, 2012, p. 21]

First of, it is very strange to state that causal dependency *is* an expert. I think it makes a lot more sense to consider causal dependency as a guide that any expert should follow. An analyst expert helps to determine the relevance of one proposition to another. In most cases, it is difficult to know which facts might be relevant, and therefore it is difficult to know for an evidential decision maker what conditional credences he should use. Understanding the causal structure of the case at hand is often easier, and is most cases the causal structure is a good guide to the relevant evidential facts.

However, in cases like Newcomb's paradox, causal dependency at first sight does not seem to be a good guide to the relevant evidential facts. What I think Price wants to show by saying that causal dependency is the expert, is that causal dependency can never be wrong. When causal and evidential considerations seem to come apart, this simply means one is mistaken about the causal structure, and one should revise their views on the causal structure. The true causal structure of the case, which is known to an analyst expert, will *always* lead the expert to the relevant evidential facts.

This leads Price to the conclusion that Newcomb cases are not cases where there is something strange about decision theory. They are simply cases where there is something strange about causation. He says:

> "In effect, the EviCausalist proposal is simply that we should take seriously in general the view of causal dependence that is thus forced on us in this particular case." [Price, 2012, p. 22]

5 Possible objections to EviCausalism

5.1 Medical Newcomb cases

As we have seen, Price's EviCausalism provides a way of causal one-boxing. Essentially, in all Newcomb cases it agrees with EDT, only it understands the relevant probabilities in such a way that they become causal. However, if it is so similar to EDT, might it not be susceptible to the same sort of counterexamples?

These counterexamples to EDT are what have become known as the medical Newcomb cases, where the most famous one is the smoking lesion. Suppose there exists a medical condition that causes both the tendency to smoke and to get cancer, and furthermore suppose that smoking does not cause cancer. Because of the common cause, smoking is evidentially related to getting cancer. EDT takes these evidential but non-causal considerations into account, and therefore prescribes abstaining from smoking as the rational thing to do in this case. Since smoking will in no way cause you to get cancer, however, this is clearly the wrong advice.

As before, the diagnosis of the mistake made by EDT is the use of conditional probabilities. In this case, EDT uses C(Getting cancer | You smoke), and because of the correlation this conditional credence is fairly high. However, this credence has nothing to do with smoking causing cancer, and therefore it is irrational to let it play such an important role. In Lewis's words, not smoking would only provide you with good news, but it would do nothing to improve your chances of staying healthy.

Now, as we have seen, Price proposes to understand a difference between P(A) and P(A|B) as causal dependency between A and B. This seems like Price is not only making the same irrational recommendations as EDT, but, even worse, also insists that this *is* a case of causal dependence. He recognizes this problem and states:

> "It would add idiocy to irrationality, surely, to try to justify this recommendation by claiming that causation should be understood in such a way that this agent can *cause herself* to lack the gene." [Price, 2012, p. 23]

The traditional response to the medical case counterargument is the tickle defense. It has served EDT well, and is also useful in the present case. The tickle defense states that by introspection there is extra information available that screens off the evidential relevance of certain factors. In the case of the smoking lesion this means that an agent knows for herself whether she feels the inclination to smoke. If so, this is an indication that she has the gene, but in this case the actual act of Smoking carries no extra information. Smoking is no longer evidentially relevant to getting cancer, and therefore EDT will no longer recommend to refrain from smoking.

Let's repeat Price's definition for causal dependence:

> "B is causally dependent on A just in case an expert agent would take $P(B|A) \neq P(B)$, in a calculation of the V-utility of bringing it about that A (in circumstances in which the agent is not indifferent to whether B)" [Price, 2012, p. 21]

If we apply this definition to the current case, we get: Cancer is causally dependent on Smoking if and only if an expert agent would take P(Cancer | Smoking) ≠ P(Cancer) when calculating the V-utility of smoking when she is trying to decide whether to smoke or not

This agent already knows for herself whether she feels the inclination to smoke, and thus whether or not she is likely to have the unwanted gene. This means that actually deciding to smoke will carry no extra information, and thus when calculating the V-utility of Smoking, P(cancer) should be equal to P(Cancer | Smoking). This means that, because of the tickle, the definition of causal dependence is not met and Smoking does not cause Cancer, as desired. EviCausalism therefore does not recommend refraining from smoking to prevent getting cancer, and is not trapped by the medical case counterexamples.

5.2 Backwards causation

Now that we have seen that EviCausalism can defend itself, we need to turn to its intuitive merits. The reasons we started this discussion after all was an appeal to intuitions. To the intuition, more specifically, that (normal) causal consideration should play a role in decision making. EDT, while giving the correct answer in the Newcomb problem, was not convincing enough because it either ignores causal links, or it amounts to backwards causation. We deemed both undesirable.

EviCausalism, however, amounts to backwards causation since it stresses that there is a real causal link between your choice and the contents of the opaque box, which is filled prior to your choice. Fortunately, EviCausalism does something more than just stipulating backwards causation without further clarification. It offers a justification for backwards causation and a criterion of when backwards causation can apply. Price writes:

> "What EviCausalism adds to this background is a proposal about the nature of causal dependence itself, such that a Newcomb problem cannot *but* be retrocausal, if there is genuine evidential dependence of the Predictor's behaviour on the agent's choice, from the agent's point of view." [Price, 2012, p. 22]

The crucial concept is this genuine dependence. Price does not say we should accept backwards causation in any old case, but only if the agent truly believes in the evidential connection. For example, in a religious formulation of Newcomb's problem, where the predictor is God, we might have genuine evidential dependence, while in a case where the predictor is merely a very skilled psychologist, we do not.

Price introduces a simple test to tell the difference. Suppose we use a random procedure to determine whether you should one-box or two-box. This random choice is made after the prediction. The key question is whether one-boxers still get rich, or, in other words, whether the prediction is still reliable. If in the random case the predictor is still capable of making a correct prediction, then we speak of genuine evidential dependence. Only exceptional Predictors will pass this test. More plausible predictors will fail utterly when they have to predict random choices, and so in those cases we do not have genuine evidential dependence and therefore also no causal dependence.

Dummett (1954) distinguishes three conditions for what he calls retrocausality. Suppose we have two events, E (effect) en C (cause), that consistently co-occur and where E takes place prior to C. For them to qualify as retrocause and effect, E has to be causally undetermined by things that happened before E. Furhermore, E cannot be a normal forward cause of C and we have to be able to explain C without any reference to E.

The Newcomb case meets all three criteria. The prediction is not causally determined by things that happened before the prediction is made, i.e. there is no normal causal chain leading up to the prediction that ensures the predictor is (almost) always right. Furthermore, the prediction does not cause the choice made, since the one who plays the Newcomb game does not know what the prediction is and therefore cannot match their choice to fit the prediction.

What Price's criterion of being able to predict a random choice adds to this, is that it ensures there is no trickery involved. If there were a causal link from prediction to choice, randomizing the choice would break that link. Price thus offers us a way to test if the evidence is really so strong that we have no choice but to accept backwards causation. EviCausalism will only amount to backwards causation if that is the case. In case of Newcomb's paradox that means it only accepts backwards causation in the case of an exceptional predictor that can predict a random choice. In all other cases, there is no genuine evidential dependence, and EviCausalism will prescribe two-boxing.

The bilking argument

The bilking argument as formulated by Black [1956] is an important argument against backwards causation. The bilking argument states that in a presumed case of backwards causation, we could always wait for the effect to happen, and then do something to prevent the cause from happening. If this is impossible, it suggests a normal, forward causal link between effect and cause. If we succeed in preventing the cause, we break the correlation between the events, and we have no longer reason to assume they are causally connected. In both cases there is no backwards causation.

In Newcomb's paradox, however, it is not so easy to apply the bilking strategy. If you get to play the game, and choose one or two boxes, you could wait till the prediction has happened. But how then would you bring about that the prediction is wrong? You could try to act against your natural inclinations, so if you would normally one-box, you now two-box instead. But if the predictor is good enough to predict a random choice, surely the predictor can foresee this plan as well. However, the conclusion that there must be a normal causal chain from the prediction to your choice because bilking proved unsuccessful does not seem warranted.

To see this we can once again use Price's definition of causal dependence. Your choice is causally dependent on the prediction just in case P(your choice | the fact some prediction is made) is different from P(your choice). In cases where you have reason to believe that these probabilities differ, EviCausalism would not have prescribed one-boxing, since those cases would not pass the test of randomizing your choice. This is precisely because in those cases there is a normal, forward causal connection between your choice and the contents of the opaque box.

That means that in the cases it was justified to one-box, it seems reasonable to assume these probabilities are the same, as long as you do not know what the prediction reads. Therefore, the

bilking argument does not work in cases where EviCausalism prescribes one-boxing. We cannot prevent the cause, but we also cannot draw the conclusion that this means there is normal forward causation.

It becomes a different story if you know what it is the predictor thinks you will do, instead of merely just knowing that some prediction has been made. It that case it becomes obvious that you can bilk, you just do the opposite of what you know is predicted.

We noted earlier that it is reasonable to assume that while still deliberating, one cannot know for sure what they will do already. Price calls this the special epistemic status of deliberation. In cases of backwards causation, this means that one cannot know about the effects in the past of what one decides. The same holds when the effects of one's choice lie in the future. If you have knowledge of the future, this knowledge could be bilked. The difference is that usually we assume we do have epistemic access to the past, but not the future.

To speak of deliberation, of a real choice, the effects of this choice, whether they lie in the past or in the future, cannot be known already. If the effects are known, there is no choice and therefore no deliberation. In the cases where the prediction is known, we thus cannot speak of deliberation, and therefore these cases are not relevant to decision theory.

To conclude, there are two possible ways in which you could try to apply the bilking argument. Either you do not know the contents of the prediction, in which cases bilking is impossible. The conclusion that this must mean there is a normal forward causal link between your choice and the contents of the opaque box, however, is not justified. The other option is that you do know what the prediction reads. In this case bilking is possible, but we cannot speak of deliberation. In both cases the bilking argument fails in showing that backwards causation is impossible.

5.3 Discontinuous strategy

Newcomb's paradox as described by Nozick leaves quite some details unspecified. Is it not given what the relevant probabilities are, just that the predictor is right most of the time. Furthermore, it is not specified how the predictor arrives at his prediction. In a way, this means that Newcomb's paradox describes a whole range of different problems.

This means that, dependent on how these details are specified, the rational policy can differ. In chapter two we have seen that for a lot of people, there is a turning point for what the rational policy should be when the predictor becomes infallible. If p is the chance that the prediction is correct, these people feel that for $p < 1$ the rational strategy is to two-box, while for $p = 1$ the rational strategy becomes to one-box. Ahmed calls this the discontinuous strategy, and argues that it is irrational.

Price's proposal of the Evicausalist can be summarized as: In most realistic Newcomb cases, it is rational to two-box. Only in exceptional cases with a very special predictor, one-boxing is the rational choice. This sounds an awful lot like the discontinuous strategy, and we have already seen that the discontinuous strategy is irrational. Is Price's proposal susceptible to the same kind of counterexample that showed the irrationality of the discontinuous strategy?

Fortunately, it is not. While both the discontinuous strategy and Price's Evicausalism have in common that only in special cases of an extremely competent predictor we wish to one-box, they think differently about what it is to be so competent.

In the case of the discontinuous strategy as described by Ahmed, being extremely competent simply amounts to being always right. The predictor has to be infallible.

In the case of Evicausalism being extremely competent is a little more subtle. The key question is whether the prediction can be circumvented by making the choice random after the prediction has been made. If this is the case, the predictor is not good enough, and two-boxing is the rational choice. If the predictor is still capable of making a good prediction, even if you decide to randomize your choice after the prediction has been made, than this predictor is good enough and one-boxing is the rational choice.

The crucial element is thus the performance of the predictor in cases of randomized choice. However, it is not necessary that the predictor never makes a mistake. Suppose you know that the prediction of a certain predictor is correct in 80% of the cases. You do a test with subjects that decide by flipping a coin after the prediction is made. If it is still the case that 80% of the predictions are correct, than this predictor is good enough in the sense needed by Evicausalism. The crucial factor is not that the predictor is always correct, but that his prediction powers are not diminished by randomized choice.

Furthermore, the scenarios where the predictor's powers are not diminished by randomized choice, are precisely the scenarios where the contents of the opaque box are causally dependent on your choice. This means that EviCausalism does not propose a discontinuity for some arbitrary predictor success rate. It proposes a discontinuity for a meaningful distinction: whether the prediction is causally dependent on your choice or not. If there is causal dependency, then one-boxing is the rational policy, while if there is no causal dependency, two-boxing is the rational choice.

6 Conclusion

The goal of this thesis was to find a way to causally justify one-boxing. This was motivated by the fact that one-boxing is clearly the best way to get rich, but that it seems undesirable to have a decision theory that ignores causal considerations.

The crucial insight is that Newcomb's paradox is an extremely unrealistic problem. It seems safe to assume that a predictor as the one described, does not exist. This means the paradox revolves about what to do in supernatural cases.

Lewis argues that in supernatural cases, inadmissible evidence allows for an exception to the principal principle. This means that in supernatural cases, rational credence does not necessarily follow objective chance.

Price advocates that something similar should hold for decision theory. In supernatural cases, it is acceptable that EDT and CDT do not coincide. Prices even takes the theory one step further and argues that in exceptional cases like Newcomb's paradox, our usual guides to the causal structure of the world can be wrong. He proposes a new way to understand causal dependency, such that causal and evidential dependence can no longer come apart.

This results in EviCausalism, a decision theory that is grounded on an evidential understanding of causal dependency. EviCausalism ensures that recommendations by EDT are always based on causal considerations. Cases like Newcomb's paradox, where traditional EDT and CDT disagree, are simply cases where our usual way of thinking about causal dependency is mistaken.

In this thesis I have closely followed Price and Lewis. However, every explicit example is mine. Both Price and Lewis hardly ever fill out their own formulas. Because of this, it is hard to grasp what their theories entail. Especially Price's EviCausalism is very hard to understand without actually specifying what the relevant probabilities are, so I have provided these to show how EviCausalism deals with Newcomb's paradox, and how it handles the smoking lesion and backwards causation. Furthermore, I added the discussion of the discontinuous aspects of EviCausalism.

For future research it might be interesting to apply EviCausalism to other problems in decision theory, such as the pauper's problem. It could also prove interesting to explore how EviCausalism handles every day decision problems, and whether the notion of causation that it proposes remains reasonable in normal situations.

To summarize, we can say that in exceptional cases, our usual guides to rational policy are faulty. In cases like that, rational credence does not necessarily follow objective chance, CDT and EDT do not need to coincide, and backwards causation might be truly plausible.

References

[1] Arif Ahmed. "Infallibility in the Newcomb problem". In: *Erkenntnis* 80.2 (2015), pp. 261–273.

[2] Frank Arntzenius. "No regrets, or: Edith Piaf revamps decision theory". In: *Erkenntnis* 68.2 (2008), pp. 277–297.

[3] Adam Bales. "The pauper's problem: chance, foreknowledge and causal decision theory". In: *Philosophical Studies* 173.6 (2016), pp. 1497–1516.

[4] Maya Bar-Hillel and Avishai Margalit. "Newcomb's paradox revisited". In: *The British Journal for the Philosophy of Science* 23.4 (1972), pp. 295–304.

[5] Max Black. "Why cannot an effect precede its cause?" In: *Analysis* 16.3 (1956), pp. 49–58.

[6] Michael Dummett and Antony Flew. "Symposium: Can an effect precede its cause?" In: *Proceedings of the Aristotelian Society, Supplementary Volumes* (1954), pp. 27–62.

[7] David Lewis. "A subjectivist's guide to objective chance". In: *Ifs*. Springer, 1980, pp. 267–297.

[8] David Lewis. "Causal decision theory". In: *Australasian Journal of Philosophy* 59.1 (1981), pp. 5–30.

[9] David Lewis. "Humean supervenience debugged". In: *Mind* 103.412 (1994), pp. 473–490.

[10] Robert Nozick. "Newcomb's problem and two principles of choice". In: *Essays in honor of Carl G. Hempel*. Springer, 1969, pp. 114–146.

[11] Huw Price. "Causation, chance, and the rational significance of supernatural evidence". In: *Philosophical Review* 121.4 (2012), pp. 483–538.

[12] Brian Skyrms. "Causal decision theory". In: *The Journal of Philosophy* 79.11 (1982), pp. 695–711.