This is a preprint of an article whose final and definitive form will be published in the Australasian Journal of Philosophy. The Australasian Journal of Philosophy is available online at: http://www.tandf.co.uk/journals/.

Bad Apples and Broken Ladders

A Pragmatic Defence of Causal Decision Theory

Adam Bales
St John's College, the University of Cambridge

Abstract

While pragmatic arguments are traditionally seen as supporting decision theory, recent discussions suggest the possibility of pragmatic arguments *against* this theory. I respond to two such arguments and clarify what it would take for arguments of this sort to succeed.

Keywords

Decision Theory, Pragmatic Arguments, Satan's Apple, Jacob's Ladder

1 Introduction

Pragmatic arguments are intended to show that rationality requires certain things of agents. In particular, pragmatic arguments show that unless an agent satisfies some specified criterion, she can be made to suffer a foreseeable and avoidable loss. Assuming rationality forbids such loss-making behaviour, it follows that rationality requires the specified criterion be met.

Historically, pragmatic arguments have been thought to support decision theory, by supporting some of its foundational assumptions. However, recently pragmatic arguments have been developed *against* decision theory. One such argument can be extracted from Arntzenius et al. [2004] and another from Peterson [2016]. If these arguments succeed then decision theory collapses.

In this paper, I'll defend decision theory against these arguments. In doing so, I'll unpack criteria for the success of pragmatic arguments.

2 Decision Theory

There are various versions of decision theory, but I'll focus on causal decision theory (CDT) for two reasons. First, it's prominent; in philosophical discussions, it's often presented as the orthodox version of decision theory. Second, it's at greatest risk from the objections to come; it most-indisputably runs into difficulties in SATAN'S APPLE, the first case to be discussed (see footnote 5). As a result, focusing on CDT makes my response to the challenges stronger by showing that they fail even in the context most conducive to their success.

While CDT itself comes in different versions, my discussion applies equally to all prominent forms. So, merely as a matter of personal preference, I'll focus on a form drawn from Joyce [1999], which appeals to the notion of expected utility (EU) defined as:

$$EU(O) = \sum_S Cr(S \backslash O) U(S \wedge O)$$

Here:

1. *O* is an option available to the agent. For example, the option of ordering pizza. So EUs are assigned to options.
2. *S* is a set of mutually-exclusive and exhaustive states of the world, specifying details relevant to the decision. So such states might specify how expensive the pizza is and how tasty it is.
3. *Cr(S\O)* is a causal credence, representing the agent's degree of belief in *S*, after accounting for *O*'s causal impact. We can think of this as the agent's credence in the (non-backtracking) counterfactual that if she were to *O* then *S* would hold.
4. *U* is the agent's utility function, which assigns real numbers to outcomes (pairs of states and options) representing their desirability. A higher utility represents a more desired outcome.

So an option's EU is a credence-weighted sum of the utilities it might lead to. CDT holds that choosing some option is rational if, and only if, this option maximises EU, relative to the agent's credence and utility functions.[1]

3 Pragmatic Arguments

A pragmatic argument against CDT would show that agents who maximise EU can be made to suffer an (avoidable, foreseeable, and uncompensated-for) loss. In so far as rationality requires us to avoid such losses, this CDT-endorsed behaviour would be irrational and so CDT would be an inadequate theory of rationality.

So far, so simple. Still, it's worth considering more slowly the ways that I qualified the loss; I said it must be avoidable, foreseeable, and uncompensated for.

Let's first consider the sense in which the loss must be avoidable. Clearly, an agent needn't be irrational for suffering a loss compared to her current circumstances. If you must either lose an arm or both an arm and a leg, you may rationally choose the former. So agents can rationally suffer loss. Pragmatic arguments succeeds only if:

> *Weak Avoidability:* the loss involved is avoidable, in that: (a) some alternative sequence of choices would leave the agent better off; and (b) the agent has the power to make each choice in this alternative sequence.

(I'll discuss the sense in which this criterion is 'weak' later)

Moving on, let's consider the sense in which the loss must be foreseeable. Even an avoidable loss could rationally be suffered if it couldn't be foreseen. Perhaps an agent believes, on good grounds, that drinking the green vial will cure her illness and drinking the blue will make things worse. So, quite rationally, she drinks the green vial. Unfortunately, she had it the wrong

[1] I focus on a normative reading of decision theory and on the deliberative conception (which treats credences and utilities as psychologically real, see Pettigrew [2015]). However, with minor variations, my discussion also applies to the preference-first conception.

way around and so her choice makes things worse. This agent has rationally suffered an avoidable loss. So, pragmatic arguments succeeds only if:

> *Foreseeability:* the loss involved is foreseeable, in that the agent knows (when she makes her decision) about the loss that will (or might) result.[2]

Finally, let's consider the sense in which the loss must be uncompensated for. An agent can rationally lose out, even when that loss is avoidable and foreseeable. Perhaps I accept a bet on a fair coin that pays out $1,000,000 on heads and costs me $1 on tails. Unfortunately, the coin comes up tails. Consequently, I lose money. This loss is avoidable (I could have refused the bet). It's foreseeable, in the above sense: it doesn't rely on me lacking knowledge of the potential loss. So a pragmatic argument succeeds only if:

> *Uncompensated:* the loss involved is uncompensated for, in that there's no potential gain from the decision that is sufficient to counterbalance the potential loss.

So if a pragmatic argument against CDT is to succeed, it needs to show that EU-maximising behaviour can lead to a loss that is avoidable, foreseeable, and uncompensated for.

4 Satan's Apple

I now turn to the first of the pragmatic arguments against CDT, which has its roots in a case from Arntzenius et al. 2004:[3]

> SATAN'S APPLE: Satan cuts an apple into a countable infinity of pieces and offers one piece at a time to Eve, such that she's offered all of the pieces within two minutes.[4] All else being equal, Eve wants more apple rather than less. Yet if she takes infinitely-many pieces then she'll be cast from Eden, and no amount of apple is worth such a fate!

CDT endorses taking each of the apple pieces.[5] To demonstrate this, I stipulate that Eve's actions are causally independent of one another: a decision about whether to take one piece has no causal influence on whether she takes others. Further, I stipulate that being cast from Eden

---

[2] Instead of knowledge, we might spell this out in terms of belief or available information. These distinctions are unimportant for my purposes.

[3] Arntzenius et al. don't take this case to undermine CDT. However, it's natural to read the case as posing such a threat and so worth exploring whether it truly does so. Note that I'm not alone in reading the case this way: Meacham [2010] also sees it as threatening CDT, and Peterson [2016: 170–171] raises the possibility that it underpins a pragmatic argument against decision theory.

[4] She'll be offered the first slice in one minute, the second thirty seconds later, the third fifteen seconds later, and so on.

[5] There's disagreement over whether evidential decision theory (EDT), a prominent alternative theory, provides this same guidance (Meacham [2010: 54–5] says yes; Arntzenius et al. [2004: 268] suggest no). I deny that EDT does so given a natural reading of SATAN'S APPLE (because on this reading, taking one slice provides strong evidence to Eve that she'll take others). However, a more contrived reading of the scenario will lead EDT to endorse taking each slice (but at the cost of the contrivance threatening the reliability of our intuitions in the case). There are also versions of decision theory that clearly don't endorse taking each slice (see Meacham [2010: 68–9] and Greene [2018]). As CDT faces the clearest challenge from SATAN'S APPLE, it's this theory that I focus on (though the defence I provide could be extended to EDT).

contributes -100 to an outcome's utility, while possessing the nth piece contributes a utility of $(\frac{1}{2})^n$.

We can now consider the decision of whether to take the nth piece. In making this decision, there are two relevant possibilities:

1. Eve might take infinitely-many other apple pieces. If she does then she'll be cast from Eden and will possess these other pieces, regardless of her current choice. As such, the utility contributed by these factors will be the same regardless of whether Eve takes the nth piece. So the only difference is that she'll gain a utility of $(\frac{1}{2})^n$ from taking this piece that she wouldn't gain otherwise.
2. Eve might take finitely-many other apple pieces. In which case, she won't be cast from Eden regardless of her current choice (as taking one more piece won't lead her to take infinitely-many overall). Further, the utility contributed by the other pieces will be the same regardless of whether Eve takes the nth piece. Again, the only difference is that Eve will gain a utility of $(\frac{1}{2})^n$ from taking this piece that she would not gain otherwise.

Either way, taking the current piece has a utility $(\frac{1}{2})^n$ greater than refusing it. So, however Eve splits her credence between these possibilities, taking the piece will have an EU $(\frac{1}{2})^n$ greater than the EU of refusing it. Consequently, CDT endorses taking the piece. Further, as this reasoning applies to each of Satan's offers, CDT will endorse taking each piece offered.[6]

Yet if Eve does so, she'll take all of the apple pieces and be cast from Eden. Further, this loss is:

1. (Weakly) Avoidable: Eve would be better off if she took finitely-many apple pieces and so remained in Eden.[7]
2. Foreseeable: Eve knows that taking all of the pieces will lead her to be cast from Eden.
3. Uncompensated For: Eve's gain (an apple) doesn't compensate her for being cast from Eden.

So SATAN'S APPLE apparently constitutes a pragmatic argument against CDT: if Eve maximises EU she'll suffer an avoidable, foreseeable, and uncompensated-for loss. If no more could be said, CDT would be in trouble.

5 Deciding at Infinity

[6] Bartha et al. [2014] argue that CDT endorses refusing some pieces in some versions of SATAN'S APPLE. However, as they themselves note, it's unclear whether their discussion applies to all versions of this scenario. So, I'll assume CDT does offer the specified guidance and will argue that, even if so, a pragmatic argument, based on SATAN'S APPLE fails.

[7] Some have argued that Eve is rationally required to take each slice [Arntzenius et al. 2004: 267; Bartha et al. 2014: 639–640). If so then her loss is avoidable only if she acts irrationally. This doesn't violate *Weak Avoidability* but might appear to undermine the sense in which the pragmatic argument is avoidable. Indeed, we might be tempted to adopt what we could call *Rational Avoidability*, a variant on *Weak Avoidability* where every choice in the alternative sequence must be rational. If so, and if we accept that Eve is rationally required to take each slice, then SATAN'S APPLE fails as a pragmatic argument. In any case, I'll assume this response fails and show that CDT can be defended even so. Thanks to a reviewer for their thoughts here.

Fortunately, more can be said. In particular, reflection on two features of SATAN'S APPLE reveals that the pragmatic argument fails.[8] I take these features one-by-one, reflecting first on the fact that SATAN'S APPLE involves an infinite sequence of decisions (Eve is offered infinitely-many apple pieces).

To see why this is important, step back from SATAN'S APPLE, and consider CDT more generally. Plausibly, this theory applies only in finite cases: cases involving finite sequences of decisions between finitely-many options, where the agent's utilities are bounded.[9]

There are powerful reasons to accept this claim. Start with the simplest: sometimes when there are infinitely-many options, no option will maximise EU (perhaps for any natural number there's an option with an EU equal to this number). CDT is ill-placed for application in such cases.[10]

Further problems arise for applying CDT in infinite cases. For example, Bernoulli's St. Peterburg paradox involves a game with an infinite EU. As such, CDT suggests it's worth paying any finite amount to play this game. Yet it's clear that the game shouldn't be valued so highly, so the game's EU is misleading.[11] Then the Pasadena paradox [Nover and Hájek 2004] and related scenarios [Colyvan and Hájek 2016] involve options with no well-defined EU. Clearly CDT is a poor fit for situations where options have no EU at all.[12]

So in infinite contexts we can have options without an EU, cases where no option maximises EU, and cases where the EU of an option is misleading. So we have grounds to restrict CDT to finite contexts. That is, we have grounds to read CDT as saying that *in finite contexts* an option is rational if and only if it maximises EU (and to read CDT as falling silent in infinite cases).

Various people might be read as objecting to such a move. For example, Nover and Hájek [2004: 246–247] argue that it's hard to justify a finitary restriction on decision theory. First, they note that infinities appear throughout physics and mathematics, including in probability theory (which is 'kindred' to decision theory) and in mathematical arguments for decision theory (based around the law of large numbers). Consequently, the finitary restriction strikes Nover and Hájek as 'high-handed' and 'at odds with the conceptual underpinnings of decision theory itself'. Second, they note that decision theory is intended to apply to idealised agents, so the finitary restriction can't be justified by claiming that humans never face infinity.[13] So it might seem illegitimate to restrict CDT to finite cases.

However, nothing that I said above commits me to claiming that a complete decision theory doesn't apply in infinite cases. All I've said is that CDT doesn't apply in such cases, and I can (and do) deny that CDT is a complete decision theory. On my view, a complete decision theory will apply in both finite and infinite cases but will reduce to CDT in the finite case.[14] This view avoids the above challenge, as it doesn't deny the legitimacy of infinities in decision theory.

---

[8] There are other ways we might defend CDT. For example, on one reading, Arntzenius et al. argue that: (a) Eve would avoid problematic behaviour if she could bind herself to future actions; and (b) pragmatic arguments lack force under such circumstances. However, it has been argued elsewhere that this defence fails [Meacham 2010; Bartha et al. 2014: 645–7]. I'll show that SATAN'S APPLE poses no challenge to CDT even if so.

[9] Alternatively, perhaps CDT applies in infinite cases but pragmatic arguments don't (see Arntzenius et al. [2004: 278–9]; Bartha et al. [2014: 644]).

[10] This assumes that at least one decision is rational in such cases. If these cases are dilemmas, where no decision is rational [Slote 1989: ch. 5], then CDT gets things exactly right: no decision maximises EU and so CDT (correctly) takes all decisions to be irrational.

[11] In Bernoulli's game a fair coin is tossed until it first lands heads. You receive a payout of 2^n, where *n* is the number of tosses.

[12] These cases reveal that CDT cannot be correctly applied in all situations. One response would be to reject CDT; another would be to restrict its scope. I take the second path. Shortly, I'll discuss why this move isn't ad-hoc.

[13] See also Bartha et al. [2014: 635, 644]; Arntzenius et al. [2004: 260].

[14] Depending on the role we take CDT to play, reduction may not be necessary. It may be enough that CDT's guidance is, in finite cases, coextensive with the guidance provided by the complete theory.

Instead, my view implies only that a decision theory that applies in finite cases might not apply, without modification, in infinite cases. This is hardly a shocking point. For a start, this view has been defended by others (see the discussion in Colyvan and Hájek [2016]). Further, looking beyond decision theory, the mathematics of infinite cases often departs from the mathematics of equivalent finite cases. In the kindred context of probability theory, for example, we appeal to finite additivity in the finite case and (the more controversial) countable additivity in the infinite case. So there's nothing odd in thinking that a mathematical theory like CDT might apply in finite cases but not, without modification, in all infinite cases. My finitary restriction remains safe.

Having made this restriction, CDT doesn't apply in SATAN'S APPLE and so no longer endorses loss-making behaviour here. Consequently, this case doesn't support a pragmatic argument against CDT. To put it another way, SATAN'S APPLE provides a pragmatic argument against CDT if this theory is interpreted as a complete theory, but not otherwise. As we already have grounds to interpret CDT as incomplete, we have grounds to reject the pragmatic argument against CDT.

6 Deciding in Dilemmas

A second defence of CDT can also be provided, by drawing on the fact that SATAN'S APPLE is, in a sense to be explored, a dilemma.

6.1 Dilemmahood

A rational dilemma is a decision scenario where all options are irrational [Priest 2002]. SATAN'S APPLE isn't the right sort of the thing to be a dilemma in this sense: it isn't a single decision scenario but rather a sequence of scenarios (each involving taking or refusing a single apple piece).[15] Nor is it clear that the members of this sequence are all dilemmas. For example, when Eve faces the first apple piece it's not clear that it's both irrational to take it and irrational to refuse it.

Still, SATAN'S APPLE is what we might call a *sequence dilemma*: any sequence of choices Eve can make is irrational.[16] To see why, consider what makes pragmatic arguments compelling. Such arguments involve an agent making a series of choices that leave her worse off than some alternative sequence. Further, the agent can identify the better alternative. Yet, if the agent can identify a better sequence of choices, surely she should carry out this sequence instead. So the force of pragmatic arguments comes from the *preferred-alternative claim*: it's irrational to make a sequence of choices if some alternative would foreseeably leave you better off (because under such circumstances, you should carry out the alternative sequence instead).[17]

[15] This is true of the diachronic version of SATAN'S APPLE considered here. Arntzenius et al. 2004: 264–5 also discuss a synchronic version of the case.

[16] This relies on sequences being, in some sense, assessable for rationality. This is a safe assumption in the current context, because pragmatic arguments rely on this same assumption. So if this assumption were rejected, the pragmatic argument would collapse (see Hedden [2015]). Further, the (second framing of the) argument in section 6.3 doesn't rely on sequences being rationally evaluable, so applies even if we reject the evaluability assumption.

[17] One might instead argue that a sequence is irrational only if it leaves the agent (foreseeably, avoidably) worse off than they started. On such a view, what matters isn't loss compared to any alternative but specifically loss compared to the status quo.

However, this seems to be an instance of the well-known human tendency to overemphasise the relevance of the status quo (see Samuelson [1988]). There's no reason to think that only loss relative to the status quo is irrational and that otherwise it's fine to knowingly make yourself worse off than you could be. So there's no reason to accept this more-restricted justification for pragmatic arguments, while rejecting the preferred-alternative claim.

Yet it follows that SATAN'S APPLE is a sequence dilemma. After all, if Eve takes infinitely-many apple pieces then she would have been better off had she instead taken, say, just two million pieces. Yet if Eve takes any finite number of pieces then she would have been better off taking one more. So however many pieces Eve takes, some alternative sequence would foreseeably leave her better off. By the *preferred-alternative claim*, all available sequences are irrational. SATAN'S APPLE is a sequence dilemma.

6.2 Applicability of CDT

This matters because, on one reading, CDT cannot be applied to scenarios that form part of a sequence dilemma. To see why, distinguish two types of theory of rationality. A *comprehensive theory* aspires to account for every feature of rationality. Meanwhile, a *guiding theory* aspires merely to provide guidance that allows agents to act rationally (without aspiring to clarify issues irrelevant to the provision of guidance). For example, imagine you can choose to gain either $1, $2, or $3. Here are two plausible normative facts: (1) you should take the $3;[18] (2) it would be better to take the $2 than the $1. A comprehensive theory would succeed only if it captured both of these facts, and others besides. However, a guiding theory would succeed as long as it captured the first fact.

Now consider traditional dilemmas, scenarios where no option is rational. A comprehensive theory must capture all feature of these scenarios, including the fact that no option is rational. However, in dilemmas there's no useful guidance to be provided, as whatever the agent does, she'll act irrationally. Consequently, guiding theories plausibly don't apply in such cases. So dilemmas cannot pose a challenge to guiding theories.

How about sequence dilemmas? In these cases, the agent cannot avoid carrying out an irrational sequence of choices. Now, perhaps guiding theories should provide guidance about how to act in a manner that is rational overall (that involves neither irrationality of choice nor of sequence). If so then there's no useful guidance to be provided in sequence dilemmas. As with traditional dilemmas, perhaps guiding theories shouldn't be applied in such cases.

So if we interpret CDT as a guiding theory and take the above view then CDT doesn't apply to SATAN'S APPLE, which is a sequence dilemma. Once again, the pragmatic argument collapses.

6.3 Applicability of Pragmatic Arguments

Still, this relies on a particular view of CDT's role and of guiding theories. So it's worth saying something further. In particular, I'll argue that even if CDT applies in sequence dilemmas, pragmatic argument don't succeed in this context.[19]

As a starting point, note that pragmatic arguments show that some sequence of choices is irrational. Normally, this is grounds to criticise a theory that leads agents to carry out that sequence. However, in a sequence dilemma, all sequences are irrational, and so carrying out an irrational sequence simply reflects the inevitable reality of the situation. It needn't reflect any flaw in the theory of choice. So pragmatic arguments fail in this context. Applying this to our specific case: as SATAN'S APPLE is a sequence dilemma, a pragmatic argument here gives no grounds for rejecting CDT.

The point can be made in different terms. Consider again the idea that the loss involved in a pragmatic argument must be avoidable. Not only is this true but perhaps avoidable loss must

---

[18] Assuming that more money is better than less.

[19] Bartha et al. [2014] reach the same conclusion, but their argument has a different shape to my own. They introduce a principle, show that this entails the rationality of taking each apple piece, and argue that this reveals that SATAN'S APPLE cannot underpin a pragmatic argument.

itself be avoidable. As above, if avoidable loss is inevitable then a theory of choice can hardly be criticised for the fact that it leads agents to such losses. So, we need to replace *Weak Avoidability* with:

> *Strong Avoidability:* the loss involved must be avoidable, in that: (a) there's some alternative sequence of choices that would leave the agent better off; (b) the agent has the power to make each choice in this alternative sequence; and (c) no third sequence would leave her even better off than the alternative.

Clause (c) ensures that there's an option that doesn't lead to an avoidable loss. While SATAN'S APPLE satisfies *Weak Avoidability* it does not satisfy *Strong Avoidability*. So, the pragmatic argument fails.

Overall, a pragmatic argument based around SATAN'S APPLE fails because CDT doesn't apply in this case (as it's an infinite case and a sequence dilemma). And it fails again because pragmatic arguments lack force here.

7 Jacob's Ladder

So I turn to JACOB'S LADDER, which Peterson [2016] takes to provide a pragmatic argument against CDT (and other forms of decision theory).[20] This scenario has three core features.

First: Jacob will be offered an infinite sequence of bets (within a finite time), such that each will be resolved by tossing a fair coin. The first bet requires him to put up a stake of x units of utility (where this value will be discussed shortly), such that this stake will be doubled if the coin comes up heads and lost if it comes up tails. Each remaining offer then requires him to put up all of his previous winnings, such that this stake will be doubled if the coin comes up heads and lost if it comes up tails. (For now, I ignore a crucial complication, by setting aside the question of whether Jacob continues to be offered further bets if he refuses a bet. I'll return to this matter shortly.)

Thus far, it seems that Jacob should be equally happy to either take or leave a given bet. If he leaves it, he'll receive his stake for certain; if he takes it, he'll have a 50% chance of receiving double his stake. The EUs of the decisions are the same.

The second feature of the scenario shakes this up by providing an incentive to bet. After all of the bets, Jacob will receive an additional payout of $\frac{1}{4}x \cdot \sum_{m=1}^{n}(\frac{1}{2^m})$, where n is the number of bets accepted. This payout: (a) gives Jacob an incentive to take each bet (it breaks the tie between betting and not betting); and (b) won't grow in an unbounded manner but rather approaches $\frac{1}{4}x$ as the number of bets increases.

The final feature of the bet specifies how x (that is, the initial stake) is chosen. In particular, Peterson is careful to specify this value so that the sequence of bets is compatible with the assumption that Jacob's utility function is bounded (in particular, that it ranges from 0 to 1). Why? Because it has been argued that rational agents must have bounded utility functions [McGee, 1999]. If so then, given that CDT applies only to rational agents, a pragmatic argument against CDT succeeds only if it involves a loss for agents with bounded utilities. So it's crucial that Peterson succeed here.

To ensure that the utilities are bounded, Peterson imagines that the coin tosses were carried out a week ago. Based on knowledge of these tosses, the value of x was set so that even if

[20] Peterson's pragmatic argument against CDT is part of a larger argument. Ultimately, Peterson concludes not that we should reject CDT but that we should reappraise the force of pragmatic arguments. However, not everyone will agree with this move; many will continue to think that pragmatic arguments represent a serious threat. As such, a pragmatic argument against CDT calls for a response on its own terms.

Jacob bets perfectly (that is, even if he bets on every occasion prior to the first tails and then stops), he won't receive more than half a unit of utility from the betting.[21] This means that on no offered bet will the potential payoff of betting exceed 1 and so Jacob can face the bets even with bounded utility. Further, while Jacob knows about this setup, it's stipulated that he doesn't know the value of x and so isn't able to determine when tails will first occur. Consequently, this knowledge doesn't change his valuations of each bet.

I can now outline the pragmatic argument against CDT. As noted above, the incentive ensures that taking each bet has a higher EU than refusing it. As a result, CDT will tell Jacob to take each bet. Yet if Jacob does so then, with probability 1, the coin will eventually land tails and Jacob will lose his initial stake of x, along with all winnings. This loss can't be compensated for via the incentive (which approaches $\frac{1}{4}x$), so Jacob has made a sequence of decisions that lead to a loss with probability 1. We have a pragmatic argument against CDT.

8 Refusing Bets

However, this argument runs into difficulties when we consider what happens if Jacob refuses a bet. Here's the initial point: for JACOB'S LADDER to undermine CDT, it must be the case that if Jacob refuses a bet then he's offered no further bets. This stipulation is required if the case is to apply to agents with bounded utility functions (as, per the above, it needs to).

To see why, set the stipulation aside and so assume that Jacob will be offered further bets, even if he refuses one. If the coin lands heads infinitely often then, for any value of x, there's a sequence of decisions that leads Jacob to a utility greater than 1. After all, by betting only when the coin comes up heads, Jacob can ensure that his initial stake is doubled arbitrarily-many times. Eventually, such doubling will lead the payoff to exceed 1, regardless of the value of x. So JACOB'S LADDER applies to agents with bounded utilities (and hence poses problems for CDT) only if the sequence of bets ends once Jacob refuses a bet. This stipulation is non-optional.[22]

However, this stipulation is fatal for the pragmatic argument. After all, when we calculate the EU of Jacob's options, we must take into account *all* of their consequences. Yet, given the stipulation, one consequence of Jacob taking a bet is that he'll be offered the next bet (and one consequence of Jacob refusing a bet is that he'll be offered no more bets). I didn't account for these consequences earlier, so the EUs haven't yet been adequately calculated.

To factor this consideration into the EUs, we need to know what'll happen if Jacob is offered future bets (as this will determine the consequences of seeking, or avoiding, such bets). More accurately, because CDT is a subjective theory, we need to know what Jacob believes will happen if he's offered future bets.

I divide the possibilities into two categories. First, let's assume that Jacob believes he'll take all future bets offered. It follows that if he takes the current bet (and so faces the future bets) then he'll eventually, with probability 1, lose his stake (and will get an incentive payment of, at most, $\frac{1}{4}x$). So the EU of taking the current bet will be, at most $\frac{1}{4}x$, once Jacob accounts for his belief that taking this bet will lead him to lose his stake.

On the other hand, if Jacob refuses the current bet then he'll be offered no further bets and so will retain his stake. As this stake is x or greater (x if this is the first bet, higher if Jacob has won previous bets), the EU of refusing the bet exceeds the EU of betting. So if Jacob is

[21] This argument assumes that if Jacob refuses one bet he's offered no further bets. I'll return to this issue shortly.

[22] Perhaps we can avoid this by: (a) stipulating that Jacob wins only finitely often; and (b) basing the value of x not just on the coin tosses but also on a perfect prediction of when Jacob will bet. Yet now the case relies on spookily-perfectly predictions of a free agent's actions (Newcomb's Problem can't be appealed to as precedence here, as it doesn't, despite standard presentations, rely on a predictor with unerring accuracy).

certain that he would accept all future bets then CDT endorses refusing the current bet.[23] The pragmatic argument then fails, as CDT doesn't tell Jacob to accept each bet.

Alternatively, imagine that Jacob isn't certain he'll accept all future bets. If so, it's possible that the EU of taking each bet will exceed the EU of refusing, so CDT may endorse taking each bet. Still, under these circumstances, the pragmatic argument collapses anyway. After all, *Foreseeability* is now violated, as Jacob doesn't realise his actions will inevitably lead him to lose his stake (he would know this only if he knew he would continue taking bets indefinitely).[24]

So, either CDT doesn't endorse betting (if Jacob is certain he'll take all future bets) or *Foreseeability* is violated (if Jacob isn't certain he'll take all future bets). Either way, the pragmatic argument fails.[25]

9 Infinities Again

Nor do the problems end there. After all, JACOB'S LADDER involves infinitely-many bets, and I've already argued that CDT doesn't apply in such cases. So the pragmatic argument runs into further trouble.[26]

Yet, here, a response is available: according to Peterson, JACOB'S LADDER can be given a finite form. To do so, we restrict the number of gambles to some finite maximum (there are to be at most a trillion bets, say). As a result, the chance of Jacob losing his stake and winnings ceases to be 1. However, as the maximum number of gambles can be set arbitrarily-high, the chance of Jacob losing all of his money can be made to be $1-\epsilon$ for an arbitrarily small ϵ. Peterson holds that for some sufficiently small ϵ it would be irrational to make choices that lead to a loss with chance $1-\epsilon$.[27]

---

[23] Plausibly, this would lead Jacob to become less confident that he would take all future bets (see footnote 25 for a discussion).

[24] Jacob knows that *if he bets indefinitely, he'll lose his stake with probability 1*. However, Jacob is ignorant of the antecedent and so ignorant of the consequent. This sort of failure of *Foreseeability* suffices to sink the pragmatic argument: it's hardly news that a rational agent might choose poorly if they're ignorant of crucial features of the future (and in Jacob's case, his own future decisions are just such a feature).

Note that if Jacob's future actions are treated as chance events then he may believe accurately about his future actions while remaining uncertain whether he'll take all future bets. Yet we then have a failure of *Uncompensated*: there's now a chance of gain to justify Jacob accepting the chance of loss.

[25] The same reasoning applies to the form of CDT developed in Skyrms [1990], on which agents first calculate EUs, but then reassess their beliefs in the light of these calculations, and recalculate the EU in the light of these new beliefs, and so on. The process continues until an equilibrium is reached, at which the agent's latest beliefs do not lead to a shift in the EUs.

For my purposes, what matters is that at the process's end, Jacob's credence function will either assign credence 1 that he'll take all future bets (in which case, CDT won't endorse taking the current bet) or will assign a credence of less than 1 to this possibility (in which case, *Foreseeability* is violated). Again, the pragmatic argument fails either way.

There may appear to be a loophole: if Jacob is initially certain that he'll take future bets but ends up unsure of this in equilibrium, it might be thought that the former fact ensures that *Foreseeability* is satisfied, while the latter leads CDT to endorse taking each bet. However, *Foreseeability* cannot be satisfied in this way. What matters in assessing this requirement is what Jacob believes when he makes his decision, not what he believed at some earlier point when he hadn't finished accounting for evidence. *Foreseeability* must be satisfied by the same belief state that justifies Jacob's decision. There is no loophole.

(Further, it's not clear that Skyrm's approach really involves the distinction between initial and final beliefs that gave rise to the apparent loophole. Instead, it might be thought that the ideal agents that CDT addresses should simply start in equilibrium (see Joyce [2012: 128]). Perhaps the dynamic process is just a useful tool for identifying equilibria.)

[26] Arguably, the case is also a sequence-dilemma, as no sequence of actions in JACOB'S LADDER is optimal from Jacob's own perspective (even if some sequence is *objectively* optimal). As rationality is a subjective notion, the case is plausibly a sequence-dilemma in the sense that matters (and so plausibly falls afoul of the discussion in section 6).

[27] Peterson [2016: 171] speaks about what Jacob *will* do, but I assume the intended claim relates to what he *should* do.

Peterson [2016: 171] admits that this isn't quite a pragmatic argument (because it doesn't involve probability 1 of a loss). However, he claims it, 'shows that a very similar type of problem can arise in finite contexts.' He concludes that the problematic nature of JACOB'S LADDER doesn't rely on its appeal to infinity.

However, contra Peterson, the finite version of JACOB'S LADDER is unproblematic; CDT provides the right guidance here. After all, as the chance of escaping ruin becomes smaller, the reward for escaping ruin becomes greater. So the chance of winning may be small but the potential payout for winning is massive (at least as a proportion of the stake). In such cases, one can rationally take a risk: a small chance of high reward can be worth a large chance of a much smaller loss. It's a virtue of CDT that it delivers this result, not a vice.

To put this another way. The finite version of JACOB'S LADDER violates *Uncompensated*, because the high chance of a small loss is compensated for by the low chance of a large gain. So JACOB'S LADDER, in its finite form, fails as a pragmatic argument. Insofar as JACOB'S LADDER poses a problem, this arises *only* from the infinite case. As CDT doesn't apply in such cases, it doesn't apply to JACOB'S LADDER. The pragmatic argument collapses.

10 General Lessons

So neither SATAN'S APPLE nor JACOB'S LADDER underpin a successful pragmatic argument against CDT. Still, other such arguments might be developed, so it's worth addressing a general question: what criteria does a pragmatic argument against CDT need to satisfy? I'll now consider this question by drawing together, and expanding upon, this paper's lessons.

10.1 General Criteria

Earlier, I introduced *Uncompensated*, *Weak Avoidability*, and *Foreseeability*. I've had nothing further to say about *Uncompensated*. However, I've replaced *Weak Avoidability* with *Strong Avoidability* to capture the fact that a pragmatic argument succeeds only when some lossless sequence is available. Absent such a sequence, it's the situation that's responsible for the agent's loss. Under such circumstances, the target of the pragmatic argument can hardly be criticised.

Finally, let's consider *Foreseeability*. I want to clarify two features of this criteria. First, per footnote 25, it's crucial that the agent be able to foresee her loss *when she makes the decisions*. It's the agent's beliefs at these times that are relevant to assessing her behaviour. It's of no consequence whether she once believed that this loss was certain, if she no longer believes this when deciding. If I believe, at age 12, that a vaccine is dangerous, I may still rationally seek it at age 19 if I've changed my views.

Second, it's worth commenting on the relationship between *Foreseeability* and another oft-discussed epistemic criterion:

> *Symmetry*: The agent being exploited in the pragmatic argument must know everything that is known by the agent doing the exploiting.

That is, Eve must know everything Satan knows; Jacob must know everything known by the bookie offering the bets. After all, as Rabinowicz [2000: 135] notes, 'it is perfectly trivial that unequal knowledge in interpersonal interactions makes room for exploitation even when the agent is fully rational.' So, it might be thought, exploitation only reveals irrational when *Symmetry* is satisfied.

Yet this isn't right. For a start, not all pragmatic arguments involve an agent doing the exploiting: nature can play the exploitative role. Further, as Briggs [2010: 13] notes, the exploiter might 'deceive an agent unwittingly' (and so her epistemic state might be irrelevant to assessing the pragmatic argument). Or *Symmetry* might be violated because the exploiter knows irrelevant facts, without this revealing anything about the pragmatic argument. So *Symmetry* isn't a necessary criterion for the adequacy of pragmatic arguments.

So why care about this criterion? Because if the exploiter knows something the exploited agent doesn't, we might worry that the exploited agent doesn't know enough for the pragmatic argument to work: perhaps the loss merely reflects ignorance, rather than irrationality. Ultimately, *Symmetry* isn't about the knowledge of the exploiter but the ignorance of the exploited.[28] And a violation of this criterion isn't a perfect sign of problematic ignorance but a useful heuristic: whenever *Symmetry* is violated, we would do well to consider whether the exploited agent knows enough for her behaviour to count as irrational. Most saliently, when *Symmetry* is violated we should double check whether *Foreseeability* is violated too.

So we have three criteria—*Uncompensated*, *Strong Avoidability*, and *Foreseeability*—and we have a heuristic—*Symmetry*—that tells us when to apply particular scrutiny to the last of these.

10.2 Specific Criteria

While the above criteria apply to all pragmatic arguments, two other criteria apply specifically to pragmatic arguments against CDT. First, pragmatic arguments against CDT must satisfy:

> *Finite:* the pragmatic argument must appeal only to finite utilities, finitely-many scenarios, finitely-many states, and finitely-many options.

After all, CDT was designed to cope with finite cases, and as a general matter, there's no reason to think that a theory designed for the finite will apply in infinite cases without modification. As CDT also faces known problems in infinite cases, we should restrict CDT's scope to finite cases (and it cannot then be threatened by infinite cases).

Further, insofar as we take CDT to be a guiding theory, pragmatic arguments against CDT must satisfy:

> *Surmountable:* It must be possible for useful guidance to be provided in the situation underpinning the pragmatic argument.

Guiding theories make no claim to capture all truths about rationality. All they aim to do is provide guidance when there's guidance to be had. In cases where guidance is impossible, guiding theories shouldn't be applied. If CDT's a guiding theory, this means it can't be threatened by pragmatic arguments that appeal to cases where guidance cannot be provided.

So the general criteria are bolstered with *Finite* and *Surmountable*, criteria that apply specifically to pragmatic arguments against CDT.

11 Conclusions

Two pragmatic arguments can be presented against CDT, one based on SATAN'S APPLE and one based on JACOB'S LADDER. However, the first fails because SATAN'S APPLE is an infinite case (which CDT doesn't apply to) and involves a sequence dilemma (where pragmatic arguments don't apply). JACOB'S LADDER is also an infinite case, despite attempts to create a finite version. The pragmatic argument here then suffers from a further challenge relating to the implications of refusing bets. So both pragmatic arguments fail.

---

[28] See Hitchcock [2004: 412]; Briggs [2010: 13].

In reaching this conclusion, I've clarified what it would take to present a successful pragmatic argument against CDT. Such an argument would need to meet three general criteria—*Foreseeability*, *Avoidability*, and *Uncompensated*—and two specific criteria—*Finite* and *Surmountable*. For now, no such argument has been presented. CDT remains safe.[29]

References

Arntzenius, Frank, Adam Elga, and John Hawthorne 2004. Bayesianism, Infinite Decisions, and Binding, *Mind* 113/450: 251–83.

Bartha, Paul, John Barker, and Alan Hájek 2014. Satan, Saint Peter and Saint Petersburg, *Synthese* 191/4: 629–60.

Briggs, Ray 2010. Putting a Value on Beauty, In *Oxford Studies in Epistemology (Volume 3)*, eds. Tamar Szabo Gendler and John Hawthorne, Oxford: Oxford University Press, 3–34.

Colyvan, Mark, and Alan Hájek. 2016. Making Ado Without Expectations, *Mind* 125/499: 829–57.

Greene, Preston 2018. Success-First Decision Theories, In *Newcomb's Problem*, edited by Arif Ahmed, Cambridge, UK: Cambridge University Press, 115–37.

Hedden, Brian 2015. Options and Diachronic Tragedy, *Philosophy and Phenomenological Research* 90: 423–51.

Hitchcock, Chris 2004. Beauty and the Bets, *Synthese* 139/3: 405–20.

Joyce, James M 1999. *The Foundations of Causal Decision Theory*, Cambridge, UK: Cambridge University Press.

Joyce, James M 2012. Regret and Instability in Causal Decision Theory, *Synthese* 187/1: 123–45.

McGee, Vann 1999. An airtight Dutch book, *Analysis* 59/4: 257–265.

Meacham, Christopher J. G. 2010. Binding and its consequences, *Philosophical Studies* 149/1: 49–71.

Nover, Harris, and Alan Hájek 2004. Vexing Expectations, *Mind* 113/450: 237–49.

Peterson, Martin 2016. Do Pragmatic Arguments Show Too Much?, *European Journal for Philosophy of Science* 6/2: 165–72.

Pettigrew, Richard 2015. Transformative Experience and Decision Theory, *Philosophy and Phenomenological Research* 91/3: 766–74.

Priest, Graham 2002. Rational Dilemmas, *Analysis* 62/1: 11–16.

Rabinowicz, Wlodek 2000. Money Pump with Foresight, In *Imperceptible Harms and Benefits*, ed. Michael J. Almedia, Dordrecht: Kluwer Academic Publisher, 123–54.

Samuelson, William, and Richard Zeckhauser 1988. Status quo bias in decision making, *Journal of Risk and Uncertainty* 1/1: 7–59.

Skyrms, Brian 1990. *The Dynamics of Rational Deliberation*, Cambridge, MA: Harvard University Press.

Slote, Michael 1989. *Beyond Optimizing: A Study of Rational Choice*, Cambridge, MA: Harvard University Press.

[29] For feedback, thanks to Johan Gustafsson. Thanks also to a number of reviewers for detailed and insightful comments; the paper is substantially better for their assistance.