

# Intentions and instability: a defence of causal decision theory

Adam Bales<sup>1</sup> 

Published online: 1 December 2018  
© The Author(s) 2018

**Abstract** Andy Egan has recently presented a prominent objection to causal decision theory (CDT). However, in this paper, I argue that this objection fails if CDT’s proponent accepts the plausible view that decision-theoretic options are intentions. This result both provides a defence of CDT against a prominent objection and highlights the importance of resolving the nature of decision-theoretic options.

**Keywords** Causal decision theory · Intentions · Options

Does decision theory evaluate what acts we should carry out, what decisions we should make, or something else entirely? One natural view is that it tells us what intentions we should, or should not, form. In this paper, I will show that Egan’s (2007) objection to causal decision theory (CDT) fails if we accept this view, along with a popular view of the nature of intentions (roughly, that in Bratman 1987). Not only does this bolster CDT but it also reveals the importance of clarifying what options decision theory evaluates. After all, we need to do so in order to resolve a broader decision-theoretic dispute.

## 1 Causal decision theory and Egan

Start with CDT. According to this theory, an option,  $O$ , is permissible if it maximises expected utility (EU), defined as follows:

---

✉ Adam Bales  
atb39@cam.ac.uk

<sup>1</sup> St John’s College, University of Cambridge, Cambridge CB2 1TP, UK

$$EU(O) = \sum_S Cr(O \setminus S)U(SO)$$

Here  $S$  refers to a set of possible states of the world.  $U(SO)$  is then a utility function, which assigns a real number to outcomes (conjunctions of a state and an option), with a higher number representing a more desirable outcome. Finally,  $Cr(S \setminus O)$  is a causal credence, which captures the causal impact of  $O$  on  $S$ . There is debate about how we should think of this credence (see Joyce 1999) but for my purposes, this can be thought of as the agent's credence in the non-backtracking counterfactual "If I were to  $O$  then the world would be in state  $S$ ".

Egan's objection to this theory can now be outlined via *the Blade Runner's Button* (a variant on Egan's *Psychopath Button*)<sup>1</sup>:

In a future where androids often commit murder, Lil faces a button that permanently deactivates all androids. Lil weakly desires to deactivate the androids but strongly desires to live.

Now Lil thinks she's unlikely to be an android. However, she knows: (a) that humans tend not to press such buttons; and (b) that androids are programmed with subconscious self-destructive tendencies and so would be likely to press. Consequently, Lil thinks that only an android would be likely to press the button.

Plausibly, Lil should not press. After all, if she does then she's almost certainly an android and so pressing will almost certainly kill her. Surely, Lil should not press if she's confident that if she does so then this will cause her death.

However, at least given a natural reading of the options evaluated by decision theory (perhaps a reading on which options are actions), CDT endorses pressing. Informally, this follows from the fact that Lil is confident that she's not an android and so confident that pressing will deactivate all androids but won't kill her (combined with the fact that pressing doesn't *cause* her to be an android). So pressing has better causal effects than not pressing and so CDT endorses pressing.

More formally, this result follows from the fact that  $Cr(\text{android} \setminus \text{press}) = Cr(\text{android} \setminus \text{refrain}) = Cr(\text{android})$  (because Lil's choice doesn't causally influence whether she's an android), combined with the fact that  $Cr(\text{android})$  is low. For concreteness, let's imagine that Lil has a credence of 0.01 that she's an android (and so a credence of 0.99 that she's not an android). Further, we can take death to contribute  $-100$  to the utility of an outcome and ridding the world of androids to contribute 5. The utilities of the outcomes here will then be as outlined in Table 1.

The EU of the two options can now be calculated as:

<sup>1</sup> I owe this *Blade-Runner* inspired variant, to Richard Holton. I rely on it to avoid discussing exterminating those with a classified mental health condition.

**Table 1** The blade runner’s button

	Android	Human
Press	− 95	5
Refrain	0	0

$$\begin{aligned}
 EU(\text{press}) &= Cr(\text{android} \setminus \text{press})U(\text{android} \wedge \text{press}) \\
 &\quad + Cr(\text{human} \setminus \text{press})U(\text{human} \wedge \text{press}) \\
 &= 0.01 * -95 + 0.99 * 5 \\
 &= 4 \\
 EU(\text{refrain}) &= Cr(\text{android} \setminus \text{refrain})U(\text{android} \wedge \text{refrain}) \\
 &\quad + Cr(\text{human} \setminus \text{refrain})U(\text{human} \wedge \text{refrain}) \\
 &= 0.01 * 0 + 0.99 * 0 \\
 &= 0
 \end{aligned}$$

As  $EU(\text{press}) > EU(\text{refrain})$ , CDT endorses pressing. Insofar as pressing is irrational, CDT is in trouble.

## 2 Optionhood and intentions

Still, let’s set this difficulty aside for a moment to ask a question about CDT. So far, I have said that CDT labels an *option* as permissible if it maximises EU. But what are these options?

As Hedden (2012) portrays things, there are two natural views here. First, perhaps options are acts: perhaps decision theory evaluates things like the act of taking a job. Second, perhaps options are decisions: perhaps decision theory evaluates things like the decision to take a job.

Hedden rejects the former view. Why? Well consider an agent who can carry out one set of acts but believes she can carry out a distinct set of acts. Which of these sets forms the agent’s options? Not the acts she can actually carry out (but believes she cannot). After all, decision theory is a theory of a subjective, action-guiding ought. However, an agent can’t be guided by a theory that takes as options something she doesn’t realise she is capable of doing. Yet, equally, her options are not those acts that she believes she can carry out (but actually can’t). After all, this would sometimes lead to the implausible conclusion that the agent ought to do something that she is unable to do. As an agent’s options cannot be acts of either sort, the prospects for an options-as-acts view are dim.<sup>2</sup>

On the other hand, Hedden endorses the view that options are decisions. Roughly, he argues that on any plausible account of decisions, an agent will know what decisions she has the ability to make. If so, then there will be no room for the

<sup>2</sup> This gives a mere sense of Hedden’s own, more detailed, argument. See Hedden (2012).

above difficulties to arise, because there will be no room for an agent to believe she can make one set of decisions but to actually be able to make some distinct set of decisions. So accepting that options are decisions allows us to avoid the above difficulty and so we have grounds to accept this view.

We can now ask a further question: what are decisions? Here, Hedden is less forthcoming. However, he hints (in Hedden 2012, p. 352n) at the possibility that decisions involve the formation of intentions. So to decide to  $\phi$  is to form an intention to  $\phi$ . This is a very natural view (indeed, it's a view that has been defended previously: cf. Raz 1975). After all, if we accept that we do sometimes form intentions then it would seem profligate to add in a distinct category of "making a decision". Instead, it is more natural to think that making a decision just is forming an intention.

So on a natural view, the options evaluated by decision theory involve the formation of intentions.

### 3 Intentions and reconsideration

Fortunately, if options involve intention formation, Egan's objection to CDT collapses. In particular, if options involve intention formation then CDT will no longer ultimately endorse pressing but rather will, as desired, endorse refraining. As such, Egan's objection to CDT relies on an assumption about the nature of optionhood that the proponent of CDT can happily reject.

In order to demonstrate this, I will first outline a prominent account of intentions [largely drawn from Bratman (1987) and Holton (2009)].<sup>3</sup> On this account, intentions establish default behaviour: if I form an intention to  $\phi$  then I will, by default,  $\phi$ . In order to overcome this default, I must both *reconsider* my intention and, in doing so, *revise* it. When should I revise an intention (once I'm reconsidering it)? According to Bratman and Holton, I should do so if it would be irrational to now form the intention. In other words, we can simply apply CDT to determine whether I should revise or retain the intention.

When should I reconsider an intention?<sup>4</sup> Well, as Holton (2009, pp. 160–162) views things, we rarely consciously decide to reconsider. So instead of assessing the rationality of a decision to reconsider, we should instead be assessing the rationality of possessing certain subconscious habits of reconsideration. For example, perhaps rational agents will have a habit of reconsidering intentions that they later discover were formed under false pretenses.

Now, again on Holton's view, there is no single, simple rule that captures rational habits of reconsideration. Instead, there are a plethora of rules of thumb, each of

<sup>3</sup> I will take the truth of this account for granted. As such, I am not providing a decisive argument that CDT deals appropriately with the *Blade Runner's Button*. Instead, I am showing that the proponent of CDT has a natural response to Egan's objection and, hence, showing that this objection lacks force unless this response can be undermined. Note that Holton's discussion is about resolutions, rather than intentions in general. However, many of his points can be naturally extended to apply to intentions more generally.

<sup>4</sup> In Sect. 5.3 I discuss the view that CDT should simply be applied in this context too.

which provide some insight into rational reconsideration. Here, I will focus primarily on one such rule (a rule that Holton himself did not discuss). To get to this, it will help to consider a pair of cases.

**DRINK DRIVING:** Gareth knows that when drunk, he tends to think the dangers of drink driving overblown and becomes tempted to drive home (so as to avoid minor inconvenience). So, before he starts drinking, he forms an intention to not drive if he has more than two beers. He hopes that this will enable him to resist the temptation to drive drunk.

**DRINK DANCING:** Intan knows that when drunk, she comes to think her drunk dancing less embarrassing than it is. So, before she starts drinking, she forms an intention to not dance if she has more than two beers. She hopes that this will enable her to resist the temptation to dance drunk.

Now, **DRINK DRIVING** is a paradigm case where rational habits will ensure that Gareth does not reconsider his intention in the face of temptation. That is, we would judge Gareth's rationality poorly if he reconsidered such a sensibly-formed intention just because of his drunken views.

On the other hand, we could easily fill out **DRINK DANCING** such that it would be rational for Intan to reconsider her intention. For example, perhaps after drinking, Intan comes to believe that she is too uptight when sober and that it would benefit her mental wellbeing to let go for once. In such circumstances, there's nothing wrong with Intan reconsidering her intention.

What makes for the difference between these cases? It's not simply a matter of impairment: both Gareth and Intan may suspect that drinking impairs their judgements to just the same extent. Instead, what is different between the cases is the agent's beliefs about what's at stake. In **DRINK DRIVING**, Gareth thought a lot was at stake when he formed the intention (because he believed drink driving to be a serious peril). However, when drunk he thought far less was at stake (because he believed that driving drunk would merely save him from a minor inconvenience). On the other hand, in **DRINK DANCING** Intan thought that comparatively little was at stake when she formed the intention (she thought it a matter of embarrassment). Further, when drunk, Intan thinks that more than minor convenience is at stake, because she thinks that letting loose would benefit her mental health.<sup>5</sup>

In order to get from this distinction to a rule of thumb for rational reconsideration, let *initial stakes* refer to how bad the agent thinks it would be, when she forms the intention, if she later abandoned the intention. Further, let *later*

---

<sup>5</sup> As a referee noted, one might suspect that sober Intan would see her later dancing preferences as involving legitimate preference change, while sober Gareth would not feel likewise about his later driving preference. Perhaps this (rather than stakes) is what distinguishes the cases. However, imagine Gareth and Intan's changing preferences result purely from changing beliefs about their skill as drunk drivers and dancers. Insofar as this belief change results from the misleading effects of alcohol, Gareth and Intan would presumably each see their changing preferences as illegitimate (insofar as these result from irrational changes to beliefs). Yet, plausibly, reconsideration remains rational for Intan but not Gareth. Appeal to stakes accounts for this; appeal to an agent's legitimacy judgements does not. So, I continue to prefer an appeal to stakes here.

*stakes* refer to how bad the agent thinks it would be, at the time of potential reconsideration, if she does not abandon the intention.

Using this terminology, we can define:

STAKES: Rational habits of reconsideration will: (1) lead agents to reconsider when later stakes are substantially more weighty than initial stakes; (2) lead to non-reconsideration when the later stakes are substantially less weighty than the initial stakes; (3) lead to either reconsideration or non-reconsideration when the initial and later stakes are of similar weight.

Clause (2) of STAKES entails that Gareth should not revise his intention. After all, here the initial stakes (potential death) are far more weighty than the later stakes (a minor inconvenience). On the other hand, clause (3) of STAKES entails that Intan can rationally either reconsider or not reconsider. After all, here the initial stakes (embarrassment) are of similar weight to the later stakes (a minor gain in mental wellbeing). So STAKES makes sense of the difference between the above cases.<sup>6</sup>

#### 4 Intentions and androids

I now return to *the Blade Runner's Button*. Here, if options involve intention formation, CDT initially endorses forming the intention to press, as above. Still, we must now ask whether a rational agent will reconsider this intention. STAKES reveals that she will. After all, the initial stakes here were that the androids would survive, rather than die. However, once Lil forms the intention to press she comes to believe that she is almost certainly an android (as only an android is likely to press). Consequently, the later stakes will be Lil's own survival. Insofar as Lil cares far more about her own survival than about ridding the world of androids, clause (1) of STAKES comes into play. Consequently, Lil should reconsider her intention.<sup>7</sup>

Further, Lil will now revise her intention. After all, Lil now believes that she is almost certainly an android and, in the light of this, CDT will endorse revising the

---

<sup>6</sup> My core argument for STAKES appeals to these concrete cases. However, we might also be sympathetic to this principle for more abstract reasons. To see why, note that in cases of potential intention reconsideration there are two relevant perspectives: the agent's perspective at the time of intention formation and the agent's perspective at the time of potential reconsideration. Now it might be felt that plausible principles of intention reconsideration will need to give each of these perspectives their due. After all, intentions can't do some of the work required of them (say, in helping us to resist temptation) if we always assess them entirely from our present (perhaps temptation-clouded) perspective. Yet it is also clear that we must not allow ourselves to be held prisoner to our past perspectives, with our current perspective being given no say. A natural question: if we must listen to both perspectives and they provide divergent guidance, how is this guidance to be balanced? STAKES provides a plausible, partial answer: we listen to the perspective that takes the most to be at stake.

<sup>7</sup> What if Lil is not rational? Well, then she will not revise here intention. However, it's hardly a surprise that agents who reason inappropriately might end up poorly off (and so this result hardly poses a problem for CDT).

intention.<sup>8</sup> Why? Because Lil's decision does not causally influence whether she's an android and given that she believes she's an android, the best expected causal effect comes from refraining rather than pressing.

More formally, this follows from the fact that  $Cr(\text{android} \setminus \text{press}) = Cr(\text{android} \setminus \text{refrain}) = Cr(\text{android})$  (because Lil's choice doesn't causally influence whether she's an android). For concreteness, we can imagine that Lil has a credence of 0.99 that she's an android, after she forms the intention to press. We can now calculate the EU of the available options as follows:

$$\begin{aligned} EU(\text{retain}) &= Cr(\text{android} \setminus \text{retain})U(\text{android} \wedge \text{retain}) \\ &\quad + Cr(\text{human} \setminus \text{retain})U(\text{human} \wedge \text{retain}) \\ &= 0.99 * -95 + 0.01 * 5 \\ &= -94 \\ EU(\text{revise}) &= Cr(\text{android} \setminus \text{revise})U(\text{android} \wedge \text{revise}) \\ &\quad + Cr(\text{human} \setminus \text{revise})U(\text{human} \wedge \text{revise}) \\ &= 0.99 * 0 + 0.01 * 0 \\ &= 0 \end{aligned}$$

$EU(\text{revise}) > EU(\text{retain})$ , so a rational agent should revise.

So Lil will now intend to refrain from pressing the button. Should she reconsider *this* intention? By clause (2) of STAKES she should not. After all, the initial stakes here were Lil's life (that is, when Lil formed this new intention, she took her life to be at stake). On the other hand, the later stakes involve the less weighty consideration of whether the androids are wiped out. So, having formed the intention to refrain from pressing, Lil should not reconsider. Consequently, she will act in accordance with the default established by this intention and so will refrain from pressing.

CDT's proponent can now respond to Egan's objection. After all, CDT will not ultimately endorse pressing in the *the Blade Runner's Button*. It may initially endorse forming the intention to do so but a rational agent will then reconsider and revise this intention. Having then come to intend to refrain, the agent will stand by this intention and so refrain. CDT is compatible with the judgement that rational agents will refrain in *the Blade Runner's Button*.

## 5 Three objections

At this point, three objections arise.

<sup>8</sup> This part of the solution echoes the deliberative response to Egan's cases discussed in Arntzenius (2008) and Joyce (2012). However, unlike those solutions, the solution to be developed here will ensure that rational agents refrain from pressing the button, even if CDT is true. Contra this, Joyce ends up suggesting that rational agents may choose either option (and Arntzenius suggests something similar). Insofar as it seems problematic to say that pressing is permissible, even if it is not required, this means that the solution presented here has an important virtue lacked by these previous solutions.

## 5.1 Objection 1: Wasted effort

First, it might be argued that there is something problematic about CDT initially endorsing pressing (regardless of what a rational agent will later do, after having formed this intention).

Perhaps this objection will strike some as self-evident: insofar as pressing seems problematic, it might be taken to be obvious that an adequate theory of choice should never endorse forming the intention to press. Yet I simply deny the force of this brute appeal to intuition. Yes, there seems to be something wrong with Lil pressing the button, given that she expects doing so to cause her death. Nevertheless, I see no reason to think that there's anything obviously wrong with *forming the intention* to press, especially if Lil expects that she will ultimately repudiate this intention. Merely forming this intention does not, in itself, have any bad consequences. So a brute appeal to intuition here does not strike me as deeply concerning.

Still, a further argument could be provided to bolster the objection. To get to such an argument, note that forming and revising an intention comes at some cost (in time and mental effort). As such, CDT's guidance here leads agents to act in an unnecessarily costly manner, given that they could simply form the intention to refrain from the get go. So there seems to be something problematic about the fact that CDT initially endorses pressing.<sup>9</sup>

However, this objection contains the seeds of its own resolution. In particular, the utilities in Table 1 presupposed that deciding to press would actually lead Lil to go on to press. On the other hand, if Lil believes that electing to press would ultimately lead her to refraining then the utility value of the pressing option within each state should be the same as the utility value of the refraining option (because in both cases, she will ultimately refrain).<sup>10</sup> Indeed, once we account for the fact that reconsidering and revising intentions is costly, the utility value of pressing in each state should be lower than the utility value of refraining (because intending to press will lead to costly reconsideration and revision). Letting the effort of reconsidering and revising contribute  $-1$  to the utility of an outcome, the utilities in *the Blade Runner's Button* will be as per Table 2

It is now clear enough that the EU of pressing will be  $-1$  and the EU of refraining will be  $0$ . So refraining maximises EU. As such, CDT will immediately endorse refraining rather than pressing, because it would be futile to form an intention that one knows one will go on to revise (rather than simply forming the later intention immediately). So CDT does not endorse pressing at any point. The current objection collapses.

<sup>9</sup> I owe this concern to an anonymous referee.

<sup>10</sup> What if Lil doesn't believe this? Well, then the following will not apply. But it is hardly surprising that an agent might act suboptimally if she either fails to act rationally or believes falsely about some crucial matter. So CDT is not threatened by the undesirability of Lil's behaviour under such circumstances.



**Table 2** The blade runner's button

	Android	Human
Press	- 1	- 1
Refrain	0	0

## 5.2 Objection 2: Countervailing habits

Another objection arises. So far, I've discussed just the rule of thumb captured by STAKES. Still, as I noted earlier, there are other rules of thumb for rational reconsideration too. This raises the possibility that accounting for these other rules might disrupt the outlined view about how Lil ought to behave.

Of course, a mere possibility is not itself grounds for great concern. Unfortunately, a more concrete version of this objection can be presented. In particular, consider the following:

FORESIGHT: Rational habits of reconsideration will not lead an agent to reconsider on the basis of changes that the agent anticipated when she formed her intention.

Why accept FORESIGHT? Well, anticipated changes were accounted for when the agent formed the intention and so it might seem like double counting to reconsider on the basis of these changes.

However, a problem now arises. After all, in *the Blade Runner's Button*, Lil knows that once she forms the intention to press the button, she will come to believe that she's probably an android. So FORESIGHT suggests she should not reconsider on the basis of coming to believe this (and so should not reconsider on the basis of the influence that this shift in belief has on the perceived stakes). So FORESIGHT appears to undermine my argument.

In responding to this concern, the first thing to discuss is the way in which we should read rules of thumb like FORESIGHT and STAKES. In particular, on the view under discussion, these principles should not be read as providing absolute rules for when intentions should be reconsidered. Rather, they should be taken to provide grounds for reconsideration or non-reconsideration (cf. Holton 2009, pp. 160–162). When such principles clash, then, we don't have a contradiction but rather a case where we need to determine which rule of thumb wins out (that is, which provides stronger grounds).

So the question at hand is how we are to weigh STAKES against FORESIGHT. Well, let's consider a well-known case where they clash (adapted from Gauthier 1997):

GIRL GERMS: A young boy, Charlie, has no interest in girls. However, he notices that many boys become obsessed with them as they grow up. Charlie carries out his research and comes to understand why this occurs. He then forms an intention to not date girls as he grows.

Now, we can imagine that a few years have passed and Charlie finds himself attracted to girls. Should he reconsider his intention to refrain from dating girls? It is, I take it, clear that he should.

What do the principles discussed above suggest. Well, FORESIGHT argues against reconsideration: Charlie knew that his views would change in just this way when he formed his intention. As such, the intention formation already accounted for this change.

On the other hand, STAKES argues for reconsideration. After all, when Charlie formed the intention, he did so in order to avoid something that he saw as icky. But as Charlie is straight, at the later time, he believes that maintaining the intention would lead him to forgo one of the fundamental experiences of human life (romantic entanglement). The later stakes substantially outweigh the earlier costs and so clause (1) of STAKES supports reconsideration.

Insofar as reconsideration is the appropriate response to this case, we have grounds to think that when STAKES and FORESIGHT clash, it is STAKES that wins out.<sup>11</sup> Consequently, reflection on FORESIGHT does not undermine my discussion of *the Blade Runner's Button*. The second objection has been addressed.

### 5.3 Objection 3: CDT and reconsideration

Throughout this paper, I have assumed that CDT cannot be used to determine whether to reconsider an intention. Instead, we must appeal to heuristics like STAKES in this context. Yet, CDT is a general theory of choice: it tells us how we ought to make any decision at all.<sup>12</sup> As such, it applies just as much to the decision of whether to reconsider an intention as to any other decision. So, it might be objected, the view in this paper is flawed.<sup>13</sup> Further, it might be worried that once CDT is applied to determine when an agent ought to reconsider, the provided solution to *the Blade Runner's Button* will collapse. If so then CDT will once again be challenged by this case.

So far, so glum. However, this objection ultimately fails. In particular, it fails because it mischaracterises the nature of intention reconsideration.<sup>14</sup> On Bratman's view, agents do not typically *decide* whether to reconsider an intention. Rather agents simply do, or do not, reconsider as a result of habits or tendencies towards doing so. So discussions of reconsideration are not discussions of decisions but of habit. As the evaluation of habitual behaviour is beyond CDT's domain, CDT does

<sup>11</sup> Of course, it could be argued that this isn't always the case: that sometimes FORESIGHT overrides STAKES. Again, however, I remain unconcerned by abstract objections. Absent an argument that FORESIGHT sometimes overrides STAKES, and indeed does so in the case at hand, I take the current objection to be unconvincing.

<sup>12</sup> The word "decision" is ambiguous between: (a) a particular option (the decision *to* go to the ball); or (b) a deliberative situation (the decision *of whether to* go to the ball). Here, I utilise the latter sense. So: CDT can guide us in any deliberative situation.

<sup>13</sup> I thank a referee for raising this issue.

<sup>14</sup> At least, it does so given Bratman's account of intentions. And, remember, I am assuming the truth of this account, and exploring what follows.

not apply to question of reconsideration. Instead, as I have assume throughout this paper, it is silent in this context.<sup>15</sup> So reflection on this matter does not undermine the offered solution to *the Blade Runner's Button*. The third, and final, of our objections fails.

## 6 Conclusions

*The Blade Runner's Button* doesn't undermine CDT if options involve intention formation, and we accept Bratman's view of intentions. Insofar as such a view of optionhood and of intentions is plausible, this means that the proponent of CDT can happily accept this view and avoid the force of Egan's objection. So much the better for CDT.

So much the better, too, for discussions of optionhood. After all, one might dispute the importance of such discussions. Does it really matter, one might ask, whether decision theory endorses the decision to exercise or the act of exercising? Either way, it endorses the exercising option and this might seem to be all that matters. On this basis, it might be thought that the question of optionhood is a fringe question, of no interest to broader decision-theoretic discussions.

However, the above argument reveals that this is too fast: the nature of optionhood has broader implications. After all, if options are actions, then Egan's objection to CDT succeeds in its initial form. On the other hand, if options involve the formation of intentions then, as demonstrated above, Egan's objection fails. So Egan's objections cannot be evaluated without an account of optionhood.

**Acknowledgements** For helpful feedback, thanks to Arif Ahmed, Richard Holton, Gary Mullen, and various anonymous reviewers.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Arntzenius, F. (2008). No regrets, or: Edith Piaf revamps decision theory. *Erkenntnis*, 68, 277–297.
- Bratman, M. (1987). *Intention, plans, and practical reason*. Cambridge: Harvard University Press.
- Egan, A. (2007). Some counterexamples to causal decision theory. *The Philosophical Review*, 116, 93–114.
- Gauthier, D. (1997). Resolute choice and rational deliberation: A critique and a defense. *Noûs*, 31, 1–25.
- Hedden, B. (2012). Options and the subjective ought. *Philosophical Studies*, 158, 343–360.

---

<sup>15</sup> One could adopt a broader view of CDT, where this theory does evaluate habitual reconsideration behaviour. My defence will not apply to such a theory. Still, I find this broad theory implausible (for reasons gestured at in Bratman 1987, p. 73; Holton 2009, pp. 3, 70–166). So, I rest content with defending the narrow form of CDT.

- Holton, R. (2009). *Willing, wanting, waiting*. Oxford: Oxford University Press.
- Joyce, J. M. (1999). *The foundations of causal decision theory*. Cambridge: Cambridge University Press.
- Joyce, J. M. (2012). Regret and instability in causal decision theory. *Synthese*, *187*, 123–145.
- Raz, J. (1975). Reasons for action, decisions and norms. *Mind*, *84*, 481–499.