Causal Decision Theory and Decision Instability[1]

Brad Armendt

Revised Jan. 23, 2019

In the face of new information, your deliberation about what to do may change course. Sometimes this can happen when your only new information comes from your deliberation itself, and from the course it has taken so far. Sometimes deliberation's path follows turns and switchbacks, and a stable decision may be hard to find. One of the best-known illustrations of this, the story of the man who met death in Damascus, appeared in the infancy of the subjective theory of rational choice known as *causal decision theory*. Causal decision theory and treatments of decision instability have long been linked.[2] Causal decision theory is much discussed presently, and objections to it come in several forms, old and new. The theory is in excellent health, but that is not my present topic. Here I will explore the use of causal decision theory, when you deliberate about what to do in *Death in Damascus* and in similar decision problems.

A straightforward and general understanding of the scope of causal decision theory is presented here. We can call it *unadorned* causal decision theory, but really, *causal decision theory* is already a fine term. When it is applied to problems like Death in Damascus, we will find that the interplay of your rational assessments and your rational beliefs during deliberation is a fascinating topic, but not a source of difficulty for causal decision theory.

---

[1] Thanks to Shyam Nair, Brian Skyrms, Simon Huttegger, Peter Vanderschraaf, Steven Reynolds, and especially to Jim Joyce for helpful discussions of this paper. My views about topics discussed here have benefited from many conversations and exchanges with others. Particular thanks to William Harper, Christopher Hitchcock, Paul Weirich, Alan Hajek, Susan Vineberg, Danny Hintze, Jaemin Jung, and Wes Anderson. Special thanks to Jim Joyce, and most of all, to Brian Skyrms.

[2] Causal decision theory and the story of the man who met death in Damascus were both introduced in Allan Gibbard and William Harper, "Counterfactuals and Two Kinds of Expected Utility," in C.A. Hooker, J.J Leach, E.F. McClennan (eds.) *Foundations and Applications of Decision Theory* (Dordrecht: Reidel, 1978); reprinted in W.L. Harper, G.A. Pearce, R. Stalnaker (eds.) *Ifs* (Dordrecht: Reidel, 1981), pp. 153-190. Brian Skyrms' example of the *mean demon* shares the form of Death in Damascus, and it appears in his early discussion of deliberation dynamics in "Causal Decision Theory," *Journal of Philosophy* LXXIX (1982): 695-711.

In what follows, I use the Death in Damascus problem to illustrate decision instability and *deliberation dynamics*. Then I consider a purported counterexample to causal decision theory, representative of others, namely Andy Egan's *Murder Lesion* problem.[3] A simple response on behalf of causal decision theory, called the *Simple Response*, shows how Murder Lesion and similar problems fail to be counterexamples, and it clarifies the general use of the theory in problems of decision instability. I then compare unadorned causal decision theory and the Simple Response to previous treatments by Frank Arntzenius and by Jim Joyce.[4] There are differences among the three, and I recommend the unadorned theory and the Simple Response. But there is also much agreement among them, particularly in the practical import of adopting them. The present effort does not press on to consider other objections to causal decision theory, but it makes room for better discussions of their merits.

Decision instability and deliberation dynamics.

Allan Gibbard and William Harper considered the issue of stability in rational choice, illustrated by the example of the man who met death in Damascus:

> Consider the story of the man who met death in Damascus. Death looked surprised, but then recovered his ghastly composure and said, 'I am coming for you tomorrow'. The terrified man that night bought a camel and rode to Aleppo. The next day, death knocked on the door of the room where he was hiding and said, 'I have come for you'.
>
> 'But I thought you would be looking for me in Damascus,' said the man.
>
> 'Not at all,' said death "that is why I was surprised to see you yesterday. I knew that today I was to find you in Aleppo."
>
> Now suppose the man knows the following. Death works from an appointment book which states the time and place; a person dies if and only if the book correctly states in what city he will be at the stated time. The book is made up weeks in advance on the basis of highly reliable predictions. An appointment on the next day has been inscribed for him. Suppose, on this basis, the man would take his being in Damascus the next day as strong evidence that his appointment with death is in Damascus, and would take his

[3] Andy Egan, "Some Counterexamples to Causal Decision Theory," *Philosophical Review* CXVI (January 2007): 93-114.

[4] Frank Arntzenius, "No Regrets, or: Edith Piaf Revamps Decision Theory," *Erkenntnis* LXVIII (2008): 277-297. James M. Joyce, "Regret and Instability in Causal Decision Theory," *Synthese* CLXXXVII (2012): 123-145.

> being in Aleppo the next day as strong evidence that his appointment is in Aleppo.[5]

To evaluate his options with causal decision theory, the man uses states of the world that form a *K*-partition. Each such state is one he takes to be outside his causal influence, and sufficiently specific given his interests. Let K_D and K_A be 'Damascus is inscribed' and 'Aleppo is inscribed'; they form a *K*-partition. If, for example, he believes it is Damascus rather than Aleppo, and his belief $pr(K_D)$ is greater than ½, then causal decision theory endorses going to Aleppo.[6] But that decision seems unstable: when the man comes to believe he is about to go to Aleppo, he has new information that influences his other beliefs, including his beliefs about what is inscribed, about K_D and K_A. Since he takes death's appointment book to be based on reliable predictions, his anticipation that he will go to Aleppo raises his belief that Aleppo is inscribed after all, so $pr_n(K_A)$ is greater than ½, and $pr_n(K_D)$ is less, which makes Damascus the better option. But when the man anticipates he is about to choose Damascus, that newer information again influences his beliefs, so $pr_{nn}(K_D)$ exceeds ½, which makes Aleppo the better option …, and so on.

What should the man do?[7] The sense of instability in problems like *Death in Damascus* arises when he can reevaluate his options in light of information arising from his

---

[5] Gibbard and Harper, *op. cit*., pp. 185-186. The story is a variant of one told by W. Somerset Maugham in *Sheppey*, 1933, and alluded to in the title of John O'Hara's *Appointment in Samarra*, 1934. Other versions of the story are far older, dating to the ninth century and probably earlier; see 'When Death came to Baghdad,' in Idries Shah (ed.) *Tales of the Dervishes* (London: Jonathan Cape, 1967). Different cities appear in various earlier versions of the story.

[6] Here I use the K-expectation version of causal decision theory due to Skyrms, *op. cit*. Causal decision theory endorses going to Aleppo in the sense that going to Aleppo's expected utility *U* is maximal, and *U* represents his preferences. To say that causal decision theory endorses action *A* is to rely on a principle that endorses actions that maximize *U*. I follow other discussions in accepting that principle; my concern here is the proper application of the principle in extended deliberation. To say that causal decision theory endorses acting at a particular moment *t*, or when one is in a particular epistemic state *e*, is to rely on another principle. More about that below.

[7] One plausible answer is toss a coin, or adopt some internal method of randomizing his choice, thereby pursuing a mixed strategy; see William Harper, "Mixed Strategies and Ratifiability in Causal Decision Theory," *Erkenntnis* XXIV (1986): 25-36. Neither pure act (remain in Damascus, go to Aleppo) is ratifiable, but a 50-50 mixture of those acts is. A good idea, perhaps, but suppose that mixed strategies are ruled out as viable options—if the man were to use one, death would know, and would interrupt his appointment-keeping to find the man wherever he is, as in Paul Weirich, "Decision Instability," *Australasian Journal of Philosophy* LXIII (1985): 465-472. Fanciful examples aside, problems that forbid or

deliberations. This is a good setting for the theory of deliberation dynamics, where updates of your beliefs inform your continuing deliberations, leading to new assessments of your options that in turn provide reasons for further belief updating. The theory can be applied to deliberations about many sorts of decision problems, simple and complex.[8]

Deliberation dynamics applies to deliberation that takes place over time.[9] At each moment, your beliefs about states of the world (*e.g.* Damascus is inscribed) underwrite your current assessments of your options. Those assessments conform, let us suppose, to subjective rational decision theory; throughout this discussion, causal decision theory is the theory in use. If we assume that the values you attach to outcomes (*e.g.* life, death) are not shifting, then when your beliefs are stable, so are your assessments of your options. But as in the case of the man who met death, your beliefs may not be stable. At a given moment, your current assessments may give you reason to change your beliefs about what you will do, and to change your beliefs about states of the world that matter to the outcome of doing it. The new beliefs underwrite new assessments, and we can entertain the trajectories of your repeatedly-revised beliefs and assessments over time. Sometimes those trajectories may display oscillations, as we imagined in the case of the man who met death.[10]

---

penalize mixed acts thereby impose a restrictive exogenous constraint on the decision-maker's options. Having said that, however, we should be careful about 'solving' a decision problem by altering it with additional options, and offering a solution to the revised problem.

[8] See Skyrms, *op. cit*., pp. 701-706. In later work, Skyrms developed an important connection between dynamic deliberation and well-founded solution concepts for noncooperative games among Bayesian players, see *The Dynamics of Rational Deliberation* (Cambridge: Harvard University Press, 1990). In decision problems with more than two options, deliberation can be significantly more complex than in Death in Damascus. Arntzenius *op. cit.* and Joyce *op. cit.* each invoke Skyrms' deliberation dynamics in their treatments of Death in Damascus and similar problems, as we will see.

[9] Including intermittent deliberations: the offer expires on the day after tomorrow; the election is on Tuesday; a plan must be in place next week; I have one hour to make a move... When deliberation takes time, further news may arrive and sway its course. But effects of outside news are ignored here.

[10] When you deliberate about different sorts of problems, your changing beliefs and evaluations follow different sorts of paths. When causal decision theory is used to evaluate your options in a *Newcomb Problem*, for example, deliberation may yield straightforward convergence to high confidence that you will take both boxes, and that the opaque box will be empty, since an increasing confidence that you will take both boxes does not lead you to think it would be better to do otherwise.

The trajectory of your beliefs will depend on the details of your dynamics. How do your beliefs about what you will do depend on the values you attribute to each option?[11] You recognize at t that one action A looks better than its alternative B, that $U_t(A) > U_t(B)$; perhaps you also recognize the difference between them, or the ratio $U_t(A) / U_t(B)$, on some particular U_t scale. How much does that lead you to increase your belief at t_+ that you will do A, your $pr_{t+}(A)$? The answers to such questions throughout your deliberation might be given by a dynamical rule. Many such rules are possible; among them are rules that *seek the good*, according to which you raise your probabilities that you will perform actions exactly when you currently regard those actions as better than their alternatives, or more precisely, as better than the *status quo*, which is your current expectation of the outcome of the problem you are deliberating about.[12] If at some moment t your beliefs lead you to regard your options as equally good, so that $U_t(A) = U_t(B)$, then your assessments give you no reason, under dynamics that seek the good, to alter those beliefs and assessments. You are at an *equilibrium* of the dynamics. Since you then regard each action as equally good, to choose one or the other is to break a tie, or 'pick' among the tied options.

Returning to Death in Damascus, then, suppose that during his deliberation, the man is attentive to his evaluations of his options, and that they inform his beliefs about what he will soon do, and about what is inscribed in Death's appointment book. Suppose that at time t_1 during his deliberation, he regards going to Aleppo as the better action, so that $U_{t1}(A) > U_{t1}(D)$, and he realizes that he does; he then raises his belief that he will go to Aleppo, $pr_{t2}(A) > pr_{t2}(D)$, and also that death will be waiting for him there, $pr_{t2}(K_A) > pr_{t2}(K_D)$.[13] Then, when he reevaluates his options at t_2 with those new beliefs, he sees Damascus as the better action, $U_{t2}(D) > U_{t2}(A)$, which gives him reason to revise his beliefs again. Under plausible assumptions, in the Death in Damascus problem deliberation that seeks the good will eventually lead the man's beliefs to a stable equilibrium, where he sees neither act as better

---

[11] Also, how *much* do your beliefs about what you will do depend on your assessments of those values? We might explore the idea that your beliefs respond to other influences too, but here I leave that for another occasion, and suppose that the belief changes are driven only by your shifting assessments of your options.

[12] The idea of dynamics that seek the good is Skyrms', see *The Dynamics of Rational Deliberation*, p. 30. Such dynamics must also raise the sum of the probabilities of all the actions better than the *status quo*. Both Arntzenius and Joyce require that dynamics for rational agents seek the good; see Arntzenius, *op. cit.*, p. 293, and Joyce, *op. cit.*, p. 132-133.

[13] The dynamic rule expresses the change in his beliefs about what he will do, $pr_{t2}(A)$ and $pr_{t2}(D)$. Accompanying changes in his beliefs about what is inscribed, $pr_{t2}(K_A)$ and $pr_{t2}(K_D)$, satisfy Jeffrey-conditionalization if his conditional probabilities such as $pr(K_A/A)$ are stable, and there are no further complexities.

than the other.[14] At that point, his tied evaluations give him no reason to further adjust the beliefs that underlie them. At the equilibrium state in the original Death in Damascus problem, $pr_{eq}(K_A)$ and $pr_{eq}(K_D)$ are both ½, as are $pr_{eq}(A)$ and $pr_{eq}(D)$. His action will be the outcome of some way of dealing with the tie between *A* and *D*. A general feature of equilibrium states, whether your deliberation leads you to them or not, is that you see your available options as equally choiceworthy, as having equal expected utility. It may also happen that you then believe that you are as likely to do one act as the other, but that need not be so in problems that lack the symmetry of Death in Damascus.

Why should the man embark on this deliberative journey? There is at least this reason: a rational choice should be based on all of your relevant beliefs at the time you make it. So, *if* you believe at time *t* that death is more likely to go to Aleppo than to Damascus, $pr_t(K_A) > pr_t(K_D)$, your evaluations at *t* of your options, $U_t(A)$ and $U_t(D)$, must use those beliefs. Or, to put it another way, a rational decision theory, such as causal decision theory, is properly used only when those evaluations do so. Is it incumbent upon you to possess such beliefs in the midst of deliberation? We will return to that question soon.

The original version of Death in Damascus is a symmetric problem, but asymmetric versions are easily given; just add an incentive against travel that makes the outcomes of staying in Damascus a little better than the corresponding outcomes of traveling to Aleppo.[15] Or, imagine that death's appointment book more reliably predicts the traveler's presence when he is in one city than when he is in the other.

Decision instability has become prominent in work on causal decision theory. One reason is that it displays the wider scope of the theory, beyond problems where causal dominance reasoning applies. Another is that problems displaying instability have been offered as *counterexamples* to causal decision theory.

---

[14] In general, given sufficient time, we can expect convergence to equilibrium from reasonable starting points. For Death in Damascus, continuous deliberation dynamics that seek the good are guaranteed to converge to equilibrium, but they will not display the oscillations I have described. Discrete-time dynamics that seek the good may well display oscillations; with plausible properties such as a dampening in the learning over time, convergence to equilibrium can be guaranteed. See Skyrms, *Dynamics*, *op. cit.*, and William Harper, "Decision Dynamics and Rational Choice," forthcoming in Billy Dunaway and David Plunkett, eds., *Meaning, Decision, and Norms: Themes from the Work of Allan Gibbard*, Maize Books. See also Greg Lauro and Simon Huttegger, "Decision Dependence and Causal Decision Theory," manuscript.

[15] Reed Richter, "Rationality Revisited," *Australasian Journal of Philosophy* LXII (1984): 392-403.

Murder Lesion and the Simple Response.

Andy Egan challenged causal decision theory with a set of examples that he judged to be counterexamples to the theory, and his challenge has received wide attention. One of the examples is the *Murder Lesion* problem:

> Mary is debating whether to shoot her rival, Alfred. If she shoots and hits [S & H], things will be very good for her. If she shoots and misses [S & M], things will be very bad. (Alfred always finds out about unsuccessful assassination attempts, and he is sensitive about such things.) If she doesn't shoot [~S], things will go on in the usual, okay-but-not-great kind of way. Though Mary is fairly confident that she will not actually shoot … she thinks that it is very likely that if she were to shoot, then she would hit [$S \mathrel{\square\!\!\rightarrow} H$]. So far, so good. But Mary also knows that there is a certain sort of brain lesion that tends to cause both murder attempts and bad aim at the critical moment. If she has this lesion, all of her training will do her no good—her hand is almost certain to shake as she squeezes the trigger. Happily for most of us, but not so happily for Mary, most shooters have this lesion, and so most shooters miss. Should Mary shoot? [notation added] [16]

Following Egan, let the utility of shooting and hitting (S & H) be 10, the utility of shooting and missing (S & M) be -10, and the utility of not shooting (~S) be 0 throughout.[17] Mary's initial beliefs are that she is unlikely to shoot, $pr_i(S) < .5$. She also thinks that if she did, she would hit, $pr_i(S \mathrel{\square\!\!\rightarrow} H) > .5$. Her belief in that causal conditional is dependent on whether or not she shoots, since shooting is correlated with having the lesion; so $pr_i(S \mathrel{\square\!\!\rightarrow} H / S) < .5$. However, her initial unconditional belief in that conditional is high, as just specified, since she initially thinks S is unlikely.

With those initial beliefs, a causal decision theory calculation will yield $U_i(S) > U_i(\sim S) = 0$, since the better outcome of S, namely S & H, is weighted by the high probability $pr_i(S \mathrel{\square\!\!\rightarrow}$

---

[16] Andy Egan, "Some Counterexamples to Causal Decision Theory," *Philosophical Review* CXVI (January 2007): 93-114, p. 97.

[17] There is symmetry in these payoffs, but perhaps not in the beliefs: *most* shooters have the lesion, but whether the same proportion of non-shooters lack it is unsaid; it's *nearly certain* that those with the lesion miss; in light of her training, Mary thinks it's *very likely* that she would hit. Nothing I say here depends upon the problem being as symmetric as Death in Damascus. In what follows, we might identify states of a K-partition (have the lesion, don't have the lesion) and use the K-expectation version of causal decision theory, as before. But here I follow Egan, who uses causal conditionals as in the Gibbard-Harper version of causal decision theory.

H), while the worse outcome *S* & *M* is weighted by the low probability $pr_i(S \,\square\!\!\rightarrow M)$. So causal decision theory endorses shooting. Egan regards that as a flawed endorsement:

> It's irrational for Mary to shoot. … In general, when you are faced with a choice of two options, it's irrational to choose the one that you confidently expect will cause the worse outcome. Causal decision theory endorses shooting ... In general, causal decision theory endorses, in these kinds of cases, an irrational policy of performing the action that one confidently expects will cause the worse outcome. The correct theory of rational decision will not endorse irrational actions or policies. So causal decision theory is not the correct theory of rational decision.[18]

The act of shooting is intuitively irrational, Egan says, and widely judged to be so.

> …we have (or at least my informants and I have) clear intuitions that it's irrational to shoot or to press, and rational to refrain in *The Murder Lesion*…[19]

There is a response to Egan's view of the example. Egan's case for the irrationality of causal decision theory's endorsement (that Mary shoot) is the intuitive irrationality of shooting. No basis for the intuition is offered, but it is not hard to feel, nor hard to explain. What is happening? Mary begins with the beliefs that she lacks the lesion, and that shooting would be effective; based on those beliefs causal decision theory endorsed shooting.[20] That is the right endorsement given her beliefs and values at that time. But in preferring shooting, she probably has the lesion and will very likely miss. So Mary comes to confidently expect, and we who contemplate her problem come to confidently expect, that shooting will cause the worse outcome. That is what the first step in the deliberative process tells her. What is Egan's intuition, if not the result of taking that step? At that point, however, when Mary has *that* belief, causal decision theory endorses *refraining*; that is what Mary's current utilities will tell her. Egan applies the endorsement that causal decision theory makes at one time (shoot) to a decision at a later time, after Mary's beliefs have changed, and he sees a flaw where there is none. The error is in the supposition that causal decision theory is forever committed to its endorsement under Mary's initial beliefs.

---

[18] Egan, *op. cit*., p. 97-98. Phrases referring to a different example omitted.

[19] *Ibid.*, p. 98.

[20] It's worth remarking that users of causal decision theory are no more prone to murder, premature death, disease, or psycho-killing than anyone else. The window-dressings of our examples should be more varied; outcomes might be prizes or penalties, large or small, rather than death. Many *games* exhibit instability: Battle-of-the-Sexes-with-a-Twin, for example. The theoretical issues apply to small stakes as well as large, a point worth remembering when deliberation has a cost.

> The resulting theory enjoins us to *do whatever has the best expected outcome, holding fixed our initial views about the likely causal structure of the world*. The following examples show that these two principles come apart, and that where they do, causal decision theory endorses irrational courses of action. (emphasis Egan)[21]

What causal decision theory really endorses is what has the best expected outcome, given our *current* views about the likely causal structure of the world.[22] This applies to us in the first person, as deliberators (use our current beliefs), and in the third person, as judges of what causal decision theory says about others (use their current beliefs). Egan's argument suffers from a mistake about what causal decision theory endorses.[23]

A second issue is that an intuition that refraining is uniquely rational has doubtful reliability. We are invited to deliberate a little bit, but not very far, about what to do, and to stop the deliberation at an arbitrary point, with no motivation given for stopping there. If Mary correctly assesses her options at that point, when she thinks she has the lesion, refraining is better. But that provides her reason to think that, as a refrainer, she likely lacks the lesion, and that belief makes shooting the better option after all (according to her), … and so on. Even if the mistake about causal decision theory's endorsements were absent, the example

---

[21] Egan, *op. cit.*, p. 96.

[22] Egan actually states the principle correctly later in his paper in a different context, p.102, but it is clear that he relies on the incorrect version throughout the paper. Without it, what is the purported counterexample?

[23] I say that the result of the first deliberative step in Murder Lesion explains Egan's intuition that refraining is rational and shooting is irrational. But accounting for intuitions about particular examples is an uncertain project, and there may also be other intuitions in play. One might be uneasy about instability itself, for example. Why use a theory that makes an endorsement, then retracts it, then reinstates it, and so on? The retraction suggests that the endorsement should not have been made in the first place. But the second recommendation is not a retraction of the first endorsement, it is an update, in light of new information. It is one thing to have an advisor who wavers in his advice for no discernable reason, another thing when the advice changes as he receives a stream of relevant information. At some point, though, you will stop listening. In any case, it is hard to see how an aversion to instability points to the rationality of one specific option rather than the other, how it points to refraining as rational, and shooting as not. Egan's intuition, and his informants', appears to be different. Joyce suggests that framing and loss aversion may be at work, and that could be so; see Joyce, "Regret and Instability," *op. cit.*, p. 135. But Egan's example emphasizes what Mary comes to confidently expect, rather than her great aversion to the worse outcome… I think the best explanation of the intuition Egan describes is the one given with the Simple Response. Whether or not that is so, the error about what causal decision theory endorses, when your rational beliefs shift during your deliberation, remains. Thanks to an anonymous referee for raising questions concerning intuitions about these problems.

would at best indicate a problem with joining causal decision theory to a special unmotivated assumption about when deliberation must end. It establishes no problem for causal decision theory, which is consistently a good guide to rational action. So says this line of response; let us call it the *Simple Response*.

What, then, does causal decision theory endorse in the *Murder Lesion* problem? If you have Mary's initial beliefs and deliberate no farther, it endorses shooting. If you have a different initial belief, that you probably have the lesion, it endorses refraining. If you have access to and appreciation of your deliberative states, and your beliefs and assessments interact in extended deliberation, causal decision theory makes a succession of endorsements at each stage of your self-reflecting dynamical deliberation. The endorsements may change, but each one is correct for your beliefs and values at the moment it is made. Your deliberation may end in a variety of ways. You may get tired, you may have other things to do, the world may interrupt, or you may reach equilibrium. If your deliberation ends in action, the rational action to perform is the one endorsed by your current beliefs, when deliberation ended. Recall that if you reach equilibrium, you see your options as equally worthy, and you shoot or refrain by dealing with the tie. There is no single action, for every deliberator, that use of causal decision theory requires or leads to. So goes the Simple Response to Murder Lesion and to similar purported counterexamples. And in general, when deliberation is unstable and wavers between options, so says unadorned causal decision theory.

Does the equilibrium rule?

According to the preceding account, causal decision theory delivers assessments of your options at each moment of your deliberative process, whether or not you are at an equilibrium. In the company of a principle that endorses actions with maximal causal expected utility U, at each moment of your deliberative process, unadorned causal decision theory endorses the action or actions that currently have maximal U. Further constraints on use of causal decision theory are neither imposed nor needed. In particular, no requirement is made that you must reach an equilibrium, or that you must look to an equilibrium prior to reaching it, in order for causal decision theory to endorse one, or some, of your options. Of course, if you do reach stable equilibrium, causal decision theory will deliver stable evaluations of your options, and will equally endorse those tied with maximal U_{eq}.[24]

[24] Since application of causal decision theory does not depend on your arriving at equilibrium, the issue of whether your starting point and your dynamics will lead you to equilibrium is interesting, but not crucial to using the theory. This idea, and more, is also expressed by Skyrms, in *Dynamics*, *op. cit.*, pp. 36-37: "It is possible—perhaps likely—that deliberation will have reached an equilibrium by the moment of truth, in which case her decision will be a best response. On the other hand, in the absence of special knowledge, it is no more likely if the moment of truth arrives before equilibrium that she will make a worse

Contrasting accounts given separately by Frank Arntzenius and Jim Joyce disagree. Each of their accounts offers an adornment to causal decision theory, and each argues that the equilibrium state is the unique perspective from which to ascertain your rational action. In a nutshell, Arntzenius' view is that, in decision problems like those we are considering, you should be in the equilibrium state when making your decision. Joyce's view is that epistemic rationality will lead your deliberation to the equilibrium state, and it is only then that you are in a satisfactory epistemic position to make assessments that should guide your choice. Joyce would say that what causal decision theory endorses all along, even before you reach equilibrium, is what it endorses then, when your epistemic state is satisfactory and you rank each viable option equally.

Let us look briefly at Arntzenius' treatment first. It is developed in the company of his suggestion for understanding how you might perform a mixed strategy, an option that is a probabilistic mixture of your pure options (go to Aleppo, remain in Damascus). Arntzenius associates mixed acts with states of belief; to act when your rational belief that you will go to Aleppo, $pr(A)$, is x and your belief that you will remain in Damascus, $pr(D)$, is $1\text{-}x$, is to perform the mixed act $(xA, (1\text{-}x)D)$.[25] After showing that deliberation that seeks the good will eventually arrive at the equilibrium state, Arntzenius asks,

> Must one really model a rational person as a deliberator who changes his credences during the deliberation? No, one need not. Indeed it is a little bit awkward to do so. After all, if one is ideally rational, then how could there be any stage at which one has the 'wrong' credences?[26]

His second question might suggest sympathy for a treatment like the one I am advocating, but Arntzenius instead recommends that we set aside non-equilibrium beliefs:

> So, as long as we are idealizing, let us simply say that a rational person must always be in a state of deliberational equilibrium. The dynamical model of deliberation that

---

response rather than a better one. The present expected utilities just calculated may not be the ones which will obtain at the moment of truth, but they are in a sense the decisionmaker's best estimate of them." In earlier work, Weirich considers, but too quickly rejects, the idea that causal decision theory applies at each moment of extended deliberation. See Weirich, *op. cit.*, pp. 466-467.

[25] Performance of the mixed act yields one or the other of the pure options. The connection between your equilibrium beliefs and the option you take raises interesting questions; I save them for another occasion.

[26] Arntzenius, *op. cit.*, p. 294.

> I gave can be taken to merely amount to a crutch to make us realize that there always exists a state of deliberational equilibrium.[27]

This, then, is the adornment to causal decision theory that Arntzenius advocates: a requirement that you must be in the equilibrium state in order to rationally assess your options. It remains true that at equilibrium, you will see your viable options as equally choiceworthy, and you will act by dealing with the tie. Arntzenius advocates dealing with the tie by performing the appropriate mixed act. His requirement seems to entail that you must hold the specific beliefs that constitute an equilibrium for your dynamics and decision problem, though ascertaining what those beliefs are, when your problem has asymmetric conditions and payoffs, is a nontrivial matter in general. Without further exploring that and other points, though, we can see that this amounts to an additional constraint, or adornment, to causal decision theory.

Let us turn next to Joyce's treatment. Consider a problem in which you have easy access to your beliefs and preferences during deliberation. Joyce argues that when free information that matters to your decision is available, you must learn it before deciding. His *Full Information* constraint requires, we might say, that you should base your evaluations on all of the *current evidence in easy reach*. Proper use of causal decision theory incorporates all of the free information that deliberation provides, and when instability is present, new information is always available until you reach equilibrium. Deliberation may take time to get there. Throughout that time, the actions that causal decision theory endorses are the ones you (will) see as best when you are at equilibrium and they are tied. The way to deal with the tie is to break it, and causal decision theory's only endorsement for the problem is to 'pick' (choose via a tie-break) one of your options.[28]

---

[27] *Ibid*. Arntzenius argues for the existence of the equilibrium under continuous dynamics that seek the good. The decision problem he explicitly considers, *Psycho Johnny*, is based on one of Andy Egan's examples. Johnny has two options, and its complexity is similar to that of Death in Damascus. More generally, in problems with more than two available options, the existence of stable equilibria under a given dynamical rule is a complex issue and not always guaranteed. See Lauro and Huttegger, "Decision Dependence," *op.cit.*, Section 4.

[28] More generally, to pick one of the options that is viable at equilibrium. In decision problems with many available actions, it may happen that at equilibrium you exclude some of them as sub-optimal, while you see others as tied with maximum expected utility. See Joyce, "Regret and Instability," *op. cit.*, pp. 126-127 and 132-134, and James M. Joyce, "Deliberation and Stability in Newcomb Problems and Pseudo-Newcomb Problems," in Arif Ahmed (ed.), *Newcomb's Problem* (Cambridge: Cambridge University Press, 2018), pp. 149-150.

Stability is an important feature of a stable doxastic equilibrium. Why think that rationality is too? Deliberation dynamics establishes a connection, when it can be shown that a sequence or flow of rational evaluations and rational belief changes leads to the equilibrium state. Joyce fills in the story with an account of the rational belief changes: their occurrence follows from your conformity to *Full Information*. With the addition of that epistemic requirement, rational deliberation may propel you to the equilibrium state, if your self-aware deliberation is cost-free. But, while Full Information may provide excellent advice, it is silent when rational deliberation is not free of cost, and it is an adornment to causal decision theory.

There is no great tension between Arntzenius' and Joyce's accounts and the one I have given.[29] Use of unadorned causal decision theory provides guidance throughout your deliberation, and its guidance is tied to your equilibrium beliefs and assessments when that is where you are. Arguments given by Skyrms, Arntzenius, Joyce, and Harper show that you will arrive there through rationality, alertness, and curiosity, when your rational deliberation is persistent and uncurtailed. It may be debated whether you misapply causal decision theory if you act before reaching it. It is surely an idealization to regard continued deliberation as cost-free (think of opportunity costs), and when it is not, acting so as to maximize your U at the end of a truncated deliberation can well be rational, contrary to the letter of Arntzenius' and Joyce's accounts.[30] But here it is worth keeping in mind that

---

[29] In one respect, there is a difference. When you eventually do act A in an unstable decision problem, you will have grounds for regretting you did so. Arntzenius, *op. cit.*, counts *foreseeable* regret, if you have it, against the rationality of the act; such foresight is avoided in the equilibrium state. If foreseeable regret is possible in a deliberation truncated at t when causal decision theory is correctly applied (that is, when current beliefs are fully used), the account I am advocating is more sanguine about foreseen regret, from which no pure action would be immune; your U_t is taken to capture your criteria for rational action. On this point, I am in agreement with Joyce, "Regret and Instability," *op. cit.*, pp. 142-143.

[30] Another way of looking at it is that such cases are exceptions to Joyce's account, which focuses on more idealized agents, rather than contrary to it. Joyce would agree that when further deliberation is sufficiently costly, it can be rational to act before reaching equilibrium. There is room for a more general account of extended deliberation, in which you repeatedly compare doing one act, doing another, and doing the act of seeking more information and then making a later decision (to act one way or another, or to seek again and then decide, etc.). Death in Damascus as described so far may be seen as embedded in a richer decision problem concerning when to do A or do D, and when to seek more information before doing either. When is it rational to seek more information? When the expected utility of further deliberation and its outcome is no less that the expected utility of acting now. A full theory of such assessments is beyond the scope of this paper; see Skyrms *Dynamics*, *op. cit.*, chapter four "The Value of Knowledge," especially pp. 101-106, for relevant discussion. But no theory will say that rationality always requires further deliberation, whatever its cost in

subjective rational decision theory interests us from both third-person and first-person points of view: as a theory that explains the choice-worthiness of actions for a decision-maker in light of her relevant beliefs and desires, and as a tool that can help us ascertain what to do, as we reflect on our own relevant beliefs and desires. When we see deliberation as a dynamic process where act assessments produce new evidence, and so on, the first-person perspective is in the foreground. But it is useful to know from the third-person perspective that in the equilibrium state, the best actions are equally good.

Heuristics can often help us, in first person deliberation, find what is rational for us do. When we understand the path to equilibrium in deliberations that are unstable, and we realize that the recommendation at equilibrium is to deal with a tie, we seem to have an excellent heuristic for deciding the problem. Using a tiebreaker is a fine way to break a tie. So is performing a mixed act of the tied options; if mixed acts are not somehow forbidden, each way is as good as the other. So here is a heuristic: just go ahead and pick.[31] Our study of rational decision making pays off! One may think that heuristics have second-class status as guides to rational action, but I think there is a serious point here. In general, decision instability may arise in problems that are asymmetric, with outcome-values that do not stand in numerically simple relationships to each other, and with causal tendencies that bear similarly non-simple relationships to each other. When that is so, a deliberator who is under a rational obligation to locate the deliberational equilibrium *via* the beliefs she would hold in that state has a very challenging task. It is hard to see how Arntzenius' and Joyce's treatments can be practically applied, except in very simple problems, unless a simplifying heuristic is available.[32] Fortunately, the heuristic to just pick among your options is simple, and it is well-supported by the arguments we have considered. It is a heuristic that all three accounts can endorse.[33]

---

resources or opportunity, and everyday experience tells us that it is often best to get on with things, and act.

[31] This is a point that has often been noticed, at least in conversation; I do not claim it as mine.

[32] In complex problems with many options, this heuristic may need help from other(s) that exclude the options that are not viable at equilibrium.

[33] You have to recognize that a heuristic applies in order to use it, and it probably takes some amount of deliberation to do so. If you have encountered similar problems, or know examples in the philosophical literature, you may recognize where deliberation will go. Others less familiar with such problems may deliberate to the equilibrium. For many of us, a couple of steps down the deliberative path conveys the character of problems like Death in Damascus; when that happens to you, and you see the relevance of *Just pick!*, use it. As with other good heuristics, there is no guarantee that you will always recognize when this one applies, or that cues will never lead you to think it applies when it doesn't. That is a chance we take, but good heuristics are not thereby useless to rational agents. Poker games

Conclusion

Unadorned causal decision theory is based on the straightforward idea that causal decision theory provides expected utilities of your options at each moment you entertain them. Its use does not depend upon the existence of a stable deliberational equilibrium. It does not insist that only one perspective is suitable for rational decision, and remain silent, or speak illegitimately, until you achieve it. Nor must your evaluations of options depend on information you currently lack, however advisable it may be that you seek such information. But the dynamical reasoning that supports the rationality of equilibrium choices, and that was used to argue for the other accounts, remains important. When rationality, or rationality plus alertness and curiosity, is unimpeded, deliberation plays out to the equilibrium, where it ends by breaking ties. The simplifying heuristic (just pick!) aids users of unadorned causal decision theory as well as users of its variants. Truncated deliberation yields a different endorsement (*do A*, or *do D*…, depending on the point of truncation) than does deliberation that reaches equilibrium (*pick between A and D*), but given the doxastic states in which the endorsements are made, neither one is counterintuitive. Whatever course deliberation takes, causal decision theory is consistently a good guide to rational action.

among rational players would be slow-moving affairs if that were so. Thanks to an anonymous referee for raising concerns about a previous presentation of the heuristic. The main point, once again, is that however deliberation ends, whether at equilibrium or not, use of unadorned causal decision theory is appropriate at each step along the way.