Face Death with Indifference

J. Dmitri Gallow

Causal decision theorists say to pursue acts which you expect to improve things, and avoid acts which you expect to make matters worse (§1). There are some choices in which, no matter how you act, you will, after acting, expect that an alternative act would have made things better. The classic example from Gibbard & Harper (1978): you must choose between traveling to Damascus and traveling to Aleppo. Death is very good at predicting your choice, so you expect Death to find you no matter where you go, though your choice does not affect Death's destination. So: if you decide to go to Aleppo, then you will expect Death to await in Aleppo, and therefore, you will expect that going to Damascus would save your life. If you decide to go to Damascus, then you will expect Death to await in Damascus, and therefore, you will expect that going to Aleppo would save your life. For all I've said so far, there is little to recommend one city over the other; but it is natural to think asymmetries can break the tie. If the bars are better in Aleppo, and if Death is more likely to find you if you go to Damascus than he is if you go to Aleppo, then all else being equal, going to Aleppo is required (§2). This thought is very tempting; I once thought it obviously correct. But I've been persuaded it is wrong. Neither city is a required destination—even if the bars are better in Aleppo, and even if Death is less likely to find you there. The reason: causalists who say that options like Aleppo are required in cases like these are committed to absurd consequences (§3). In particular, they will violate weak versions of the *independence of irrelevant alternatives* (§3.1) and *normal-form extensive-form equivalence* (§3.2). These violations in turn lead to irrational and exploitable behavior like paying to have options presented to you in a certain order, and paying to change your decision once it's been made, for no apparent reason at all (§3.3). If this behavior is irrational, then it is irrational for causalists to shun Damascus for its worse bars and better predictability. So I conclude that neither city is a required destination—that we should face Death with indifference.[1]

Draft of January 9, 2019; Word count: 8,153
Comments appreciated ✉: jdmitrigallow@pitt.edu

[1] Strictly speaking, my conclusion is merely that neither city is uniquely permissible; this could be because we should be indifferent between them, but it could also be because the options are incomparable, or because neither is permissible.

1 Causal Decision Theory

1.1. Decision Problems. I will assume that, when you are making a choice, you have some set of available *acts* $\mathcal{A} = \{A_1, A_2, \ldots, A_N\}$ between which you must choose. That is: you must choose one and only one of the acts. When making this choice, there is some set of *states of nature* $\mathcal{K} = \{K_1, K_2, \ldots, K_M\}$, which, for all you know, may obtain. Exactly one of the K_i obtains, though you know not which; nor are you in any position to influence which obtains. Though you do not know which K_i obtains, you do have opinions, represented with a probability function Pr, defined over both $\mathcal{A}$ and $\mathcal{K}$. Finally, you have a *value function* $\mathcal{V}$, which tells you how valuable selecting each act would be, in each state of nature.

I assume that your value function satisfies the following identity, for any proposition ϕ and any partition $\{\psi_i\}_i$,[2]

$$\mathcal{V}(\phi) = \sum_i \Pr(\psi_i \mid \phi) \cdot \mathcal{V}(\phi\psi_i)$$

In particular, for any state of nature $K \in \mathcal{K}$ and the partition $\mathcal{A}$,

$$\mathcal{V}(K) = \sum_{A \in \mathcal{A}} \Pr(A \mid K) \cdot \mathcal{V}(AK)$$

And likewise, for any action $A \in \mathcal{A}$ and the partition $\mathcal{K}$,

$$\mathcal{V}(A) = \sum_{K \in \mathcal{K}} \Pr(K \mid A) \cdot \mathcal{V}(AK)$$

1.2. Newcomb. Some—known as *evidential decision theorists*—think that this quantity, $\mathcal{V}(A)$, provides a measure of the *choiceworthiness* of an act A. Causal decision theorists disagree, because of cases like the following:

> Newcomb
> You are on a game show. Before you are two boxes, labelled 'L' and 'M' (for 'less' and 'more'). You may take one, and only one, of the boxes. Money was placed in the boxes on the basis of a reliable prediction. If it was predicted that you would take L, then \$100 was placed in box L, and \$110 was placed in box M. If it was predicted that you would take M, then \$0 was placed in box L and \$10 was placed in box M. These predictions are 90% reliable—that is, conditional on you selecting box X, the chance that it was predicted that you would select X is 90%. But nothing you do now will affect how much money is in the boxes.

[2] Notation: where ϕ and ψ are propositions, I write $\ulcorner\phi\psi\urcorner$ for the conjunction of ϕ and ψ. For our purposes, a *partition* is a set of propositions such that it must be that exactly one of the propositions in the set is true.

$$\begin{array}{c|cc} \mathcal{V}(RC) & K_L & K_M \\ L & 100 & 0 \\ M & 110 & 10 \end{array} \qquad \begin{array}{c|cc} \Pr(R|C) & L & M \\ K_L & 0.9 & 0.1 \\ K_M & 0.1 & 0.9 \end{array}$$

Table 1: Values and Probabilities for Newcomb. The matrix on the left shows the value of taking the row action R while in the column state C, $\mathcal{V}(RC)$. The matrix on the right shows the probability that you are in the row state R, given that you've taken the column act C, $\Pr(R \mid C)$.

We can represent this decision problem with the two matrices shown in table 1. There are two relevant states of nature. Either it was predicted that you would take box L, 'K_L', or it was predicted that you would take box M, 'K_M'. I suppose that your values are linear in dollars, so that your values for each act in each state are as shown in the $\mathcal{V}$-matrix on the left of table 1. The matrix on the right says: given that you choose box L, you're 90% likely to be in state K_L and 10% likely to be in state K_M. And, given that you choose box M, you're 10% likely to be in state K_L and 90% likely to be in state K_M.

In Newcomb, the value of L exceeds that of M, since

$$\begin{aligned} \mathcal{V}(L) &= \Pr(K_L \mid L)\mathcal{V}(LK_L) + \Pr(K_M \mid L)\mathcal{V}(LK_M) \\ &= 0.9 \cdot 100 + 0.1 \cdot 0 \\ &= 90 \\ \text{while} \quad \mathcal{V}(M) &= \Pr(K_L \mid M)\mathcal{V}(MK_L) + \Pr(K_M \mid M)\mathcal{V}(MK_M) \\ &= 0.1 \cdot 110 + 0.9 \cdot 10 \\ &= 20 \end{aligned}$$

So evidential decision theorists advise you to take box L. But notice that, no matter what was predicted, taking box M will get you strictly more money. In each state of nature, taking box M will get you \$10 more than taking L will. Notice also: if you were to learn which prediction was made, the value of M would exceed the value of L, and the evidential decision theorists would advise you to take M—*no matter what* you learned. If you were to learn K_L, the value of M would exceed the value of L by 10. And if you were to learn K_M, the value of M would exceed the value of L by 10. Evidential decision theorists therefore violate a principle of deontic reflection: they recommend acts which they know their better informed, future selves will wish they had not chosen.[3]

We may dramatize this violation of deontic reflection in the case of Newcomb. Suppose that the evidential decision theorist faces Newcomb, and they are playing, not for themselves, but rather for a poor orphan boy, Oliver. While

[3] See Arntzenius (2008)

they are not allowed to look in the boxes, Oliver is. He is there with them as they choose. He is allowed to offer the evidentialist advice about which box to choose, but he is not allowed to tell them the contents of the boxes. He looks inside, and says: 'Please, choose box M'. (Of course he does—the evidentialist knew that's what he'd say, no matter what he saw). The evidential decision theorist ignores Oliver's advice, and chooses box L instead. They tell him: 'If you were able to tell me what the boxes contain, I would agree with you, and I would choose M, no matter what you told me. But, since you haven't told me what's in the boxes, I must take box L.' At this point, the producers of the game show—who are really pulling for Oliver—intervene. They say: 'If you allow him, Oliver may tell you what the boxes contain.' The evidential decision theorist does not allow him. They say: 'If I allow you to tell me what's in the boxes, then I will end up taking box M. But currently, I think that's worse than choosing L. So I think it's better for me to not know.' The producers try a different tack. They say: 'Alright, if you don't listen to what Oliver has to say about the contents of the boxes, then we'll take \$60 away from whatever Oliver wins (perhaps leaving him with a bill to pay).' The evidential decision theorist knows that, if they listen to Oliver, they'll take box M. The value of box M is \$20. On the other hand, if they don't listen, they'll take box L. The value of box L is \$90. Minus the \$60 lost by not listening, not listening gets Oliver a net \$30. So, in order to keep Oliver quiet, they'll take \$60 away from him.[4]

Imagine yourself as Oliver, pleading with the evidential decision theorist to take the box that you can see contains an additional \$10. They are choosing only for your benefit. You are telling them that M is the box which will most benefit you. They believe you. They know that box M will benefit you the most. Yet they refuse to take it. They moreover refuse to take the information you are trying to give them, even though they know that this information is not in any way misleading, that it will teach them what is objectively in your best interest, and that their learning this information is objectively in your best interest. To keep themselves from learning this information, they are willing to take \$60 away from you—though, again, their only concern is maximizing *your* welfare. Does this look like the behavior of a rational agent? The causal decision theorist thinks not, and I agree. And so I think that $\mathcal{V}$ does not give an adequate measure of the choiceworthiness of an act.

1.3. Utility and Improvement. According to the causal decision theorist, we should measure the choiceworthiness of an act, A, not by its *value*, $\mathcal{V}(A)$, but rather by what we may call its *utility*, $\mathcal{U}(A)$, where

$$\mathcal{U}(A) \stackrel{\text{def}}{=} \sum_K \Pr(K) \cdot V(AK)$$

[4] See Wells (forthcoming)

The difference between $\mathcal{V}$ and $\mathcal{U}$ is that, while $\mathcal{V}(A)$ measures the *conditional* expected value of the act A (conditional on A being chosen), $\mathcal{U}$ measures the *unconditional* expected value of A. That is, it measures the expected value of A from the perspective you currently occupy, and not the perspective you *will* occupy, once you have chosen.

One important feature of the measure $\mathcal{U}$ is that its values can depend upon how confident you are that you will end up selecting each available act (call these your *act probabilities*). For instance, in Newcomb, your probability that it was predicted that you would take L or M depends upon how likely you think you are to end up taking L or M. If l is your probability that you will take L (and therefore, $1-l$ is your probability that you will take M), then

$$\begin{aligned} \Pr(K_L) &= \Pr(K_L \mid A)l \ + \ \Pr(K_L \mid M)(1-l) \\ &= 0.8l + 0.1 \\ \text{and} \qquad \Pr(K_M) &= \Pr(K_M \mid L)l \ + \ \Pr(K_M \mid M)(1-l) \\ &= 0.9 - 0.8l \end{aligned}$$

For this reason, the utility of taking L and M will similarly depend upon how likely you are to take L or M:

$$\begin{aligned} \mathcal{U}(L) &= \Pr(K_L)100 + \Pr(K_M)0 \\ &= 80l + 10 \\ \text{and} \qquad \mathcal{U}(M) &= \Pr(K_L)110 + \Pr(K_M)10 \\ &= 80l + 20 \end{aligned}$$

So: as your probability for choosing L goes up, so too does the *utility* of choosing L. And as your probability for choosing M goes up, the *utility* of choosing M goes down. It is for this reason that the evidential decision theorist advises you to choose L. The evidential decision theorist says to choose whichever option will give you the best news about its utility. But, no matter what your probability for choosing L is, the utility of M will exceed the utility of L. It is for this reason that the causal decision theorist advises you to choose M. The causal decision theorist says to choose whichever option has the highest utility, and to ignore any news the performance of that act will provide about its own utility.

For my purposes, it will be convenient to introduce another, equivalent, formulation of causal decision theory. According to this formulation, CDT says to choose whichever act is expected to do the most to *improve* the world. If you are in state K, then $\mathcal{V}(AK) - \mathcal{V}(K)$ is a measure of the extent to which the act A improves the world. You don't know which state you're in, but the expectation

$$\mathcal{I}(A) \stackrel{\text{def}}{=} \sum_K \Pr(K)[\mathcal{V}(AK) - \mathcal{V}(K)]$$

tells you the degree to which you *expect* A to improve the world.[5,6] If $\mathcal{I}(A)$ is positive, then A is expected to improve the world. If $\mathcal{I}(A)$ is negative, then A is expected to make the world worse. Maximizing expected improvement is equivalent to maximizing utility. Given any two acts, X and Y, the utility of X is at least as great as the utility of Y iff the expected improvement of X is at least as great as the expected improvement of Y: $\mathcal{U}(X) \geqslant \mathcal{U}(Y)$ iff $\mathcal{I}(X) \geqslant \mathcal{I}(Y)$.[7] So we may understand causal decision theory as saying that you should always choose the act which you expect to do the most to improve the world.

Just as you may evaluate the expected improvement of an act from the perspective you currently occupy, so too may you evaluate the expected improvement of an act, A, from the perspective you would occupy, were you to choose another act, B. (I mean: the perspective you would occupy *immediately* after choosing B, before learning anything else.) From this perspective, you will have learned that you have performed B, so your probability for each state K will be $\Pr(K \mid B)$. If you are in state K and you have chosen B, then $\mathcal{V}(AK) - \mathcal{V}(BK)$ is a measure of the extent to which performing A instead would improve the world. Immediately after choosing B, you wouldn't know which state you're in, but the expectation

$$\mathcal{I}(A \mid B) \stackrel{\text{def}}{=} \sum_K \Pr(K \mid B)[\mathcal{V}(AK) - \mathcal{V}(BK)]$$

tells you the degree to which you would *expect* A to improve the world. So $\mathcal{I}(A \mid B)$ gives a measure of the degree to which you would expect A to improve things, were you to choose B. Given the quantities $\mathcal{I}(A \mid B)$, for each pair of acts A and B, we may calculate $\mathcal{I}(A)$ as follows.[8]

$$\mathcal{I}(A) = \sum_{B \in \mathcal{A}} \mathcal{I}(A \mid B) \cdot \Pr(B)$$

---

[5] $\mathcal{I}$, unlike $\mathcal{U}$ and $\mathcal{V}$, is a *ratio* scale, rather than an *interval* scale. This is as it should be, for there is a theoretically significant zero point: the point of zero expected improvement.

[6] The quantity $\mathcal{I}(A)$ depends upon your act probabilities in two places. As we've already seen, if there are correlations between your act and the state of nature, then, as your act probabilities change, so too will the probabilities $\Pr(K)$ in $\mathcal{I}(A)$. Additionally, the term $\mathcal{V}(K)$ depends upon your act probabilities, since $\mathcal{V}(K) = \sum_A \Pr(A \mid K)\mathcal{V}(AK)$. As you become more confident that you will perform A, $V(K)$ will approach $\mathcal{V}(KA)$, and therefore, A's expected improvement will approach zero. (This may appear odd at first; but notice that, the more confident you become that you will perform A, the more you integrate the goodness of your performing A into your evaluation of K, and the less performing A stands to *improve* your evaluation of K.)

[7] To see that this is so, first note that $\mathcal{I}(A) = \mathcal{U}(A) - \mathcal{U}(\top)$, where $\mathcal{U}(\top) = \sum_K \Pr(K)\mathcal{V}(K)$. Thus: $\mathcal{U}(X) \geqslant \mathcal{U}(Y)$ iff $\mathcal{U}(X) - \mathcal{U}(\top) \geqslant \mathcal{U}(Y) - \mathcal{U}(\top)$ iff $\mathcal{I}(X) \geqslant \mathcal{I}(Y)$.

[8] In much of what follows, I will spare the reader the tedium of deriving everything explicitly in the main text. For those who wish to check the math, some advice: multiply the matrix $\mathcal{V}(RC)$ by the matrix $\Pr(R \mid C)$. This gives the matrix $\mathcal{U}(R \mid C)$, of the utility of the row act R, from the perspective you'll occupy immediately after performing the column act C. From here, you may use the identity $\mathcal{I}(R \mid C) = \mathcal{U}(R \mid C) - \mathcal{U}(C \mid C)$ to calculate the matrix $\mathcal{I}(R \mid C)$. The identity in the body can then be used to easily calculate the unconditional improvements, $\mathcal{I}(R)$.

For ease of exposition, going forward, I will refer to '$\mathcal{I}(A \mid B)$' as the degree to which *A improves upon B*. If $\mathcal{I}(A \mid B)$ is positive, then I will say, categorically, that A improves upon B. (Note that, for all acts A, $\mathcal{I}(A \mid A) = 0$, so no act improves upon itself.) In Newcomb, M improves upon L. You know that, no matter what was predicted, M will get you \$10 more than L does. For this reason, $\mathcal{I}(M \mid L) = 10$. And L does not improve upon M. No matter what was predicted, L will get you \$10 less than M does. For this reason, $\mathcal{I}(L \mid M) = -10$. So M improves upon L, and L does not improve upon M.

Terminology: if you think that rational choice in a decision problem is determined by the values $\mathcal{I}(A \mid B)$, for each pair of acts A and B (perhaps together with your act probabilities), then I will call you a *causalist*.[9]

2 Improvement Instability

In Newcomb, taking box M certainly does more to improve the world than taking box L does. This fact does not vary with your act probabilities. Sometimes, which act does the most to improve things *does* vary with your act probabilities. In some cases, this can lead to *instability* in CDT's recommendations. Consider the following decision problem:[10]

> Death in Damascus
> You must choose to travel to either Aleppo or Damascus. You know that, tomorrow, Death will look for you in one of these cities. If you and Death are in the same city, you will die; if you and Death are in different cities, you will live. Your choice does not causally affect where Death awaits, but Death has made a reliable prediction about where you will go. The probability that Death is in Damascus, given that you are in Damascus, is 90%. Whereas the probability that Death is in Aleppo, given that you're in Aleppo, is only 60%. Aleppo additionally has nicer bars than Damascus, so that, if you must meet Death, you'd rather spend your final night on earth in Aleppo.

[9] The evidential decision theorist will not count as a causalist, on this definition, since the value of an act is underdetermined by the values $\mathcal{I}(A \mid B)$. To see this, consider a variant of Newcomb in which, if it was predicted that you would take L, then there is \$0 in L and \$10 in M; whereas, if it was predicted that you would take M, then there is \$100 in L and \$110 in M. In this case, the value of M exceeds the value of L, but it will still be the case that $\mathcal{I}(M \mid L) = 10$, and $\mathcal{I}(L \mid M) = -10$.

[10] *Cf.* Gibbard & Harper (1978). The original Gibbard & Harper example had symmetric probabilities and values for Aleppo and Damascus. Egan (2007) uses asymmetric Death in Damascus cases to argue against CDT. In Egan's cases, one of the two acts—the *safe* act—will always yield the status quo, whereas the other act—the *risky* act—will be either very good or very bad, depending upon a factor outside of your control, but which is highly correlated with your act. If you select the safe act, you should expect the risky one to have very good consequences. But, if you select the risky act, then you should expect it to have very bad consequences. This case also leads to instability. Egan believes that you should choose the safe act.

$$\begin{array}{c|cc} \mathcal{V}(RC) & K_A & K_D \\ A & 10 & 1010 \\ D & 1000 & 0 \end{array} \qquad \begin{array}{c|cc} \Pr(R|C) & A & D \\ K_A & 0.6 & 0.1 \\ K_D & 0.4 & 0.9 \end{array}$$

Table 2: Values and Probabilities for Death in Damascus.

Whether you live or die, and the quality of the bars you frequent, are the only factors relevant to your decision. You prefer living to dying, and you prefer the bars in Aleppo. If we let 'A' be the act of going to Aleppo, 'D' be the act of going to Damascus, and use 'K_A' and 'K_D' for Death's being in Aleppo and Damascus, respectively, then your values and probabilities are as shown in table 2. As in Newcomb, the degree to which you expect the acts A and D to improve things will depend upon your act probabilities. Let a be your probability that you will go to Aleppo, so that $1-a$ is your probability that you will go to Damascus. Then,

$$\mathcal{I}(A) = 810(1-a) \qquad\qquad \mathcal{I}(D) = 190a$$

If you are more than 81% likely to go to Aleppo, $a > 0.81$, then you will expect that going to Damascus would do more to improve things than going to Aleppo, $\mathcal{I}(A) < \mathcal{I}(D)$. If you are less than 81% likely to go to Aleppo, $a < 0.81$, then you will expect that going to Aleppo would do more to improve things than going to Damascus, $\mathcal{I}(A) > I(D)$. When you are exactly 81% likely to go to Aleppo, $a = 0.81$, you will expect that both A and D would improve things to the same degree, $\mathcal{I}(A) = \mathcal{I}(D)$.

This is illustrated in figure 1. There, the line between A and D is of length 1, and each point on this line corresponds to a possible distribution of your act probability—the probability you choose act X is given by the point's distance from the side opposite X (see figure 1a). The arrows in figure 1b show the direction of increasing expected improvement. If your probability for A is less than 81%, then A has a higher expected improvement than D; if your probability for A is greater than 81%, then D has a higher expected improvement than A. The designated point in figure 1b is the *equillibrium* act probability, at which point both A and D are expected to improve things to the same degree.

Suppose that, when you begin deliberation, you think you're as likely to go to Aleppo as you are to go to Damascus, $a = 50\%$. Then, CDT will tell you that you must go to Aleppo. Suppose that you listen to CDT, and that you choose to go to Aleppo. Once you have decided this, you will learn that you have. When you learn this, you should revise your subjective probabilities by becoming more confident that A is true. But then, your subjective probability for A will rise above 81%. And then, CDT will tell you that you must go to Damascus. Suppose that you listen to CDT's new advice. Then, you will learn that you have. When

(a) (b)

FIGURE 1: Deliberational Dynamics for DEATH IN DAMASCUS. In figure 1a, a point on the line between A and D represents an assignment of probabilities to your choosing to go to Aleppo or Damascus. Points closer to A correspond to higher probabilities for A. Points closer to D correspond to higher probabilities for D. In figure 1b, arrows show the direction of increased expected improvement. If $a > 0.81$, then D is expected to improve more than A. If $a < 0.81$, then A is expected to improve more than D. $a = 0.81$ is the *equillibrium* act probability, at which point $\mathcal{I}(A) = \mathcal{I}(D)$.

you learn this, you should revise your subjective probabilities by becoming more confident that D is true. But then, your subjective probability for A will drop below 81%. And then, CDT will tell you that you must go to Aleppo.

Each available act in DEATH IN DAMASCUS improves upon the other. Because $\mathcal{I}(A \mid D) = 810$, A improves upon D. And because $\mathcal{I}(D \mid A) = 190$, D improves upon A. In cases like these, the advice of CDT is *unstable* in a way that leads to diachronic rational dilemmas. If CDT is true, then, in DEATH IN DAMASCUS, there is no option which will remain rational for you to perform after you've resolved to perform it. For every available act, giving yourself evidence that you will perform that act is giving yourself evidence that it is irrational, and the alternative is required.[11] This claim should not be confused with the following: giving yourself evidence that you will go to Damascus is giving yourself evidence that going to Aleppo would be objectively best; and giving yourself evidence that you will go to Aleppo is giving yourself evidence that going to Damascus would be objectively best. Due to Death's ability to predict your choices, this is simply a feature of the case.[12] What is distinctive about CDT's judgment in DEATH IN DAMASCUS is that it tells you, as you are going to Aleppo, that going to Aleppo is impermissible, and that going to Damascus is required. And, as you are going to Damascus, it tells you that going to Damascus is impermissible, and going to Aleppo required.

What strikes me as most disturbing about CDT's instability here is that this instability is entirely *predictable*. Before you decided to go to Aleppo, you knew that, by doing so, you would give yourself evidence that Death awaits in Aleppo. If you know for certain that choosing A will give you some information, then, upon choosing A, you shouldn't be *surprised* by acquiring this new information.

[11] Terminological note: I say that X is 'required' if and only if X is the only permissible option. Thus, as I use it here, being required definitionally entails being permissible. If there are rational dilemmas, in which no option is permissible, then I say that no option is required.

[12] Though see CANTWELL (2010), who suggests that, in cases like DEATH IN DAMASCUS, you should update, not by *conditioning*, but rather by *imaging* on your action—see, *e.g.*, LEWIS (1976) and JOYCE (1999) for more on imaging. If you update in this way, then you will not give yourself evidence that the foregone city is objectively best.

This information should be *expected*. If the information was expected, then it should have already been taken into consideration in your deliberation, and so it shouldn't give you any reason to reconsider your decision. CDT treats the information that you'll likely go to Aleppo as though it gives you a reason to reconsider your decision to go to Aleppo, even though CDT *knew* that you would end up receiving this information when it advised you to go to Aleppo in the first place. This is disturbing because it makes it appear that CDT is not taking into consideration all of the relevant information when it is giving you advice.

There are deliberational versions of CDT which give you the following advice: as you deliberate about what to do, you should raise your probability for acts which you expect to improve things, and lower your probability for acts which you expect to make matters worse. In this way, you will be driven to the *equilibrium* act probability $a = 81\%$. From this deliberational vantage point, both A and D will be expected to improve the world to the same degree. At this point, deliberational variants of CDT either fall silent (telling you only which act *probabilities* are rational, and not which *acts* are rational), or they advise you to select the *mixed act* of going to Aleppo with an 81% probability and going to Damascus with a 19% probability, or else they tell you to be indifferent between Aleppo and Damascus.[13] (Insofar as they fall silent or counsel indifference, deliberational variants of CDT are consistent with everything I will have to say here; though the mixed act interpretation of deliberational CDT will be targeted by my arguments in §3.)

A very natural thought about how to act in cases like these is that you should select whichever act has the highest probability in equilibrium. That is, when you face a choice between two options, each of which improves upon the other, you should select the option which *most* improves upon the other.[14]

Choose the Most Improvement: Given a choice between two options, X and Y, if X improves upon Y more than Y improves upon X, $\mathcal{I}(X \mid Y) > \mathcal{I}(Y \mid X)$, then X is required.

In Death in Damascus, Choose the Most Improvement says that going to Aleppo is required. Note that Choose the Most Improvement only applies in cases where you face a choice between *two* options. It says nothing about how to choose when there are three or more options available. The advice of Choose the Most Improvement in cases like Death in Damascus is endorsed by Egan (2007), Briggs (2010), Gustafsson (2011), and Spencer & Wells (2017), and

[13] See Skyrms (1990), Arntzenius (2008), and Joyce (2012).

[14] To see that these two characterizations align, note that an *equilibrium* act probability, x, will have $\mathcal{I}(X) = \mathcal{I}(Y)$, or $\mathcal{I}(X \mid Y)(1-x) = \mathcal{I}(Y \mid X)x$. Thus, $x > 0.5$ iff $\mathcal{I}(X \mid Y) > \mathcal{I}(Y \mid X)$. (Recall the identity $\mathcal{I}(X) = \sum_A \mathcal{I}(X \mid A) \cdot \Pr(A)$, and consult the definition of *equilibrium* on page 12.)

the principle is explicitly endorsed in full generality by Wedgwood (2013) and Barnett (ms).

We may argue directly for Choose the Most Improvement from these two plausible principles:

Symmetry Indifference: Given a choice between two options, X and Y, if X improves upon Y to the same degree that Y improves upon X, $\mathcal{I}(X \mid Y) = \mathcal{I}(Y \mid X)$ (so that the unique equilibrium act probability is a 50% probability of selecting X and a 50% probability of selecting Y), then neither X nor Y is required.

Sweetening: If, given a choice between X and Y, neither X nor Y is required, then, if X^+ is more valuable than X in every state of nature, then, given a choice between X^+ and Y, X^+ is required.

Suppose that X improves upon Y more than Y improves upon X. Then, there is some possible option X^- such that X is more valuable than X^- in each state of nature, and, given a choice between the two, X^- improves upon Y to the same extent that Y improves upon X^-. By Symmetry Indifference, given a choice between X^- and Y, neither is required. By Sweetening, then, given a choice between X and Y, X is required. Thus, Choose the Most Improvement is true. It's hard for causalists to deny Symmetry Indifference, and there are powerful arguments for Sweetening.[15] So Choose the Most Improvement is a compelling and plausible principle.

For another application of Choose the Most Improvement, consider Extreme Equilibrium.

> Extreme Equilibrium
> Before you are two boxes: box A and box B. Box A contains a guaranteed \$10. If it was predicted that you would select box B, then box B was left empty. If it was predicted that you would select box A, then \$100 was placed in box B. These predictions are 90% reliable—conditional on you taking box X, the chance that it was predicted that you would take X is 90%. (Though nothing you do now will affect how much money is in the boxes.)

I assume that your values are linear in dollars so that your values and probabilities for Extreme Equilibrium are shown in table 3. The degree to which A and B are expected to improve things will depend upon your act probabilities. If a is your probability that you will take A, then:

$$\mathcal{I}(A) = 0 \qquad\qquad \mathcal{I}(B) = 80a$$

[15] See, *e.g.*, Hare (2010).

$\mathcal{V}(RC)$	K_A	K_B	$\Pr(C\|R)$	A	B
A	10	10	K_A	0.9	0.1
B	100	0	K_B	0.1	0.9

Table 3: Values and Probabilities for Extreme Equilibrium.

So long as you are not certain to take B, taking B is expected to improve things more than taking A is. However, once you have resolved to take B, you will expect both A and B to improve things to the same degree. In this respect, Extreme Equilibrium differs from Newcomb. In Newcomb, no matter your act probabilities, M has a greater expected improvement than L does. In contrast, in Extreme Equilibrium, while B has a greater expected improvement than A for *almost* all act probabilities, for the extreme act probability $a = 0$, both A and B have equal expected improvement. In Newcomb, as you pursue the option which you expect to do the most to improve things, it *remains* the option which you expect to do the most to improve things. In Extreme Equilibrium, in contrast, as you pursue the option which you expect to do the most to improve things, the degree to which you expect it to do more good than the alternative shrinks to zero.

A quick word on terminology: if you've followed the advice of the deliberational causal decision theorist, raising your probability for acts which you expect to improve things, and lowering your probability for acts which you expect to make matters worse, and you've reached stable act probabilities, where no further revisions are called for,[16] then I'll say that your act probabilities are *settled*. If, with these settled act probabilities, two or more acts have equal expected improvement, then I'll say that your act probability distribution is an *equilibrium*.[17] Thus, in Newcomb, when you are certain to take M, your act probabilities are settled, but they are not in equilibrium, since the expected improvement of M will still exceed that of L. In contrast, in Extreme Equilibrium, when you are certain to take B, your act probabilities are both settled and in equilibrium. And I'll say that a decision problem leads to *instability* iff your settled act probabilities are an equilibrium. Thus, both Death in Damascus and Extreme Equilibrium lead to instability. (Once you're certain to take B, you'll expect A to improve things to the same degree; however, were you to take A you'd expect taking B to

[16] See Skyrms (1990) for the details.

[17] Thus, what I call a 'settled' act probability distribution is what Skyrms (1990) calls an 'equilibrium'. I introduce the distinction because I wish to distinguish act probability distributions like the one in Newcomb (which is settled, but not in equilibrium), from the one in Extreme Equilibrium (which is settled and in equilibrium). These act probability distributions are importantly different, though both of them assign a 100% probability to one act. My conclusion will be that, in Extreme Equilibrium, neither option is required; though, in Newcomb, taking box M is required.

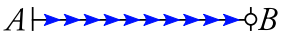


FIGURE 2: Deliberational Dynamics for EXTREME EQUILLIBRIUM So long as $a > 0$, $\mathcal{I}(B) > \mathcal{I}(A)$. When $a = 0$, $\mathcal{I}(B) = \mathcal{I}(A)$.

improve the world more.)

The evidential decision theorist says that it is permissible to take either A or B in EXTREME EQUILIBRIUM.[18] However, a causalist who accepts CHOOSE THE MOST IMPROVEMENT will disagree. They will say that, because B improves upon A and A does not improve upon B—*i.e.*, $\mathcal{I}(B \mid A) = 80$ and $\mathcal{I}(A \mid B) = 0$—it is permissible to choose B and impermissible to choose A. In fact, in the case of EXTREME EQUILIBRIUM, the same conclusion follows from a strictly weaker principle:

CHOOSE EXTREME EQUILIBRIA: Given a choice between two options, X and Y, if the unique equilibrium act probability is a 100% probability of choosing X, then X is required.

In EXTREME EQUILIBRIUM, the only equilibrium act probability is a 100% probability for selecting B. So CHOOSE EXTREME EQUILIBRIA says that B is required. Note that CHOOSE EXTREME EQUILIBRIA only applies in cases where you face a choice between *two* options. It says nothing about how to choose when there are three or more options. CHOOSE EXTREME EQUILIBRIA follows from the mixed act understanding of deliberational CDT, assuming that we equate the act X with the 'mixed' act of performing X with 100% probability.

CHOOSE EXTREME EQUILIBRIA is incredibly intuitive and plausible. I have, however, come to believe that it is false. (Since it is false, so too is CHOOSE THE MOST IMPROVEMENT.) What I've come to believe instead is that we should respond to cases like DEATH IN DAMASCUS and EXTREME EQUILIBRIA with *indifference*.

INSTABILITY CALLS FOR INDIFFERENCE: Given a choice between two options, X and Y, if both $\mathcal{I}(X \mid Y) \geqslant 0$ and $\mathcal{I}(Y \mid X) \geqslant 0$, then neither X nor Y is required.

Thus, in DEATH IN DAMASCUS, neither Aleppo nor Damascus is required; and, in EXTREME EQUILIBRIUM, neither box A nor box B is required.[19]

---

[18] The reader may be inclined to agree because they see A as a *safe* option—one which is guaranteed to pay out \$10 no matter what. Such attitudes towards risk may be rational—see BUCHAK (2013)—but let us suppose for the nonce that you are not at all risk-averse. Debates about risk-aversion cross-cut our debate here.

[19] Recall: as I use the term here, 'required' means 'uniquely permissible'. Thus, it is consistent with INSTABILITY CALLS FOR INDIFFERENCE that you should take the options X and Y to be *incomparable*, rather than being indifferent between them; it is also consistent with the principle that neither X nor Y is permissible.

	K_A	K_B	K_C
A	0	0	100
B	100	0	0
C	0	100	0

	A	B	C
K_A	0.8	0.1	0.1
K_B	0.1	0.8	0.1
K_C	0.1	0.1	0.8

Table 4: Values and Probabilities for Improvement Condorcet

My reason for reaching this conclusion is that Choose Extreme Equilibria leads to absurd consequences in a particular decision problem—a decision problem which, due to its resemblance to Marquis de Condorcet's voting paradox, I will call an '*improvement Condorcet paradox*' (though, to be clear: I take this to be a *veridical* paradox). I will show that, in this decision problem, Choose Extreme Equilibria leads to violations of weak versions of the principles *the independence of irrelevant alternatives* and *normal-form extensive-form equivalence*. Moreover, in these kinds of cases, any agent who acts in accord with Choose the Most Improvement will end up predictably poorer than someone who abides by Instability Calls for Indifference, even when they face the problem in the same circumstances.

3 An Improvement Condorcet Paradox

Consider the following decision problem:[20]

> Improvement Condorcet
> Before you are three boxes, labeled 'A', 'B', and 'C'. You may take one and only one of the boxes. A $100 bill was placed in one of the boxes on the basis of a prediction about how you would choose. If it was predicted that you would choose A, the money was put in box B. If it was predicted that you would choose B, the money was put in box C. If it was predicted you would choose C, the money was put in box A. These predictions are 80% reliable.

Your values and probabilities for this problem are as shown in table 4. Which box you expect to do the most to improve things depends upon your act probabilities. Let 'a', 'b', and 'c' be your probabilities that you will take box A, B, and C, respectively. Then:

$$\mathcal{I}(A) = 70c \qquad \mathcal{I}(B) = 70a \qquad \mathcal{I}(C) = 70b$$

If you are most likely to take A, then B will be expected to do the most to improve things. If you are most likely to take B, then C will be expected to do the most to

[20] *Cf.* Ahmed (2012).

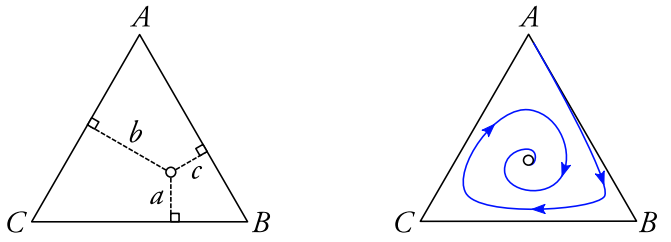


Figure 3: Deliberational Dynamics for Improvement Condorcet. Figure 3a: an equilateral triangle of height 1. A point in the triangle corresponds to an act probability distribution over A, B, and C. The probability assigned to X is given by the length of a perpendicular line from the point to the side opposite the vertex X. Figure 3b: if you begin with a high probability for A and adjust your act probabilities to move in the direction of maximal expected improvement, then you will follow a spiral path like the one shown, eventually arriving at the equilibrium act probability of $a = b = c = 1/3$.

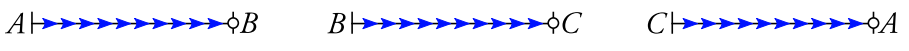
improve things. And if you are most likely to take C, then A will be expected to do the most to improve things. Each available act improves upon one alternative and is improved upon by the other, in equal measure.

In this decision problem, there is a unique equilibrium act probability of $a = b = c = 1/3$. No matter your initial act probabilities, if you increase the probabilities of acts which you expect to improve things more, and decrease the probabilities of acts which you expect to improve things less, then you will be driven to this equilibrium. (See figure 3.)

Improvement Condorcet is so-named because of its resemblance to the Marquis de Condorcet's voting paradox. In Condorcet's voting paradox, there are three candidates—call them 'I', 'J', and 'K'. One third of the voters have the preference ordering $K \succ J \succ I$; one third have $J \succ I \succ K$; and one third have $I \succ K \succ J$. Then, in a one-on-one contest between I and J, J would receive a two-thirds majority; in a one-on-one contest between J and K, K would receive a two-thirds majority; and, in a one-on-one contest between K and I, I would receive a two-thirds majority. Something analogous happens in Improvement Condorcet, if we assume Choose Extreme Equilibria. Suppose you are given a choice between A and B—C is taken off of the menu (note, however, that even though you are guaranteed to not take C, there is still a 10% probability that it was predicted you'd take C). In that case, your act probability for C, c, is constrained to be zero, and the expected improvements for A and B are:

$$\mathcal{I}(A) = 0 \qquad\qquad \mathcal{I}(B) = 70a$$

With C removed from the menu, your choice is exactly like the choice from Extreme Equilibrium. So long as you are not certain to take B, taking B is



(a) (b) (c)

FIGURE 4: Deliberational Dynamics for a choice between A and B (figure 4a), a choice between B and C (figure 4b), and a choice between C and A (figure 4c) in IMPROVEMENT CONDORCET.

expected to improve things more than taking A is. Once you have resolved to take B, you will expect both A and B to improve things to the same degree. B improves upon A, $\mathcal{I}(B \mid A) = 70$, and A does not improve upon B, $\mathcal{I}(A \mid B) = 0$. The unique equilibrium act probability is a 100% probability of choosing B. (See figure 4a.) So CHOOSE EXTREME EQUILIBRIA says that, given a choice between just A and B, B is required. Suppose, on the other hand, that A is removed from the menu, and you are given a choice between B and C. In that case, your act probability for A, a, is constrained to be zero, and the expected improvements of B and C are:

$$\mathcal{I}(B) = 0 \qquad\qquad \mathcal{I}(C) = 70b$$

Once again, the unique equilibrium act probability is a 100% probability of choosing C (see figure 4b), and CHOOSE EXTREME EQUILIBRIA says that, given a choice between just B and C, C is required. Similarly, when B is removed from the menu, and you are given a choice between C and A, the expected improvements of C and A will be:

$$\mathcal{I}(C) = 0 \qquad\qquad \mathcal{I}(A) = 70c$$

And the unique equilibrium act probability will be a 100% probability of choosing A (see figure 4c). So CHOOSE EXTREME EQUILIBRIA says that A is required.

3.1. THE INDEPENDENCE OF IRRELEVANT ALTERNATIVES. If we assume that IMPROVEMENT CONDORCET is not a rational dilemma (*i.e.*, if we assume that *some* option is permissible), then CHOOSE EXTREME EQUILIBRIA leads to a violation of a principle known as *the independence of irrelevant alternatives* (or just 'IIA').[21]

INDEPENDENCE OF IRRELEVANT ALTERNATIVES (IIA): If, given a choice between X and Y, Y is not a permissible option, then, given a choice between

[21] The principle in the body is weaker than a usual formulation of IIA, which says that, if it is not permissible to select X off of one menu (which includes X), then X is also not permissible to select off a *larger* menu. There are additionally other, logically independent, principles which go by the name 'independence of irrelevant alternatives'. (For instance, the principle which goes by this name in ARROW (1950)'s famous impossibility result is logically independent of the principle in the body. See RAY (1973).)

X, Y, and Z, Y is not a permissible option.

By Choose Extreme Equilibria, *every* option in Improvement Condorcet is impermissible in a one-on-one choice with some alternative. So, if *some* option is permissible,[22] we will have a violation of IIA. For illustration: suppose that A is a permissible choice in Improvement Condorcet. By Choose Extreme Equilibria, given a choice between A and B, A is impermissible. So A is not a permissible choice when you are presented with the restricted menu $\{A, B\}$, but it *is* a permissible choice when you are presented with the larger menu $\{A, B, C\}$. And this contradicts IIA. The same goes if we say that B or C is permissible instead. For Choose Extreme Equilibria says that B is impermissible on the restricted menu $\{B, C\}$, and C is impermissible on the restricted menu $\{C, A\}$.

When thinking about principles like IIA, it is important to be clear about how options are individuated. Some apparent counterexamples to the principle are not genuine counterexamples, because they conflate distinct options. John invites you to dinner. You are presented with two options: having dinner with John, and having dinner by yourself. You regard having dinner with John as clearly better than having dinner by yourself. Then, John says: 'Or we could go to the Trump rally and grab food afterwards.' You're now faced with three options: having dinner with John, going to the Trump rally with John, or having dinner by yourself. You now regard having dinner by yourself as the best possible option.[23] Have you violated IIA? No. When you are offered the Trump rally as an alternative, you learn something about what dinner with John would be like. This changes your evaluation of the dinner, and makes it a *different* option than the one with which you were presented before. Let me be clear, then, about how I individuate options. For the purposes of understanding the principle IIA, X and X^* are *the same option* iff: a) you give them the same value, in every possible state of nature—that is: $\mathcal{V}(XK) = \mathcal{V}(X^*K)$, for every state of nature K; and b) your subjective probability distribution over states, conditional on X, is the same as your subjective probability distribution over states, conditional on X^*—that is: $\Pr(K \mid X) = \Pr(K \mid X^*)$, for every state of nature K. So understood, Choose Extreme Equilibria leads to violations of IIA.[24]

3.2. Normal-Form Extensive-Form Equivalence. Abiding Choose Extreme Equilibria will lead you to violate a weak principle of *normal-form extensive-form equivalence* (or just 'NEE').

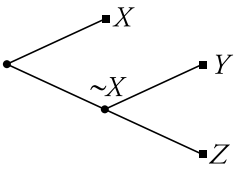
---

[22] By the symmetry of the case, we should conclude that *every* option is permissible, but we need not assume this in order to make the present point.

[23] Sen (1993).

[24] *Cf.* Wedgwood (2013), Bassett (2015), and Barnett (ms), for discussion of how Wedgwood's *benchmark theory* and Barnett's *graded ratifiability* theory lead to violations of IIA.

Normal-Form Extensive-Form Equivalence (NEE): If you are certain to remain rational and your beliefs and values are certain to not change, then, if it is permissible to not choose X when given a choice between X, Y, and Z, then, given a choice between X and going on to choose between Y and Z, it is permissible to not choose X.

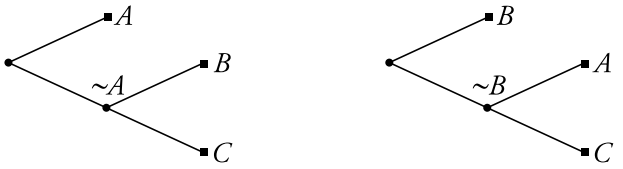


The antecedent of NEE is important. Suppose you think that your beliefs or values might change after choosing $\sim X$ and before choosing between Y and Z. Then, it may be rational to choose X now in order to take the decision out of the hands of your not entirely trustworthy future self. Likewise, if you fear that your future self will not choose rationally, this could give additional reason to select X at stage one. However, restricted to cases where you are certain to retain your beliefs, values, and rationality, NEE is very plausible.

Consider now the following two choices:

> A or ~A
> Money was distributed between boxes A, B, and C as in Improvement Condorcet. At stage 1, you are given a choice to either take box A or to not. If you take box A, then you receive its contents. If you don't take A, then at stage 2, you choose between B and C. (See figure 5a.) You are certain to retain your beliefs, values, and rationality throughout.
>
> B or ~B
> Money was distributed between boxes A, B, and C as in Improvement Condorcet. At stage 1, you are given a choice to either take box B or to not. If you take box B, then you receive its contents. If you don't take B, then at stage 2, you choose between A and C. (See figure 5b.) You are certain to retain your beliefs, values, and rationality throughout.

Assume Choose Extreme Equilibria. Then, in A or ~A, if you choose $\sim A$ at stage 1, at stage 2, you will choose C. So, at stage 1, you face a choice between A and C. By Choose Extreme Equilibria, A is required at stage 1. In B or ~B, if you choose $\sim B$ at stage 1, then, at stage 2, you will choose A. So, at stage 1, you face a choice between B and A. By Choose Extreme Equilibria, B is required at stage 1.



(a) A or ~A (b) B or ~B

Figure 5

We can now show that, assuming *some* option is permissible, Choose Extreme Equilibria violates NEE in Improvement Condorcet. For B is either permissible or it is not. Suppose it is. Then, NEE says that $\sim A$ is permissible in A or ~A. Choose Extreme Equilibria, on the other hand, says that $\sim A$ is impermissible, contradicting NEE. Suppose on the other hand that B is impermissible. Then, it is permissible to not choose B. In that case, NEE says that $\sim B$ is permissible in B or ~B. Choose Extreme Equilibria, on the other hand, says that $\sim B$ is impermissible, contradicting NEE. Either way, Choose Extreme Equilibria contradicts NEE.[25]

3.3. Predictable Long-run Poverty. If you additionally accept Choose the Most Improvement, then your preferences in Improvement Condorcet may be exploited to lose you money in the long run. Suppose that, instead of taking a box yourself, you select a box with the aid of an assistant. You tell the assistant which box to take, but it is the assistant who makes the final choice. (You keep the money. Note also that the reliable predictions are now about which box your assistant will end up selecting.) By the symmetry of the case, you see no reason to favor any box over the others, and you tell your assistant to take box A. Before your assistant departs, they get an idea. They say: 'Are you sure? I'll give you an opportunity to change to box B (but not box C—I'm taking that off the menu). In exchange for changing your mind, I'll require thirty dollars.' (You are certain that they will take this decision to be final, they will take the box you decide upon, and that there's no longer any way to get them to take C.) At this point, you face a new decision: not between A, B, and C, but instead between staying with A and changing to B and losing thirty dollars. The expected improvements of the available acts are:

$$\mathcal{I}(A) = 30(1-a) \qquad\qquad \mathcal{I}(B) = 40a$$

[25] It's worth noting that some deliberational versions of CDT which recommend indifference in Death in Damascus will, in other cases, violate NEE. For instance, Joyce's favored formulation of deliberational CDT violates NEE in Ahmed (2014)'s Dicing with Death decision problem. I'm inclined to regard this as a mark against that decision theory. See Joyce (2018) for discussion.

In this new decision, the equilibrium act probability is $a = 3/7$, or about a 57% probability of taking box B. A now improves upon B by thirty dollars, $\mathcal{I}(A \mid B) = 30$, but B improves upon A by 40 dollars, $\mathcal{I}(B \mid A) = 40$. By Choose the Most Improvement, B is required. So the principle advises you to hand your assistant thirty dollars to have them take B instead. But you could have had B in the first place, for free. How could your assistant's offer give you reason to switch?

Nothing changes if we suppose that you know in advance that your assistant will make you an offer of this kind. No matter which box you initially select, the assistant will be able to offer you a trade for another box, at a cost of $30, which you will see as favorable, assuming Choose the Most Improvement. There's no initial selection which will prevent your future self from making the trade.

Note that, if you make the trade, then you will likely end up losing money overall. You have only a 10% chance of winning $100—an expected return of only $10. But you've just handed over three times that amount. In the long run in which you make this decision over and over again, with your assistant offering the trade each time, you will lose $20 on average. You could have instead ended up with $10 on average, if only you'd refused the assistant's trade.

Causalists are used to making less money in certain decision problems. For instance, anyone who takes box M in Newcomb will predictably make less money, over the long run, than someone who takes box L. The usual causalist reply is convincing: this is true, but only because those who take L will typically be *provided* with more money than those who take box M. Being afforded greater opportunities for wealth is no sign of rationality; nor is being afforded fewer opportuntities for wealth a sign of irrationality. So predictable poverty in Newcomb is no sign of irrationality.[26] A comparable defense is not available here. In this case, it was not an unfortunate environment which led to your poverty; it was your poor decisions. Someone who was indifferent between A and B when given a choice between the two would predictably make more money facing exactly the same problem, in exactly the same circumstances.

Any decision theory which validates Choose the Most Improvement will advise you to pay to have the options presented to you in a certain order—even when you're certain to retain your beliefs, values, and rationality throughout. For instance, consider Pay or A:

> Pay or A
> Money is distriuted between boxes A, B, and C as in Improvement Condorcet. At stage 1, you may either pay $30, P, or not, $\sim P$. If you pay, then, at stage 2, you will face the decision B or ~B. If you do not, then, at stage 2, you will face A or ~A. (See figure 6.)

If you abide Choose Extreme Equilibria, you will choose A in A or ~A. So,

[26] See, *e.g.*, Gibbard & Harper (1978), Lewis (1981), Joyce (1999), and Wells (forthcoming).

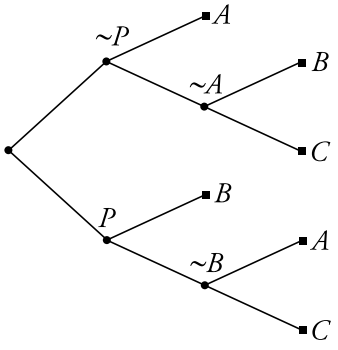


Figure 6: Pay or A

if you don't pay, you will end up choosing *A*. If you abide Choose Extreme Equilibria, you will choose *B* in B or ~B. So, if you pay, you will end up choosing *B*. So, at stage 1, you face a choice between paying $30 and taking box *B* and not paying and taking box *A*. This is the same choice you faced with your assistant. And, again, Choose the Most Improvement tells you to pay the $30.

Again, paying likely leads to you losing money overall. Whether you play A or ~A or B or ~B, the expected return is $10. And you've just handed over three times that amount. In the long run in which you face Pay or A over and over again, you will lose $20 on average. Again, someone who was indifferent between *A*, *B*, and *C* when given a choice between any two would predictably make more money when facing exactly the same choice in exactly the same circumstances. So this predictable poverty is not a consequence of poor opportunities; it is a consequence of poor choices.[27]

4 Conclusion

With the Improvement Condorcet case, we have seen two impossibility results. Firstly: assuming that *some* choice is permissible in the case, there is no decision theory which satisfies both Choose Extreme Equilibria and the independence of irrelevant alterntives (IIA). Secondly: assuming that some choice is permissible in Improvement Condorcet, there is no decision theory which satisfies both Choose Extreme Equilibria and a weak principle of normal-form extensive-form equivalence (NEE). Further, if you additionally abide Choose the Most Improvement, these violations of IIA and NEE will lead you to make decisions which will predictably lead you into poverty—not because these decisions are correlated with inevitable poverty, but rather because these decisions *cause* poverty.

[27] We may not be bothered by these consequences if we accept *The Binding Principle*—see Arntzenius et al. (2004) and Meacham (2010). According to this principle, if an objection to a decision theory only arises for agents who are incapable of binding themselves (forcing their future selves to make a certain decision), then it is no objection at all.

An agent who accepts Instability Calls for Indifference will, in precisely the same decision problems, in precisely the same circumstances, walk away with riches in the long run.

So, I counsel rejecting Choose Extreme Equilibria and Choose the Most Improvement. In their place, I recommend Instability Calls for Indifference. Thus, I recommend saying that neither Aleppo nor Damascus is uniquely permissible in Death in Damascus, and neither box *A* nor box *B* is uniquely permissible in Extreme Equilibrium. As we saw in §2 above, Choose the Most Improvement is entailed by Symmetry Indifference and Sweetening. Since Instability Calls for Indifference entails Symmetry Indifference, I counsel rejecting Sweetening, too. There are strong arguments in favor of Sweetening—see, for instance, Hare (2010). Still, the costs of denying Sweetening strike me as less steep than the costs of accepting it; so I am inclined to reject it.

References

Ahmed, Arif. 2012. "Push the Button." *Philosophy of Science*, vol. 79 (3): 386–395. [14]

—. 2014. "Dicing with Death." *Analysis*, vol. 74 (4): 587–592. [19]

Arntzenius, Frank. 2008. "No regrets, or: Edith Piaf revamps decision theory." *Erkenntnis*, vol. 68: 277–297. [3], [10]

Arntzenius, Frank, John Hawthorne & Adam Elga. 2004. "Bayesianism, Infinite Decisions, and Binding." *Mind*, vol. 113 (450): 251–283. [21]

Arrow, Kenneth J. 1950. "A Difficulty in the Concept of Social Welfare." *The Journal of Political Economy*, vol. 58 (4): 328–346. [16]

Barnett, David James. ms. "Graded Ratifiability." [11], [17]

Bassett, Robert. 2015. "A Critique of Benchmark Theory." *Synthese*, vol. 192 (1): 241–267. [17]

Briggs, Rachael. 2010. "The Metaphysics of Chance." *Philosophy Compass*, vol. 5 (11): 938–952. [10]

Buchak, Lara. 2013. *Risk and Rationality*. Oxford University Press, Oxford. [13]

Cantwell, John. 2010. "On an alleged counter-example to causal decision theory." *Synthese*, vol. 173: 127–152. [9]

Egan, Andy. 2007. "Some Counterexamples to Causal Decision Theory." *Philosophical Review*, vol. 116 (1): 93–114. [7], [10]

Gibbard, Allan & William L. Harper. 1978. "Counterfactuals and Two Kinds of Expected Utility." In *Foundations and Applications of Decision Theory*, A. Hooker, J.J. Leach & E.F. McClennan, editors, 125–162. D. Reidel, Dordrecht. [1], [7], [20]

Gustafsson, Johan E. 2011. "A Note in Defence of Ratificationism." *Erkenntnis*, vol. 75: 147–150. [10]

Hare, Caspar. 2010. "Take the Sugar." *Analysis*, vol. 70 (2): 237–247. [11], [22]

Joyce, James M. 1999. *The Foundations of Causal Decision Theory*. Cambridge University Press, Cambridge. [9], [20]

—. 2012. "Regret and instability in causal decision theory." *Synthese*, vol. 187 (1): 123–145. [10]

—. 2018. "Deliberation and Stability in Newcomb Problems and Pseudo-Newcomb Problems." In *Newcomb's Problem*, Arif Ahmed, editor. Oxford University Press, Oxford. [19]

Lewis, David K. 1976. "Probability of Conditionals and Conditional Probablities." *Philosophical Review*, vol. 85: 297–315. [9]

—. 1981. "'Why ain'cha rich?'." *Noûs*, vol. 15 (3): 377–380. [20]

Meacham, Christopher J.G. 2010. "Binding and Its Consequences." *Philosophical Studies*, vol. 149 (1): 49–71. [21]

Ray, Paramesh. 1973. "Independence of Irrelevant Alternatives." *Econometrica*, vol. 41 (5): 987–991. [16]

Sen, Amartya. 1993. "Internal Consistency of Choice." *Econometrica*, vol. 61 (3): 495–521. [17]

Skyrms, Brian. 1990. *The Dynamics of Rational Deliberation*. Harvard University Press, Cambridge, ma. [10], [12]

Spencer, Jack & Ian Wells. 2017. "Why Take Both Boxes?" *Philosophy and Phenomenological Research*. [10]

Wedgwood, Ralph. 2013. "Gandalf's solution to the Newcomb Problem." *Synthese*, vol. 190 (14): 2643–2675. [11], [17]

Wells, Ian. forthcoming. "Equal Opportunity and Newcomb's Problem." *Mind*. [4], [20]