Frustration and Delay

Abstract. The paper describes a situation in which Causal Decision Theory (CDT) advises the agent to decline a free offer of $1,000, with the foreseeable result that the agent is $1,000 poorer than if she had taken the offer. I take this absurd consequence of CDT to constitute a refutation of it.

1. The game show *Frustration!* – popular around the start of the twenty-sixth century – faced contestants with two opaque boxes labelled A and B and the following instructions:

> *Frustration*. You must take just one box – you get to keep whatever you find in it. Before you choose, you should know that five minutes ago we conducted a brain scan to determine which box you would ultimately take on receiving these instructions. The scan detects a short-lived electrical signature – the *A-signature* – whose presence (or absence) predicts that you will take Box A (Box B). These predictions are at least 99% accurate.[1] If the scan detected the A-signature then we put nothing in Box A and $1M in Box B. If the scan detected no A-signature then we put the $1M in Box A and nothing in Box B. Now do you want to take Box A or Box B?[2]

Of course most contestants won nothing and went away frustrated; and as for the ≤1% of contestants who made $1M, the producer – the Betelgeuse Broadcasting Corporation (BBC) – found that it could cover these costs with advertising revenues when the show proved – inexplicably – to be wildly popular.

Soon though, it ran into trouble: not because of viewing figures, which remained buoyant, but because of a shortage of *contestants*. The game was aptly named – losing it *was* frustrating. And almost everyone *did* lose. After one particularly bad run of twenty-four shows without a winner, what had once been a flood of applicants threatened to dry up altogether. Of course the prospect of winning $1M was still tempting enough to some, but most of them chose instead to appear on the better-known, longer-running and more lucrative rival show *Newcomb!*[3] Instead of airing it every night as originally planned, the producers were soon forced to cut *Frustration!* back to one episode per week.

2. The solution seemed obvious: *pay* people a small incentive, say $1,000 ($1K), to come on the show. Unfortunately, monetary inducements to come on philosophy-based game shows had been prohibited by interplanetary treaty following the, scandals surrounding the twenty-third century show *Trolley!* and its ill-fated sequels.

Eventually somebody had a better idea: instead of giving contestants $1K up-front, why not make the money part of the game itself? That is: instead of offering contestants a straight choice between Box A and Box B, give them a *third* option to take $1K, no strings attached, before deciding. This (perfectly legal) device got enthusiastic approval, and that evening the game show, now called *Frustration and Delay*, confronted its contestants with two opaque boxes and the following, amended instructions:

> *Frustration and Delay*. You must take just one of these boxes – you get to keep whatever you find in it. Before you choose, you should know that five minutes ago we conducted a brain scan to determine which box you would ultimately do on receiving these instructions. The scan detects a short-lived electrical signature – the *A-signature* – whose presence (or absence) predicts that you will ultimately take Box A (Box B). These predictions are at least 99% accurate. If the scan detected the A-signature then we put nothing in Box A and $1M in Box B. If the scan detected no A-signature then we put the $1M in Box A and nothing in Box B.

> You have *three* options. You can take Box A now. You can take Box B now. *Or* you can take this $1K, no strings attached, *before* choosing between A and B. In case you think there is some catch, we can assure you (following rigorous tests) that taking the money makes no difference to your signature or to your choice. Subjects with the A-signature who take the $1K are as likely to take Box A as are subjects with the A-signature who turn it down; and the same goes for subjects who lack the A-signature. Nor does taking the $1K make any difference to what is in either box: this has already been fixed and there is nothing anyone can do to change it now. So it is a free $1K: taking it makes no difference to whatever else happens.

As you might expect, the effect was immediate and dramatic. Whereas before most contestants had left empty-handed, now most were leaving with $1K. Numbers revived and before long a new episode of *Frustration and Delay* was being beamed out every night.

But this success quickly raised concerns about cost. With ten contestants per episode, each episode of *Frustration and Delay* was expected to cost an additional $10K, on the natural assumption that everyone would take the $1K up front. After all, why *wouldn't* you take it? If you were going to miss out on the million anyway, then the only effect would be to make you $1K better off than if you hadn't taken it. Similarly, if you were one of the lucky few who was going to choose a box contrary to prediction (and so win $1M), then taking the $1K first wouldn't make any difference to that either, and so again its only effect would be to make you $1K better off.

3. Everyone was therefore surprised and relieved when it turned out that about a third of contestants chose *not* to take the initial offer of $1K but rather chose Box A or Box B directly.

These contestants, who almost always left the show with nothing at all, soon became objects of widespread curiosity as well as derision. Upon enquiry it turned out that they all followed a philosophical theory of rational choice known as *Causal Decision Theory* (CDT). This theory (as they explained) advises you to choose from amongst options by evaluating their causal *effect* on what you care about. For each option, consider all the possible hypotheses about what outcome that option *would* bring about if it *were* realized. Let us suppose that you can associate with each such outcome a numerical score corresponding to how much you want it to be realized. The *causal utility* of an option, which is the measure of its merit, is a weighted average of the scores of each possible outcome, where the weight associated with each possible outcome by each option is your confidence that taking that option would bring about that outcome.[4]

In *Frustration and Delay*, three options are initially available:

A: Take Box A directly

B: Take Box B directly

X: Take the $1K and delay the choice

What would the effect of each option be? As far as you are concerned, the effect of directly taking Box A depends only on where the $1M is. If it is in Box A, then the effect of directly taking Box A is that you make $1M. If it is in Box B, then the effect of directly taking Box A is that you leave with nothing.

Similarly, the effect of directly taking Box B depends only on where the $1M is. If it is in Box A, then the effect of directly taking Box B is that you leave with nothing. If it is in Box B, the effect of directly taking Box B is that you make $1M.

Finally, the effect of first taking the $1K depends on *two* things: (i) which box you ultimately take, and (ii) where the $1M is. (And whether you take the $1K makes no difference to either of *those* things.) If you ultimately take Box A and the $1M is in Box A, then the effect of taking the $1K is that you make $1K plus $1M. If you ultimately take Box A and the $1M is in Box B, then the effect of taking the $1K is that you make just $1K. If you ultimately take Box B and the $1M is in Box A, then the effect of taking the $1K is again that you make just $1K. And if you ultimately take Box B and the $1M is in Box B, then the effect of taking the $1K is that you make $1K plus $1M.

It follows that there are two issues on which the causal effect of all three options supervenes: (i) whether you ultimately take Box A or Box B and (ii) where the $1M is. There are therefore four possible hypotheses each of which determines what each of your options would bring about: we may label these H_1 to H_4.

H_1: You ultimately take Box A; the $1M is in Box A

H_2: You ultimately take Box A; the $1M is in Box B

H_3: You ultimately take Box B; the $1M is in Box A

H_4: You ultimately take Box B; the $1M is in Box B

The dollar values of the effects of each option under each hypothesis are as in the following table.[5]

	H_1: **Take A** **$1M in A**	H_2: **Take A** **$1M in B**	H_3: **Take B** **$1M in A**	H_4: **Take B** **$1M in B**
$\boldsymbol{A}$	M	0	M	0
$\boldsymbol{B}$	0	M	0	M
$\boldsymbol{X}$	$M+K$	K	K	$M+K$

Table 1

It is clear from this table how the causal utility of each option depends on your confidence in each of H_1, H_2, H_3, H_4. For instance, suppose that you are very confident that H_1 is true. Then you are confident (i) that you will in fact ultimately take Box A and (ii) that the $1M is in Box A. By (ii), you are confident that if you were to take Box A directly then you would make $1M. By (ii) again, you are confident that if you *were* to take Box B directly (though you probably will not) then you would make nothing. And in light of (i) and the fact that taking the $1K offer delays but does not change your ultimate choice of box, you are confident that if you were to take the $1K first then you would end up with $1M on top of the $1K bonus.[6]

Clearly though, you are in fact overwhelmingly confident that the true hypothesis is either H_2 or H_3. For you are overwhelmingly confident that the brain-scan is accurate i.e. that it detected the A-signature if you will ultimately take Box A, and that it detected no A-signature if you will ultimately take Box B. And you are certain that if the scan detected the A-signature then the $1M is in Box B, and that if the scan detected no A -signature then the $1M is in Box B. So almost certainly, either you will take Box A and the $1M is in Box B, or you will take Box B and the $1M is in Box A.[7]

But if so, then at least one of A and B has a causal utility exceeding that of X. For instance, if you are certain of H_2 then you are certain that you will end up taking Box A and that the $1M is in Box B. In that case, directly taking Box B would have maximal causal utility:

taking the \$1K and delaying would net you \$1K but result in your taking the empty Box A, because taking the \$1K makes no difference to your choice between the boxes. Similarly, if you are certain that H_2 is true then you are certain that you are of Type B and that the money is in Box A. In that case, directly taking Box A would have maximal causal utility: taking the \$1K and delaying would net you \$1K but result in your taking the empty Box B, because it would make no difference to your choice between the boxes.

What if you are close to 50% confident that H_2 is true, and close to 50% confident that H_3 is true? In that case you are nearly 50% confident that taking Box A directly would get you \$1M, because you are nearly 50% confident that the \$1M is in Box A. You are also nearly 50% confident that taking Box B directly would get you \$1M, because you are nearly 50% confident that the \$1M is in Box B. But you are nearly 100% confident that taking the \$1K and delaying the choice between Box A and Box B would get you just \$1K, because you are nearly 100% confident both that your ultimate choice has been correctly predicted and that delay would make no difference either to your choice of box or to the prediction of it. So in this case *both* taking Box A directly *and* taking Box B directly have greater causal utility than taking the \$1K and delaying.

More generally, if you are confident that one of H_2 and H_3 is true then *whatever* the exact distribution of your confidence between these two hypotheses, taking one of the boxes directly always maximizes causal utility; taking the \$1K before choosing between them *never* does.[8] So Causal Decision Theory invariably advises its followers not to take the \$1K on offer at the first stage of this game.[9] [10]

4. A strange theory indeed! – But a lucrative one, if not for its followers then at least for the producers of *Frustration and Delay*, which soon displaced *Newcomb!* as the leading philosophical decision theory-based game show across the Virgo supercluster.

Speaking of *Newcomb*, there was a certain irony in the fact that one often heard the following points in connection with that game.

- Taking the $1K would never *make* any difference to whether one also made $1M
- Declining it would therefore amount to giving up a free $1K: anyone who declined it would be $1K poorer than if they had accepted it.
- Everyone knew all this in advance.

Followers of CDT had often taken it to be absurd on these grounds to decline the $1K on offer in *Newcomb*. The irony was that these points all applied equally to the $1K on offer in *Frustration and Delay*, which CDT did recommend declining. That advice, and indeed CDT itself, soon came to seem as arbitrary and unmotivated to many of its former defenders as it had all along to its longstanding opponents.[11]

References

Ahmed, A. 2014. Dicing with death. *Analysis* 74: 587-92.

Armendt, B. 2019. Causal Decision Theory and decision instability. *Journal of Philosophy* 116: 263-77.

Arntzenius, F. 2008. No regrets, or: Edith Piaf revamps decision theory. *Erkenntnis* 68: 277-297.

Gibbard, A. and W. Harper. 1978. Counterfactuals and two kinds of expected utility. In Hooker, C., J. Leach and E. McClennen (ed.), *Foundations and Applications of Decision Theory*. Dordrecht: Riedel: 125-62. Reprinted in Gärdenfors, P. and N.-E. Sahlin (ed.), *Decision, Probability and Utility* (1988). Cambridge: CUP.

Hájek, A. 2016. Deliberation welcomes prediction. *Episteme* 13: 507-28.

Joyce, J. 1999. *Foundations of Causal Decision Theory*. Cambridge: CUP.

———. 2012. Regret and instability in Causal Decision Theory. *Synthese* 187: 123-45.

———. 2018. Deliberation and stability in Newcomb problems and pseudo-Newcomb problems. In Ahmed, A. (ed.), *Newcomb's Problem*. Cambridge: CUP: 138-59.

Levinstein, B, and N. Soares. Forthcoming. Cheating death in Damascus. *Journal of Philosophy*.

Lewis, D. K. 1979. Counterfactual dependence and time's arrow. *Nous* 13: 455-76

———. 1981. Causal Decision Theory. *Australasian Journal of Philosophy* 59: 5-30. Reprinted in his *Philosophical Papers Vol. II*. Oxford: OUP 1986: 305-339.

Liu, Y. and H. Price. 2019. Heart of DARCness. *Australasian Journal of Philosophy* 97: 136-50.

Nozick, R. 1969. Newcomb's problem and two principles of choice. In N. Rescher (ed.), *Essays in Honor of Carl G. Hempel*. Dordrecht: D. Reidel: 114-46. Reprinted in P. Moser (ed.), *Rationality in Action: Contemporary Approaches*. Cambridge: CUP 1990: 207-34.

Rabinowicz, W. 2002. Does practical deliberation crowd out self-prediction? *Erkenntnis* 57: 91-122.

Richter, R. 1984. Rationality revisited. *Australasian Journal of Philosophy* 62: 392-403.

Skyrms, B. 1980. *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws*. New Haven: Yale UP.

———. 1990. *Dynamics of Rational Deliberation*. Cambridge, MA: Harvard UP.

Spohn, W. 1977. Where Luce and Krantz really do generalize Savage's decision model. *Erkenntnis* 11: 113-34.

Sobel, J. H. 1989. Partition theorems for causal decision theories. *Philosophy of Science* 56: 71-93.

Von Neumann, J. and O. Morgenstern. 1953. *Theory of Games and Economic Behaviour*. Third ed. Princeton: Princeton UP.

---

[1] More precisely: at least 99% of subjects who choose Box A display the A-signature in advance, and at least 99% of subjects who choose Box B do *not* display the A-signature in advance.

[2] This game is a version of *Death in Damascus* (Gibbard and Harper 1978: 372). More realistically: imagine that smoking harms you iff you possess a gene that inclines you to smoke.

[3] The Newcomb protocol (cf. Nozick 1969: 208) was as follows. The contestant faces two boxes A and B, A being opaque and B transparent and with \$1K in B, and the following instructions:

> *Newcomb*. You must take *either* just Box A, *or* Box A and Box B (and you keep what you take). Before you choose, you should know that five minutes ago we conducted a brain scan to determine what you would ultimately do. The scan detects a short-lived electrical signature – the *E-signature* – whose presence (absence) predicts that you will take Box A only (both boxes) with at least 99% accuracy. If the scan detected the E-signature then we put \$1M in Box A. If the scan detected no E-signature then we put nothing in Box A. Now do you want to take just Box A, or Box A and Box B?[3]

Of course most people who took only A left with \$1M and most people who took both boxes left with only \$1K. The two-boxing subjects all reasoned *ex ante* that taking only A would be throwing away a free \$1K; and they almost all consoled themselves *ex post* that if *they* hadn't taken both boxes they'd have ended up with nothing.

[4] More formally, we can define causal utility in terms of *dependency hypotheses*.

Suppose an underlying set W of possible worlds over whose power set a probability function Cr is defined that represents your subjective degrees of belief. Let there be the following partitions of W: a set $O = \{O_i | i \in I\}$ of options and a set $Z = \{Z_j | j \in J\}$ of possible outcomes, and let there be a value function $v: Z \to \mathbb{R}$ taking outcomes to scores. (For present purposes we can identify the score of an outcome with the dollar value of your prize in that outcome.) For any option O_i and outcome Z_j, let the proposition $O_i > Z_j$ be true at a world w if and only if the closest O_i-world to w is a Z_j-world, where in particular the closest O_i-world to any world w matches w on all particular matters of fact that are causally independent of O_i and on its laws as far as possible.

Now for any function $k: I \rightarrow J$, let the corresponding *dependency hypothesis* H_k be a proposition of the form $\bigwedge_{i \in I} O_i > Z_{k(i)}$. So if we write $K = J^I$ for the set of functions from I to J, the set $H = \{H_k | k \in K\}$ forms a partition, and for any $O_i \in O$ and $H_k \in H$ we have an associated outcome $Z_{k(i)} \in Z$. The causal utility of an option O_i is defined as $U(O_i) = \sum_{k \in K} Cr\,(H_k) v\big(Z_{k(i)}\big)$. Given any set of options CDT advises you to choose the option (or any option, if there is more than one) that maximizes U.

This definition of U-score roughly follows Lewis (1981: 313), one simplification being that the counterfactual $>$ is defined here in explicitly causal terms, unlike the account in Lewis 1979. The original idea behind CDT was due to Stalnaker (1980). For alternative definitions to which the present argument also applies, see Gibbard and Harper 1978: 345f.; Skyrms 1980: 133; Sobel 1989: 73; Joyce 1999: 161.

[5] In the Lewisian terms of n. 4, we can set up the problem, and the dependency hypotheses, as follows. The option set is $O = \{A, B, X\}$. The outcome set is $Z = \{Z_0, Z_K, Z_M, Z_{M+K}\}$ where Z_n is the outcome that you win \$$n$. So there are $4^3 = 64$ possible dependency hypotheses. But many of these can be ignored: for instance, you have zero credence in any of the sixteen dependency hypotheses that entail $A > Z_{M+K}$, because if you were to take Box A directly then you would have no prospect of making more than \$1M. Similar reasoning rules out a further 44 dependency hypotheses, leaving the following four, each of which is equivalent to one of the four hypotheses specifying which box you will ultimately take and where the money is as set out in Table 1:

$$H_1: A > Z_M\,.\,B > Z_0\,.\,X > Z_{M+K}$$

$$H_2: A > Z_0\,.\,B > Z_M\,.\,X > Z_K$$

$$H_3: A > Z_M\,.\,B > Z_0\,.\,X > Z_K$$

$$H_4: A > Z_0\,.\,B > Z_M\,.\,X > Z_{M+K}$$

Assuming value linear in money, we can substitute these values into the equation for U in n. 4 to give:

$$U(A) = MCr(H_1 \vee H_3)$$

$$U(B) = MCr(H_2 \vee H_4)$$

$$U(X) = K + MCr(H_1 \vee H_4).$$

---

[6] I have used causal dependency hypotheses that describe the agent's ultimate choice; and this might cause concern if you think that there is some incoherence in the idea that an agent can assign subjective probability to an option that she is currently contemplating (Spohn 1977). I doubt that such assignments *are* incoherent (see e.g. Rabinowicz 2002, Hájek 2016; for a response see Liu and Price 2019). But in any case, we can instead use causal dependency hypotheses that describe the state of the world (and the agent) *prior* to the current decision.

To this end, let S_A say that the brain scan (which we assume infallible) detected the A-signature in the brain of the agent, and let S_B say that no A-signature was detected. Assume that the A-signature makes the chance high that you will ultimately take Box A, and that its absence makes this chance low. Then the propositions S_A and S_B can serve as dependency hypotheses, because fixing whether the A-signature was present fixes both the location of the \$1M – the \$1M is in Box A iff it was present – *and* the *ex ante chance* of ultimately taking Box A, because if it was present then you are likely ultimately to take Box A.

To apply these dependency hypotheses we need to amend the set of outcomes to include (objective) lotteries over monetary amounts. Suppose that given that you *have* the A-signature your *ex ante* chance of ultimately taking Box A $c \approx 1$. Similarly suppose that given that you *lack* the A-signature your *ex ante* chance of ultimately taking Box A is low: suppose for convenience that it is $1 - c \approx 0$ (though any small enough quantity will do). I'll write $L^p_{1-p}(X, Y)$ for a lottery that has a chance of p of winning \$$X$ and chance $1 - p$ of winning \$$Y$. Then the possible dependency hypotheses can be written as conjunctions of counterfactuals as follows:

$$S_A\colon A > Z_0 \,.\, B > Z_M \,.\, X > L^c_{1-c}(K, M + K)$$

$$S_B\colon A > Z_M \,.\, B > Z_0 \,.\, X > L^c_{1-c}(K, M + K)$$

To see why *both* hypotheses entail $X > L^c_{1-c}(K, M + K)$, note that if e.g. the A-signature was detected, then the \$1M is in box B but there is a high chance c that you will ultimately take Box A; and delaying the choice by taking the \$1K has no tendency to reduce this chance. So if the A-signature was detected, then taking the \$1K commits you to a lottery in which you have a very high chance of winning \$1K overall (the \$1K you have just taken plus nothing in Box A) and a very low chance of winning \$1K+\$1M overall. Similarly, if the A-signature was *not* detected, then taking the \$1K commits you to a lottery in which you have a very high chance of winning \$1K overall (the \$1K you have just taken plus nothing in Box B, which you are now likely to take) and a very low chance of winning \$1K+\$1M overall.

On von Neumann and Morgenstern's assumptions concerning the values of lotteries (1953: 16-27), and assuming linear value for money, the value of the lottery that X involves is as you would expect:

$$v\left(L^{c}_{1-c}(K, M+K)\right) = K + M(1-c) \approx K.$$

Applying the formula in n. 4:

$$U(A) = MCr(S_B)$$

$$U(B) = MCr(S_A)$$

$$U(X) = K + M(1-c)$$

Since $Cr(S_A) + Cr(S_B) = 1$ it follows that either $U(A) > U(X)$ or $U(B) > U(X)$ i.e. CDT rejects the $1K.

[7] The formal argument is as follows. Because you know that the scan is ≥99% accurate, you know that ≥99% of Type A persons will face a situation in which there is $1M in Box A, and ≥99% of Type B persons will face a situation in which there is $1M in Box B. Write T_A for the proposition that you will ultimately take Box A and T_B for the proposition that you will ultimately take Box B. Write M_A for the proposition that the $1M is in Box A and M_B for the proposition that the $1M is in Box B. Then if (as we can assume) your credences track these frequencies, we have:

(i) $Cr(M_A|T_A), Cr(M_B|T_B) \geq 0.99$

(ii) $H_2 \vee H_3 = M_A T_A \vee M_B T_B$ by the definition of H_2, H_3.

(iii) $Cr(H_2 \vee H_3) = Cr(M_A T_A) + Cr(M_B T_B)$ by (ii), since $M_B T_A \cap M_B T_B = \emptyset$

(iv) $Cr(H_2 \vee H_3) = Cr(M_A|T_A)Cr(T_A) + Cr(M_B|T_B)Cr(T_B)$ by (iii) and the probability calculus.

(v) $Cr(H_2 \vee H_3) \geq 0.99\left(Cr(T_A) + Cr(T_B)\right) = 0.99$ by (i) and (iv).

[8] Proof: consider $D =_{\text{def}} U(A) + U(B) - 2U(X)$. Then

(i) $D = MCr(H_1 \vee H_3) + MCr(H_2 \vee H_4) - 2K - 2MCr(H_1 \vee H_4)$

$= MCr(H_2 \vee H_3) - MCr(H_1 \vee H_4) - 2K$, by the argument in n. 5.

(ii) $Cr(H_2 \vee H_3) \geq 0.99$ by the argument in n. 6.

(iii) $Cr(H_1 \vee H_4) \leq 0.01$ by (ii).

(iv) $D \geq 0.98M - 2K > 0$ by (i) and (iii).

(v) $U(A) + U(B) > 2U(X)$ by (iv) and the definition of D

(vi) $U(A) > U(X)$ or $U(B) > U(X)$ by (v)

But (vi) means that CDT always prefers at least one of A and B to X.

[9] For comparison, *Evidential* Decision Theory (EDT) recommends the option that maximizes news value, where the news value of an option, in the terms of n. 4, may be written $V(O_i) = \sum_{k \in K} Cr\,(H_k|O_i)v\big(Z_{k(i)}\big)$. The news value of an option represents how pleased the agent would be to learn that he has realized it: EDT therefore differs from CDT in so far as it takes account of the evidential bearing of an option on states (for example, the location of the $1M) that it does nothing to affect. Taking Box A directly is very strong evidence that the $1M is in Box B, and taking Box B directly is very strong evidence that the $1M is in Box A. So $Cr(H_2|A) \approx 1$ and $Cr(H_3|B) \approx 1$; therefore $V(A), V(B) \approx 0$. But taking that $1K and delaying tells you nothing about which box you will eventually choose; doing so leaves you highly confident that *either* you will ultimately take Box A and the $1M is in Box B *or* you will ultimately take Box B and the $1M is in Box A. Therefore $Cr(H_2 \vee H_3|X) \approx 1$ so $V(X) \approx K$. So $V(X) > V(A), V(B)$ i.e. EDT always recommends first taking the $1K over directly taking either box.

[10] The same argument works for all versions of CDT mentioned in n. 4. The underlying reason is that CDT takes a very different attitude towards options that are under my *present* control (like A, B and X) as opposed to those in the *future*, like the future options of taking Box A and taking Box B that would I arise if I were now to take the $1K. 'Viewed in prospect, acts in future decisions are treated not as current options, but as potential outcomes lying causally downstream of your current choice. As with anything not under your current control, CDT assesses future acts using their current news values' (Joyce 2018: 146; 'news value' here means V as defined in n. 9). The current *news value* of my taking Box A (Box B) *after delaying* is about $1K, because news that I will do this strongly indicates that the $1M is in Box B (A).

The argument also works against the *deliberational* versions of CDT recently proposed by Arntzenius (2008) and Joyce (2012), following Skyrms 1990. (For discussion of the differences between Arntzenius and Joyce, and some criticisms of the deliberational approach from a perspective that is sympathetic to CDT, see Armendt 2019.) .According to this theory, the upshot of rational decision-making is an equilibrium in which you (a) have a probability distribution over which outcome you will realize such that (b) on this distribution, CDT reckons each option given non-zero probability to be at least as good as any other option. To calculate the equilibrium in the present problem, note that by the argument of n. 7, $Cr(H_2 \vee H_3) \geq 0.99$ on *any* distribution of your confidence over the options A, B, X. By the argument of n. 8, we have $U(A) > U(X)$ or $U(B) > U(X)$ in any such distribution, and therefore $Cr(X) = 0$ in any deliberational equilibrium. So deliberational CDT rejects the $1K offer in this problem.

[11] I should distinguish the case studied in this paper from two other alleged counterexamples to CDT.

In a mildly asymmetric version of *Death in Damascus* due to Richter that Levinstein and Soares have recently discussed (Richter 1984; Levinstein and Soares forthcoming), the arrangement resembles *Frustration* except that there is (you know) an additional $1000 in Box B. Here it might seem obvious that taking Box B is the only rational choice; but CDT (if it advises anything at all) appears also to endorse taking Box A. At least, if you are slightly more confident that the $1M is in Box A than that it is in Box B, which in a deliberational equilibrium (se n. 10) you will be, CDT is indifferent between these options. But arguably if those are your credences, then taking Box A *ought* to look as good as taking Box B: since you are slightly more confident that Box A is worth $1M, you should be just willing to take it, even if that means foregoing the $1K associated with Box B.

Similarly, *Dicing with Death* (Ahmed 2014) modifies *Frustration* by offering a third option, which in present terms amounts to paying a small amount for the use of a randomizing device that chooses unpredictably between Box A and Box B. Ahmed objects to CDT that it rejects this option. Joyce's response is that if you are 50-50 about the whereabouts of the $1M, you have 50% confidence that taking Box A directly gives you a 100% chance of winning $1M and 50% confidence that it gives you a 100% chance of winning nothing; and the same goes for the option of taking box B directly. As for randomization, you have 100% confidence that it gives you a 50% chance of winning $1M and a 50% chance of winning nothing. Setting aside irrational Ellsberg-type preferences ('ambiguity aversion'), these gambles *should* all look equally good. 'In terms of your subjective estimates of [probabilities of winning $1M], all three acts offer the same thing. So, paying to [randomize] would be paying for what you already take yourself to have' (Joyce 2018: 156).

But whatever its merits against *Dicing with Death* or asymmetric *Death in Damascus,* this response is ineffective against *Frustration and Delay*. As in *Dicing with Death*, you are 50-50 about the whereabouts of the $1M, you have 50% confidence that taking Box A (or B) directly would give you a 100% chance of winning $1M, and you have 50% confidence that it would give you a 100% chance of winning nothing. But you *also* have 50% confidence that taking Box A (B) *after taking the $1K* would give you a 100% chance of winning $1M and 50% confidence that it would give you a 100% chance of winning nothing; and the same goes for the option of taking

Box B after taking the $1K. Taking the $1K now gives you a later choice between gambles that are just as good, by Joyce's own standards, as the ones available now. Taking Box A or Box B directly, as CDT recommends, would in effect be giving up a $1K bonus for a gamble that you would have got anyway.