Counterfactuals, Indeterminacy, and Value: A Puzzle

Andrew Peet & Eli Pitcovski

October 25, 2021

Abstract

According to the Counterfactual Comparative Account of harm and benefit (CCA), an event is overall harmful (/beneficial) for a subject to the extent that this subject would have been better (/worse) off if it had not occurred. In this paper we present a challenge for CCA. We argue that if physical processes are chancy in the manner suggested by our best physical theories, then CCA faces a dilemma: If it is developed in line with the standard approach to counterfactuals, then it delivers that the value of any event for a subject is indeterminate to the extreme, ranging from terribly harmful to highly beneficial. This problem can only be avoided by developing CCA in line with theories of counterfactuals that allow us to ignore a-typical scenarios. Doing this generates a different problem: when the actual world is itself a-typical, problematic implications will emerge. For example, we will sometimes get the result that the counterfactual nonoccurrence of an actual benefit is itself a benefit. An account of overall harm bearing either of these two implications is deficient. Given the general aspiration to account for deprivational harms and the dominance of CCA in this respect, theorists of harm and benefit face a deadlock.

1 Introduction

What makes an event overall harmful or beneficial for a subject? The most prevalent approach to overall harm/benefit is the Counterfactual Comparative Account (CCA),[1] according to which an event E is overall harmful (/beneficial)

[1] Among others, adherents include: Feldman (1991), Broome (1999, 2004), Bradley (2009), Feit (2015), Purves (2014, 2019), Klocksiem (2012)

for a subject S to the extent that S is worse (/better) off than they would be had E not occurred.[2]

In this paper, we present a trilemma between the following options:

1. Reject CCA.

2. Accept that it will almost always be indeterminate whether an event E is harmful to a subject, and that this indeterminacy is extreme in the sense that the value of every event is indeterminate between a huge range of values: from terribly harmful to highly beneficial.

3. Reject several appealing principles concerning the relationship between harm, benefit, and comparative degrees of benefit.

More specifically, we argue that if the world is indeterministic, CCA and the standard approach to counterfactuals jointly impose option 2. We call this the *Indeterminacy Problem*. We further argue that CCA and any of the non-standard approaches to counterfactuals capable of saving CCA from the untenable implications of option 2, jointly impose option 3. We call this the *Asymmetry Problem*.

The paper proceeds as follows: in section 2 we provide some background about indeterministic laws and counterfactuals. In section 3 we outline CCA and introduce the *Indeterminacy Problem* in more detail. In section 4 we consider responses to the *Indeterminacy Problem* and argue that they all succumb to the *Asymmetry Problem*. In section 5 we consider our options. Indeterminacy can be made more palatable by introducing degrees of truth. However, we argue that this gives rise to a version of the *Asymmetry Problem*. Biting the bullet on the *Asymmetry Problem* is also unattractive. It poses a burden for practical reasoning, and involves divergent judgments for intuitively similar cases. We consider the possibility that the *Asymmetry Problem* can be ignored as it only arises in weird marginal cases, and we argue that this is not the case. Rejecting CCA seems like the best option, but as things stand, doing so creates a theoretical vacuum, given that CCA is the only account on offer capable of dealing with deprivational harms such as the harm of death.

[2] CCA is standardly formulated in terms of the nearest $\neg E$ *world*. As we will explain shortly, this is a simplification that, when eliminated, generates a form of indeterminacy. However, the resulting indeterminacy is not especially worrying.

2 Background: Counterfactuals and Physics

We begin with a familiar problem from the theory of counterfactuals. The standard view of counterfactuals tells us that a counterfactual of the form "if A had happened then C would have occurred", (or, $A \mathrel{\square\!\!\rightarrow} C$), is true iff all of the nearest A-worlds are also C-worlds. The A-worlds that are nearest to the actual world will be those that i) match the actual world until (shortly before) the antecedent time, and, ii) match the actual world in physical law (but not necessarily in matters of particular fact) after the divergence (see Stalnaker, (1968), Lewis (1973), Bennett (2003)).[3,4]

The source of the problem is that according to certain popular interpretations of quantum mechanics a system's wave function does not fully determine the evolution of the system. Rather, it delivers objective probabilities of locations.[5] We can picture this as follows: suppose we are considering a system of particles in a three dimensional space. We can think of the space as being divided up into a large three-dimensional grid, with each particle being assigned its own cell. We can specify the state of the system at a time by specifying the location of each particle on this grid. We can then model the history of the system from an initial starting point at t_1 by specifying the state of the system at each subsequent time. If the laws governing the system are deterministic then, at least in principle, we would be able to derive the state of the system at each subsequent time from the state of the system at an initial starting time t_1 together with the laws governing the system. Moreover, counterfactuals of the form "had the system been in state S_1 at t_1, it would have been in state S_2 at t_2" will be unproblematic. That is, for any given starting point S_{x}, there

[3]Lewis's account of closeness differs slightly as he avoids appeal to temporal ordering in an attempt to provide a counterfactual theory of the direction of time. He also later modified his account (1986) in response to the problem we discuss here. We'll discuss Lewis's modified view in section 4.

[4]If we assume determinism the divergence is typically thought to require a small 'miracle' - i.e. a minor deviation from the laws of nature shortly before the antecedent time. Alternatively the determinist can hold that the initial conditions of the universe must have been slightly different (see Dorr (2016)).

[5]Any interpretation of quantum mechanics according to which the wave function delivers objective probabilities of locations will have this result. The most obvious example is the GRW theory (see https://plato.stanford.edu/entries/qm-collapse/). Not all interpretations of quantum mechanics will have this result. For example, Bohmian approaches treat quantum mechanical systems as deterministic. Ultimately, if it turns out that a deterministic interpretation of quantum mechanics is correct this would do little to undermine the argument we present here. After all, it is certainly possible for an indeterministic interpretation to be correct. So, there will be some possible world w which is like ours but governed by indeterministic laws. It would be problematic for a theory of harm if, for example, it entailed that genocides that occur at w are not determinately harmful to their victims.

will be a single state S_{x+n} such that, had the system been in S_x it will be in state S_{x+n} after n transitions. However, if the laws governing a system are not deterministic, we will not be able to derive the state of the system at each subsequent time from the state of the system at an initial starting time t_1.

To illustrate, suppose that the laws governing our system tell us that for any given state S_1, a given particle could move to any cell adjacent to the cell it occupies at S_1. The laws do not settle which adjacent cell it will move to. If this is the case then we won't be able to derive the state of the system at subsequent times from the state of the system at an initial starting time. We will only be able to assign probabilities to various possible outcomes for the system. For example, suppose the particles in our system are reasonably evenly distributed at t_1, and suppose the system goes through a series of n transitions. If each particle has an equal probability of entering any adjacent cell during any given transition, we will not be able to predict what state particles in the system will be in, but there will be a much higher probability that they will remain reasonably evenly distributed than there is of them congregating in one small corner of our grid. If we assign velocity and direction to our particles at each given state, we should sometimes expect a different result, but the basic idea would be the same.

We can think of real physical systems along similar lines. The wave function for a given particle takes anything that affects the behaviour of particles (initial position, velocity, direction, spin etc.) and delivers the probability of the location of the particle at subsequent times. The wave function for a given particle will not determine a specific location for the particle at any given time. The wave function for a system takes the initial locations, velocities, directions etc. of all the particles in the system and delivers probabilities for the state of the total system at subsequent times. So, given the initial state of a system, we cannot derive the state of the system at any given subsequent time.

This has some surprising consequences: Suppose the system we are considering is the actual world. And suppose I let go of a ball. There is, it turns out, a small chance that the ball would remain suspended in the air. Take S_1 to represent the initial state of the world before I drop the ball. There will be a set of possible developments of the system corresponding to the outcome whereby the ball remains suspended in the air. The probability of the system developing in any of these ways will be vanishingly small. Nonetheless, some of these developments of the system will be assigned a non-zero probability. Now suppose that I don't let go of the ball. Instead, I consider what would happen

if I had done so. I think the following:

(1) If I had let go of this ball it would have fallen to the ground.

(1) strikes us as true. However, according to the standard view of counterfactuals its truth requires that the ball falls at *all* the nearest worlds at which I let go. There is at least one world at which the ball remains suspended in the air. But is this world among the nearest worlds to actuality? Well, this world can match the actual world exactly up to the antecedent time. Moreover, there is no violation of physical law (even after I let go). So it should be among the nearest possible worlds at which I let go. But that would render (1) false. The problem generalizes and seems to render the vast majority of ordinary counterfactuals false (Hajek (MS), Hawthorne (2005)). This problem has far reaching implications for CCA.

3 Overall Harm: Sources of Indeterminacy

Taking into account every way in which events can harm and benefit us (not just *pro tanto* but overall), what determines the extent to which some event is, all things considered, harmful or beneficial for a subject? The most popular response to this question is by means of the Counterfactual Comparative Account (CCA). On a rough first pass formulation, it is the view that:

> **CCA** An event E is harmful for a subject S iff S is worse off in the actual world (in which E occurs) than in the nearest possible world in which E does not occur (the nearest $\neg E$ world). The degree to which an event E is harmful for S is the degree to which S is worse off in the actual world than in the nearest possible world in which E does not occur.[6] (For an account of overall benefit, replace 'harm' with 'benefit' and 'worse' with 'better'. We will henceforth not systematically include the "beneficial/benefit" qualification and, when not specifically addressing benefit, focus on harm).

Worse and better off are cashed out in terms of the aggregated intrinsic value (for S) of the totality of states at each relevant world. So, for instance, (using

[6]There are various slightly different formulations of the view, some of which do not account for degrees of harm, and only some of which account for overall benefit. Given the purposes of this paper, we will stick to CCA. For similar formulations, see Bradley (2009, p.50), Broome (1999, 2004), Feldman (1991, p.150).

negative value for intrinsic harm), a subject is worse off at some world w_1 than she is at some other world w_2, iff the aggregate *intrinsic* value of states in w_1 is lower than the aggregate *intrinsic* value of states in w_2; if her states in w_1 are altogether more *intrinsically* bad for her than her states in w_2. Accounts of overall harm can be complemented by an account of intrinsic harm. Here, however, we will focus on accounts of overall harm.[7]

Most CCA theorists adopt the simplifying assumption that there will always be some closest $\neg E$ world that we can compare to actuality (see, for example, Feldman (1991) and Feit (2002)). When this assumption is eliminated value-indeterminacy quickly looms. Assume determinism for the time being, and consider the following case:

> **Driver** Ruth is driving in the woods. She is feeling spontaneous, so she decides that when she reaches the next junction she will toss a coin. If the coin lands heads she will turn right. If it lands tails, she will turn left. Unbeknownst to Ruth, if she does indeed turn right, it is very likely that she will find a treasure. If she turns left, it is very likely that she will be kidnapped and enslaved by a cruel gang. As it turns out, Ruth suddenly dies before she ever reaches the junction.

The nearest possible worlds at which Ruth lives and the coin lands heads are just as close to actuality as those at which she lives and the coin lands tails.[8] So, it seems the CCA theorist must claim that it is indeterminate whether Ruth's death harms her (assuming that kidnapping and enslavement leads to a life not worth living). After all, CCA asks us to consider the closest world at which Ruth lives on. Yet, there is no single nearest world at which she lives. And the various worlds at which she lives have highly divergent intrinsic values for Ruth.

We don't find this indeterminacy to be especially worrying. It seems per-

[7]Intrinsic harm and intrinsic benefit are often understood in terms of levels of well (/ill)-being. This is common to hedonist accounts of well-being (Feldman (2004), desire based accounts, like Heathwood (2019; 2014) and objective list theories, such as Harman (2004), Griffin (1986), Finnis (2011)). (Woodard (2013) is an exception in this respect). CCA (and other views about overall harm, rather than intrinsic harm) can remain neutral with respect to theories of well-being. Bradley (2012) takes this axiological neutrality to be a desideratum for an adequate theory of overall harm.

[8]Note that determinism does not entail that there is a fact of the matter regarding whether the coin would have landed heads or tails had it been flipped. If determinism is true then for Ruth not to have died the conditions leading up to the antecedent time must have been slightly different (either due to a small localized violation of the laws of nature, or a slight difference in the initial conditions of the universe). Different divergences from actuality here will lead to different coin flip outcomes. And no such divergence is privileged such that, had Ruth not died, her survival would be due to one particular divergence rather than some other.

fectly reasonable for the CCA theorist to simply accept that the value of an event will typically be indeterminate (Feit (2002) gestures in this direction). To determine the value of an event, E, for a subject, S, we will look at the aggregate intrinsic value of states in each nearby $\neg E$ world for S, and extract the range of values between which E is indeterminate by determining how much (intrinsically) better or worse each of these worlds is for S in comparison to the actual world. For example, suppose that in the worlds at which Ruth finds the treasure the states in the portion of life following the coin toss have an aggregated intrinsic value of 5, and the worlds at which she is kidnapped have a value of -5. CCA theorists can hold that the value (i.e. harm/benefit) of Ruth's death is indeterminate between the values 5 and -5.

Assuming determinism, this seems like the right result. However, things start to look a lot worse for CCA when we take into account the kind of physical indeterminism thought to undermine the standard account of counterfactuals. Recall our earlier example: suppose I am holding onto a ball. Our best physical theories tell us that at some very small proportion of the nearest worlds at which I let go of the ball it remains suspended in the air. Compare this to a subject S's untimely demise at the age of 20. At the majority of nearby worlds at which they persist their life will be worth living. However, there will typically be some small proportion of nearby worlds at which the system develops in a disastrously a-typical manner. For example, there will typically be some nearby world at which an unlikely quantum fluctuation damages the pain center of their brain causing them to live in excruciating pain for the rest of their life (call worlds at which such improbable and bizarre quantum events occur "quantum weirdness worlds").

At this world S's life will not be worth living. Let us suppose that the relevant portion of S's life at this world has a value of –10, and let us add it to the set of equally nearby worlds at which S persists. When S's life up to the time of death has a value X, suppose that the values of the nearest worlds at which S lives are as follows: w_1=X+2, w_2=X+4, w_3=X+6, w_4= X-10. So, the value of S's death is indeterminate between the values 10 and -6. It is no longer determinately the case that S's death is harmful to them.

This problem generalizes. For *any* event E that we think of as bad for a subject there will typically be some nearest $\neg E$ possibility which is far worse for the subject. So, it will rarely if ever be the case that any event is determinately harmful for a subject. The same is true of benefit. For any event E that we would normally take to be beneficial for a subject S there will typically be some

nearby quantum weirdness world at which they are far better off following E than they are in actuality. So, no event will ever be determinately beneficial for a subject either. Indeed, for any given event E there will typically be some nearby $\neg E$ quantum weirdness worlds at which the subject is far better off than they are in actuality, and some at which they are significantly worse off than they are in actuality. So the value of any event E will typically be indeterminate between a very wide range of values. Thus, the value of any event for a given subject will be radically indeterminate. This is the *Indeterminacy Problem*.

4 Overcoming Indeterminacy leads to asymmetry

4.1 Alternative accounts of counterfactuals, alternative versions of CCA

Our problem is as follows: CCA tells us that an event E is harmful for a subject S iff the nearest worlds at which E does not occur are better for S than the actual world. According to the standard view of counterfactuals, the nearest worlds at which an event E does not occur will be those that match the actual world until (shortly before) the occurrence of E, and which match the actual world in physical law. However, due to the chancy nature of physical laws, there will typically be a very small number of very weird possibilities (quantum weirdness worlds) among the closest worlds at which E does not occur. And some of these possibilities will be very bad for the subject. As a result events will rarely if ever be determinately harmful for a subject.

As we saw in section 2, this problem is derived from a well-known problem for counterfactuals. The literature on counterfactuals is replete with responses to this problem. So, the natural next step is to consider whether any of these responses offers hope to CCA. There are a number of approaches that do help CCA avoid the indeterminacy problem. However, they all give rise to a problem of their own: they all force us to give up on some key principles regarding the relationship between harm, benefit, and degrees of benefit.

To see this, it will be helpful to briefly survey some of the responses to the physical chanciness argument for counterfactual error theory, and consider analogous responses to the *Indeterminacy Problem* for CCA.

The responses can be divided into three categories. Firstly, there are those

that hold that for any counterfactual $A \mathrel{\square\!\!\rightarrow} C$ there is always a single closest A world to actuality. In cases of indeterminacy it will simply be a brute metaphysical fact that one world is closest (call this BRUTALISM). Consider again our sample counterfactual:

(1) If I had let go of this ball it would have fallen to the ground.

BRUTALISM tells us that there will always be a single nearest world in which I let go of the ball, and that the truth of (1) will depend on what happens at this world. If the ball falls to the ground in this single nearest world then (1) is true. Since the probability of the ball remaining suspended in the air is extremely low, the chances are that the ball will fall to the ground at the nearest world at which I let go. After all, there is nothing to choose between the various nearby worlds at which I let go of the ball, it is simply a brute fact that one of them is the nearest world. And, the ball falls to the ground at the vast majority of these worlds. So, the probability that the ball falls at the nearest world to actuality is extremely high. As a result, the probability of (1) being true in any given case is also extremely high. BRUTALISM is advocated by Hawthorne (2005) and Stefansson (2018) (similar approaches are advocated by Moss (2013) and Schultz (2014)).

The second species of response ("MOST WORLDS") holds that $A \mathrel{\square\!\!\rightarrow} C$ is true iff C is true at the majority of the nearest A worlds (Bennett (2003)). Or, alternatively, that it is true if the objective chance of C is high conditional on A (Leitgeb (2012a, 2012b)). MOST WORLDS resolves the problem of counterfactual error theory as follows: although there is some small number of nearby worlds at which the ball remains suspended in the air, the probability of this happening is vanishingly low. At the vast majority of nearby worlds at which I let go of the ball it falls to the ground. So, (1) is true.

The third species of response revises the modal closeness relation. For example, Lewis (1986) incorporates the notion of a "quasi-miracle" into his account of modal closeness. A quasi miracle is a "remarkable" low probability event. And, ceteris paribus, if a world contains a quasi-miracle this renders it more modally distant than an otherwise similar world that does not contain the quasi-miracle. A ball's remaining suspended in the air is a remarkable low probability event. So, any world in which this occurs will be more modally distant than an otherwise similar world at which the ball falls to the ground. Williams (2008) provides a similar view: he builds the notion of "typicality" into the ordering relation on worlds. Typicality is a function of the probability of properties of the

event given the laws governing a world. To use Dodd's (2011) example, suppose we have a series of 10000 coin flips. Any series of flips is just as probable as any other. Yet, the series "all heads" will be less typical than any series with a roughly equal distribution of heads and tails. This is because the property "all heads" is a great deal less probable than the property "50/50 heads and tails". A-typical events function for Williams just as quasi-miracles do for Lewis. So, any world at which the series of coin flips lands all heads will be more modally distant than a world at which the series lands roughly 50/50 heads/tails. Similarly, any world in which our ball remains suspended in the air will be more distant than an otherwise identical world at which it drops to the ground. After all, the latter outcome is far more typical. Call this MODIFIED NEARNESS CCA.[9,10]

With these pictures on the table we are able to consider analogous modifications of CCA. BRUTALIST CCA can stick with our earlier formulation, viz.:

> **CCA**An event E is harmful for a subject S to the extent that S is worse off in the actual world than in the nearest possible world in which E does not occur.

BRUTALIST CCA avoids indeterminacy by holding that, for any event E, there will always be a single nearest world at which E does not occur. In the vast majority of cases this single nearest $\neg E$ world will not contain a low probability quantum event that radically alters the intrinsic value of our subject's life.[11]

[9]Although he is not primarily concerned with counterfactual error theory, Gundersen's (2004) statistical normality approach to counterfactuals delivers a similar result. This result could also be achieved by building non-statistical normality (ala Smith (2016)) into the similarity relation.

[10]There is one species of approach we have not mentioned here: contextualism (e.g. Ichikawa (2008), Lewis (2016), & Sandgren & Steele (2020)). Contextualists hold that a counterfactual $A \mathrel{\square\!\!\rightarrow} C$ is true iff C is true at all the most contextually relevant nearby A worlds, and that what we have been calling "quantum weirdness worlds" are rarely contextually relevant. In order to help with the indeterminacy problem for CCA we will need an answer to the question "why are quantum weirdness worlds not contextually relevant when talking about harm?" An answer to this question will involve identifying some property that worlds relevant to discussions of harm possess, and that quantum weirdness worlds lack. This will have to be some property that precludes weird quantum events. Possible properties include typicality (Lewis, 2016, p 306), or fittingness for the ceterus paribus laws of the domains of inquiry relevant to the conversation (Sandgren and Steele 2020). Thus, contextualist views end up being very similar to the modified nearness approach (except instead of trying to find the closeness ordering relation at play in all contexts, we are identifying a restriction on the relevantly close worlds that is at play in contexts where we are discussing harm). The problems we raise for the above approaches thus carry over straightforwardly to contextualism.

[11]BRUTALIST CCA faces what we take to be an insurmountable challenge in addition to the problems it shares with MOST WORLDS CCA and MODIFIED NEARNESS CCA: Take a paradigmatically harmful event E. The brutalist approach predicts that the vast majority of the time E

MOST WORLDS CCA will hold that an event E harms a subject iff most nearby $\neg E$ worlds are better for the subject than the actual world. Low probability quantum events that radically alter the intrinsic value of a subject's life will occur at very few nearby worlds. So most ascriptions of harm or benefit will receive their intuitive truth values.

Finally, MODIFIED NEARNESS CCA will hold that an event E harms a subject S iff all the nearest worlds at which E occurs are better for S than the actual world, with nearness being thought of in terms of typicality or some similar property that excludes quantum weirdness worlds. This will render the value of most events indeterminate in the manner spelled out for the standard account in the previous section. However, since quantum weirdness worlds are no longer amongst the nearest worlds we no longer get the result that no event is ever determinately harmful for a subject. For example, suppose our subject S dies at the age of 20. The lives they live in the various nearby worlds in which they don't die vary in quality. Yet, in most cases there will no longer be any typical nearby worlds at which they live a life worse than death.

4.2 The asymmetry problem

So, we have three responses to our problem. Unfortunately, each of these responses forces us to reject some appealing, and in some cases platitudinous principles regarding harm, benefit, and degrees of harm/benefit. The first principle is as follows:

> **Counterfactual Symmetry** If an event E is (actually) overall beneficial for a subject S, then had E failed to occur this would have harmed S (swapping "beneficial" with "harmful" and "harmed" with "benefited" yields the equivalent principle for counterfactual benefit).

COUNTERFACTUAL SYMMETRY is motivated by the following assumptions:

1. The prevention of an overall beneficial event constitutes a harm. That is, if a subject is *deprived* of a benefit, this harms the subject.

will harm the victim. However, it also allows for cases in which E is massively beneficial to the victim just because the arbitrary non-torture world that happens to be closest to actuality is a quantum weirdness world. Assuming that we cannot know that p simply on the basis of p's having a high probability (a platitude in contemporary epistemology), BRUTALIST CCA implies that we can never actually know that horrific world events such as genocides and famines harmed their victims. For all we know such events massively benefited their victims. Indeed, it is consistent with BRUTALIST CCA that no event in human history has ever been harmful, and that every event in human history had been beneficial to everyone involved. This, we think, is an unacceptable consequence.

2. If an actual event E overall benefits a subject, and the subject would not have received the same benefits had E not occurred, then E's failure to occur would have prevented the subject from receiving the benefits of E. For example, suppose that Sally needs to catch a flight in order to get to a job interview on time. She catches the flight, aces the interview, and lands the job. It seems clear in this case that, had Sally not made her flight, this would have deprived her of the opportunity to land the job.

The motivation for COUNTERFACTUAL SYMMETRY is independent from any particular view of harm. COUNTERFACTUAL SYMMETRY says that counterfactual preventions of actual overall benefits should count as (counterfactual) harms. It does not simply tell us that since Sally is better off than she would have been had she missed the flight, she would have been worse off had she missed the flight. Rather, it is committed to the slightly stronger idea that because Sally actually overall benefits from catching the flight, she would have been harmed had she missed it.

To see that COUNTERFACTUAL SYMMETRY must be rejected by the lights of BRUTALIST CCA, MOST WORLDS CCA and MODIFIED NEARNESS CCA, consider the following example:

> **Buster** Up until t_1 Buster is living a terrible life, a life not worth living. Moreover, it looks as if things are set to remain this way. At t_1 he decides to end it all: he approaches a cliff ledge and jumps. The fall would certainly kill him. However, as he jumps a passerby grabs him and pulls him to safety. As it happens, Jeff Bezos is walking by and stops to see what the commotion is about. As Bezos stands and stares at Buster, at t_2, an extremely improbable quantum fluctuation slightly alters Bezos' brain chemistry causing him to give Buster ten billion dollars. Buster uses the money to live in luxury, receive world class therapy, explore the world, gain a world class education, and contribute to countless charitable causes. He ends up living an exceptionally worthwhile life.

It was fortunate for Buster that he was saved. Things went extremely well for him in the actual world; he does far better than in the nearest world(s) in which he is not saved. Would it have been overall harmful for him not to have been saved? According to COUNTERFACTUAL SYMMETRY the answer is yes. But according to the modified versions of CCA outlined above he would not have been overall harmed had he not been saved. Indeed, he would have

benefited from not being saved. We'll demonstrate the inevitability of this result for the modified versions of CCA by considering each approach in turn.[12]

We are evaluating the harm that would have resulted from Buster's falling to his death. This is a counterfactual event. BRUTALIST CCA tells us that an event harms a subject iff the nearest world at which it fails to occur is better for a subject than the world at which it occurs. At the actual world (α) Buster does not fall. So, we need to look at the nearest world to α at which Buster falls (w_1) and assess the counterfactual "had Buster not fallen, he would have been better off" relative to this world. In order to do so we need to compare w_1 to the nearest world to w_1 at which Buster is saved (w_2). If $w_2 = \alpha$ then the fall clearly would have harmed Buster. After all, α is far better for Buster than w_1. However, there is nothing to guarantee that $w_2 = \alpha$. Indeed, there are countless worlds near to w_1 at which Buster is saved. And there is nothing to choose between these worlds - nothing to make it more probable that one such world rather than any other is the closest world to w_1 at which Buster lives. So, the probability that $w_2 = \alpha$ is extremely low. Moreover, at the vast majority of worlds nearest to w_1 at which Buster was saved he is worse off than he is in w_1. After all, in w_1 Buster's life was not worth living, and it looked set to remain that way. It was only due to an extremely improbable chance event that his life turned around in the actual world. Thus, there is a very high probability that the counterfactual "if Buster had been saved he'd have been better off" is false relative to w_1. So, according to BRUTALIST CCA, there is a very high chance that falling to his death would not have harmed Buster (indeed – it would likely have benefited him) despite the fact that he is greatly benefited by the fact that he was saved.

BRUTALIST CCA implies that Buster's death would very likely not have harmed him. Most worlds and modified nearness approaches entail that his death would not have harmed him full stop. MOST WORLDS CCA tells us that it would have been bad for Buster not to have been saved iff, at the vast majority

[12]The core assumption underlying our argument is that we can determine the truth values of counterfactuals like "If E had occurred, then this would have overall harmed S" by simply combining our best theory of harm with our best semantics for counterfactuals. CCA tells us that an event E harms a subject S whenever the subject would be better off had E not occurred. So, "If E had occurred, this would have harmed S" is true iff S is better off at the nearest E worlds (relative to actuality) than they are at the nearest $\neg E$ worlds (relative to the E worlds in question). It may be that, as an anonymous referee suggests, COUNTERFACTUAL SYMMETRY loses its intuitive appeal when spelled out like this. If this is right then the problem is either CCA or the standard accounts of counterfactuals. We suspect the problem lies with CCA. Regardless, as we will soon see, COUNTERFACTUAL SYMMETRY is not the only plausible principle the advocate of CCA will have to give up on.

of nearby worlds at which he dies, he'd have been better off had he not died. That is, it would have been bad for Buster not to have been saved iff the vast majority of nearby worlds at which he dies are such that they are worse for Buster than the vast majority of *their own* nearby worlds at which he survives. But this will not be the case. After all, it is only due to an extremely improbable quantum event that Buster's life turned around. Without this extremely improbable event (or something of its sort) Buster's life would not have been worth living. Such an event will not occur in the vast majority of worlds at issue. So, relative to the vast majority of nearby worlds at which he dies it will be the case that had he lived he'd have been worse off. So, according to MOST WORLDS CCA, had Buster died this would not have harmed him. In fact, it would have been good for him.

The same problem arises for modified nearness approaches: MODIFIED NEARNESS CCA tells us that it would have been determinately bad for Buster not to have been saved iff, at all nearby worlds at which he dies, he'd have been better off had he not died. That is, iff all the nearby worlds at which he dies are such that they are worse for Buster than all of *their own* nearby worlds at which he survives. Once again, this will not be the case. After all, it was only due to a bizarre low probability event that Buster's life turned around at the actual world. Since such events make for increased modal distance on the modified nearness account, no such world will make it into the set of relevant worlds at which he survives.[13]

So, every version of CCA capable of avoiding the *Indeterminacy Problem* leads to failures of COUNTERFACTUAL SYMMETRY.

This is a problem. However, COUNTERFACTUAL SYMMETRY is not uncontroversial. Many have criticized CCA for classifying cases that are intuitively mere failures to benefit as outright harms. For example, suppose I was intending to

[13]To put this another way, suppose that the nearest worlds to actuality at which Buster is not saved are $w_{\alpha 1}$, $w_{\alpha 2}....w_{\alpha n}$. Suppose that the nearest worlds to $w_{\alpha 1}$ at which he is saved are w_{1_1}, $w_{1_2}....w_{1_n}$, and that the nearest worlds to $w_{\alpha 2}$ at which he is saved are w_{2_1}, $w_{2_2}....w_{2_n}$ etc. In order to judge whether Buster's death would have benefited him we need to compare $w_{\alpha 1}$ to w_{1_1}, $w_{1_2}....w_{1_n}$, and $w_{\alpha 2}$ to w_{2_1}, $w_{2_2}....w_{2_n}$ etc. According to MOST WORLDS CCA Busters death would have benefited him iff the majority of worlds near to actuality at which he dies ($w_{\alpha x}$) are better than the majority of *their* nearest worlds at which he lives ($w_{\alpha x_1}$, $w_{\alpha x_2}....w_{\alpha x_n}$). But, since Buster's life was set to be not worth living, and since the change was due to a highly improbable event, this procedure yields the verdict that Buster's death would have benefited him.According to MODIFIED NEARNESS CCA Buster's death would have benefited him iff *all* of the nearest worlds to actuality at which he dies ($w_{\alpha x}$) are better than all of *their* nearest worlds at which he lives ($w_{\alpha x_1}$, $w_{\alpha x_2}....w_{\alpha x_n}$). Since Buster's life was set to be not worth living, and since the change was due to a highly improbable and a-typical event, this procedure yields the verdict that Buster's death would have benefited him.

surprise my friend with a ticket to a concert. However, I change my mind and decide to keep the ticket for myself. Many feel that my friend is not harmed in this case. Yet, they would have been better off had they been given the ticket. So, it looks as if proponents of CCA must say that their failure to receive the ticket harms them (see Bradley (2012), and Shiffrin (2012)). This criticism might be thought to clash with the motivation for COUNTERFACTUAL SYMMETRY. After all, COUNTERFACTUAL SYMMETRY was motivated by the thought that if one is deprived of a benefit one is harmed.

There are two points worth noting here. Firstly, although mere failure to benefit does not constitute a harm, it is clear that we are sometimes harmed when we fail to receive a benefit. For example, suppose that I don't change my mind about the ticket. I retain my intention to give it to my friend. However, a thief takes the ticket from me during the night. Here it seems clear that the thief's action harms my friend. And this is because it prevents my friend from receiving the ticket. We feel that **Buster** is like this. Buster's falling to his death is an event (like the ticket theft), and this event would, if it had occurred, have prevented Buster from benefiting from the weird quantum mechanical event that occurs in Jeff Bezos' brain. Thus, we think it is natural to conclude that Buster would have been harmed had he fallen to his death. His death would not merely have resulted in a failure to benefit. This seems even clearer if we consider a modification: suppose that, rather than intending to kill himself, Buster is just enjoying the view. A psychotic passer by decides to push him off the cliff, but another passer by prevents this from happening. We think it is clear that the psychotic passer by's action, had it not been prevented, would not have caused a mere failure to benefit. It would have harmed Buster. So, even if COUNTERFACTUAL SYMMETRY is too strong, **Buster** illustrates that our modified versions of CCA deliver unpalatable results.

Secondly, it is not clear to what extent Shriffin and Bradley's concerns really touch the motivation for COUNTERFACTUAL SYMMETRY. After all, COUNTERFACTUAL SYMMETRY was motivated by the thought that an event E harms a subject if it *prevents* them from receiving a benefit. And it is not clear that in cases of mere failure to benefit the subject is actually *prevented* from receiving a benefit. A benefit merely fails to materialize (Feit (2017), and Purves (2019)).[14]

[14]Importantly, even those willing to classify my change of mind as an event that harms my friend in the example above, will not classify every chance to benefit that is not taken as a harm. Following Hanser's (2008) discussion, Feit (2017) notes "...there is controversy in the literature over whether omissions are events. There are reasons to doubt the claim that such things are actions, and hence events. In order to harm someone, moreover, an event must be

Nonetheless, the notion of prevention is slippery. And **Buster** is a far fetched case. Intuitions regarding whether or not his death would have harmed him may vary. So we should not place too much weight on COUNTERFACTUAL SYMMETRY by itself.

Thankfully, we don't have to. There is a second, even more pressing problem that arises for our modified versions of CCA: all of our modified CCA theories are forced to deny one of the following two principles:[15]

> **Degree of Benefit** An event E at a world w benefits a subject S to the degree that w is better for S than the closest worlds where E fails to occur.[16]

> **Comparative Benefit** If E is good for a subject S, and S would not have been even better off had E not occurred, then E benefits S more than the non-occurrence of E would have benefited S.

To see why one of these principles must be rejected, consider the following variation on **Buster**:

> **Overall Value** Suppose that S is living a life that, absent intervening factors, is set to have -100 overall value. A process starts that would, if not prevented, bring S's life up to +150 overall value. An event E occurs that halts this process. But a bizarre quantum mechanical event occurs that could not have occurred without E, and this event bestows +300 overall value. So, S's life ends up with +200 overall value, rather than the +150 overall value S's life would have had if E had failed to occur.

According to every version of CCA we have considered, E benefits S. After all, S's life has +200 overall value. And, had E not occurred, S would not have

such that, had it not occurred, she would have been better off. Suppose that I see you on the street and fail to give you $100. Suppose that I do this by saying Hello to you. CCA does not imply that I have harmed you. After all, it need not be the case that I would have given you $100 if I had not said Hello. If this is not the case (as I hereby stipulate, and as I think is typical in such cases), then I have not harmed you." For other responses to the problem of failures to benefit as harms see Klocksiem (2012), and Hanna (2016)).

[15] Both principles concern benefit. Equivalent, equally plausible principles hold with respect to harm. The same problem will arise for these principles as well. Neither principle makes any presuppositions about the relationship between harm and benefit.

[16] This principle, integral to the version of CCA presented in section 3, is not essential to counterfactual comparative accounts of value. Nevertheless, it is hard to see how the advocate of CCA could deny it. Imagine we are comparing two independent events E_1 and E_2 at a world w, and considering which benefits S more. Suppose w has a value of 50 for S. Suppose that the nearest $\neg E_1$ worlds have a value of 45 for S, and the nearest $\neg E_2$ worlds have a value of 30 for S. Then we should say that E_2 benefits S more than E_1.

been better off. According to COMPARATIVE BENEFIT E benefits S more than $\neg E$ would have benefited S. This all seems fairly reasonable. But now consider the counterfactual "had E not occurred, this would have benefited S". According to MODIFIED NEARNESS CCA and MOST WORLDS CCA this counterfactual is true. After all, the nearest $\neg E$ world has a value of +150 for S and, according to MODIFIED NEARNESS CCA and MOST WORLDS CCA, the nearest E worlds to the nearest $\neg E$ worlds have a value of -100 for S. So, according to these approaches, if E had not occurred, this would have benefited S. BRUTALIST CCA similarly tells us that if E had not occurred this would almost certainly have benefited S. Moreover, the difference in value between the nearest $\neg E$ worlds and their nearest E worlds is 250. And the difference between the actual world and the nearest $\neg E$ world is only 50. Thus, according to DEGREE OF BENEFIT $\neg E$ would have benefited S more than E actually benefited S. So, COMPARATIVE BENEFIT and DEGREE OF BENEFIT together tell us that E benefits S more than $\neg E$ would have benefited S, and $\neg E$ would have benefited S more than E actually benefited S. Thus, if the *Indeterminacy Problem* is to be avoided, either COMPARATIVE BENEFIT or DEGREE OF BENEFIT must go.

CCA theorists are in a bind: if we formulate CCA by analogy to the standard approach to counterfactuals then it will seem that no event is ever determinately harmful for a subject (indeed, the value of any event will typically be radically indeterminate). If we modify CCA in order to eliminate quantum weirdness worlds from consideration then we get bizarre asymmetries in value between an actual event E, and its counterfactual failure to occur. Since every theory of counterfactuals will either give weight to divergent a-typical worlds or it will not, every version of CCA will either run into the radical *Indeterminacy Problem*, or it will run into the *Asymmetry Problem* (respectively).

5 Evaluating the Options

At this point we have three options:

1. Accept radical value-indeterminacy.
2. Accept counterfactual value asymmetries.
3. Reject CCA.

We close by briefly highlighting the costs of each option, tentatively suggesting that the best hopes lie in the development of alternatives to CCA.

Firstly, could we simply accept that no event is ever determinately harmful (or beneficial) for any subject? Putting aside the fact that this approach is most naturally coupled with counterfactual error theory (an unattractive position), the indeterminacy approach might not look so bad. It is not as if there is no distinction in harmfulness between, say, severe tooth ache and a pleasant dinner. Most approaches to indeterminacy allow that it can come in degrees: two propositions p and q may each be indeterminate in truth value. Yet it may still be that p has a higher degree of truth than q. With this in mind, we can simply maintain that "tooth ache is harmful" has a high degree of truth, and "tooth ache is beneficial" has a low degree of truth. Likewise "the pleasant dinner was harmful" has a low degree of truth, and "the pleasant dinner was beneficial" has a high degree of truth.

This can be captured in supervaluational terms. Suppose we are considering the sentence "E was harmful to S". CCA tells us that this is true iff the nearest $\neg E$ world is better for S than the actual world. But there is no single nearest $\neg E$ world. There are many equally nearby $\neg E$ worlds. Thinking supervaluationally, our precisifications will assign one of the nearest $\neg E$ worlds w as *the* nearest $\neg E$ world to w. We can assess "E was harmful to S" relative to each such precisification. It will be true on a precisification iff the nearest world to actuality on that precisification is better for S than actuality. We can say that it is determinately true iff all of the nearest $\neg E$ worlds are better for S than the actual world (i.e. it is true on all precisifications). We can say that it is determinately false if none of the $\neg E$ worlds are better for S than the actual world (i.e. it is false on all precisifications). And we can say it is indeterminate otherwise. Moreover, we can calculate its degree of truth by dividing the total number of nearby $\neg E$ worlds at which S is better off than at the actual world by the total number of nearby $\neg E$ worlds. So, if there are 100 nearby $\neg E$ worlds, and S is better off at 50 of them, "E was harmful to S" has a degree of truth of 0.5.[17,18]

Most of the events we typically think of as paradigms of harm will be very close to being determinately harmful. And surely if E is very close to being determinately harmful we should simply act as if it is harmful.[19] The problem is

[17] In reality it is not quite so straightforward as there will always be an infinite number of nearby $\neg E$ worlds. We will ignore this complication in what follows.

[18] See Lewis (1970), Edgington (1997), Kamp (1975), Cook (2002), and Williams (2011) for discussion of supervaluational degrees of truth.

[19] This is not quite what our best theories of decision making under indeterminacy tell us. For example, Williams (2014) argues that we should randomly select a precisification, and perform the action with the highest expected utility conditional on that precisification. This

that this simply generates a new instance of the *Asymmetry Problem*. Consider Buster's botched suicide again. The following sentence has a high degree of truth: "It was overall good for Buster that he was saved". After all, the vast majority of nearby worlds at which Buster is not saved are worse for Buster than the actual world. So, it is true on the vast majority of precisifications that it was good for him that he was saved. Thus, we should act as if it was good for Buster that he was saved. Now consider the counterfactual: "Had Buster not been saved, this would have been overall good for him". To assess this sentence we need to assess "had Buster been saved this would have been good for him" relative to the nearest world(s) to actuality at which he is not saved. We can do this since our precisifications will fix not only the nearest world to actuality at which Buster is not saved (w_1), but also the nearest world to this world at which he is saved (w_2)[20]. On the vast majority of precisifications "had Buster been saved, this would have been good for him" will be false at w_1 (that is, on the vast majority of precisifications the world assigned as w_2 will be worse for Buster than the world assigned as w_1). So, on the vast majority of precisifications "Had Buster not been saved, this would have been good for him" is also true. So, it has a very high degree of truth, and we should act as if it is true. A similar problem arises for degrees of benefit. In our **Overall Value** case both "E benefits S more than $\neg E$ would have benefited S", and "$\neg E$ would have benefited S more than E actually benefited S" both have a high degree of truth, and we should act as if they are both true. Thus, the *Asymmetry Problem* arises again.

So, what exactly is wrong with asymmetry? Do the problems go beyond weird intuitions about counterfactual harms and benefits in marginal cases? We think they do. Our original Buster case was rather extreme: Buster's life after being saved took a massive turn for the better, and this was due to a highly unusual quantum event (the sort of event that might never occur during the entirety of human history). But far less extreme versions of the case can be given. A-typical or "remarkable" events (in the Lewis/Williams sense) do occur

predicts that, if p has a high degree of truth, we will almost always rationally act as if p is true. However, it allows that if a precisification on which p is false is randomly selected it will be permissible to act as if $\neg p$. This, together with the indeterminacy approach to harm, entails that it will sometimes be rational to act as if a subject's being tortured to death is not harmful to them. This is hard to swallow. It may be possible to avoid this result by combining Williams's approach with a threshold view whereby if p's degree of truth surpasses some threshold it is always rational to act as if p.

[20] A single precisification will assign a determinate value to every expression in a language. So, it will assign a determinate value both to "E harmed S" relative to a world, and to "had E not occurred, this would have harmed S" relative to a world.

in actuality.[21] And these events can have major effects on people's lives. For example, in December 2020 the lottery numbers "5, 6, 7, 8, 9, 10" were drawn in South Africa's national lottery. This highly improbable and a-typical event was massively beneficial to the 20 people with these tickets.

Suppose that instead of being given billions by Bezos, Buster buys one of these tickets shortly after being saved. Suppose furthermore that he doesn't win a huge amount. He wins enough to pay for some medical treatments, to get some therapy, and to pay off some of his debts. After his win his life goes from being far worse than death to just about worth living. Our asymmetry generating versions of CCA still entail that Buster benefits from being saved (the actual world is better for him than the (vast majority of) the nearest world(s) at which he is not saved), and they still entail that it would have been good for him had he not been saved: if a-typicality or remarkableness makes for modal distance, then the nearest "saved" worlds to the nearest "not saved" worlds to actuality will not be worlds in which his "5, 6, 7, 8, 9, 10" ticket is drawn. After all, this event is remarkable/a-typical. MOST WORLDS CCA doesn't avoid the problem as Buster will lose in the majority of the nearest "saved" worlds to the nearest "not saved" worlds. And BRUTALIST CCA doesn't avoid the problem either as the probability of the nearest "saved" world to the nearest "not saved" world being one at which he wins will be vanishingly small. Moreover, in this version of the case the difference in intrinsic value between actuality and the nearest worlds at which Buster is not saved is pretty small. However, the difference in value between the nearest "not saved worlds" and *their* nearest "saved" worlds is very high. So, it is hard to avoid the conclusion that Buster would have benefited a lot more from not being saved than he did from being saved.

Not only is this is highly counterintuitive, it also raises difficult questions about practical reasoning: judgements of the form "ϕing would be more beneficial than ψing" play important roles in our practical reasoning. If we judge that ϕing would be more beneficial than ψing then, surely, it is rational for us to ϕ. However, in light of the *Asymmetry Problem*, it is not clear that this is the case. After all, it could be that ϕing would be more beneficial than ψing even when ψing results in us being better off than we would have been had we ϕ'd. So, if we ϕ whenever our options are ϕ and ψ and we know that "ϕing would

[21] Indeed, the fact that a-typical or "remarkable" events do regularly occur in actuality is one of the major reasons to be skeptical of remarkableness/typicality based approaches to counterfactuals (Hawthorne (2005))

be more beneficial than ψing" we will sometimes end up performing the action that yields the least optimal outcome.

Another reason not to simply bite the bullet and accept the predictions of our modified versions of CCA is that doing so will deliver unpalatable asymmetries when we compare deterministic and non-deterministic worlds. Consider the following case:

> **Indeterministic Flight** Alfi needs to get to New York as soon as possible. The quicker he gets there, the greater the rewards. If it takes him more than ten hours he will be punished. There is an airplane he can catch that will get him to New York in two hours. Alternatively, he can drive. This will take him fifteen hours. Alfi makes his flight. However, due to some low probability quantum event, the plane crashes and Alfi dies. Alfi lives in a world where the physical laws are non-determinstc.

If we accept any of our modified versions of CCA then we have to accept:

> **harm**Alfi would have been harmed if he had missed the flight (or, he would probably have been harmed if he had missed his flight).

Some readers might not find **Harm** highly counterintuitive in and of itself. However, compare **Indeterministic Flight** to the following case:

> **Deterministic Flight** Alfi needs to get to New York as soon as possible. The quicker he gets there, the greater the rewards. If it takes him more than ten hours he will be punished. There is an airplane he can catch that will get him to New York in two hours. Alternatively, he can drive. This will take him fifteen hours. Alfi makes his flight. However, due a pilot error, the plane crashes. This error was inevitable as the pilot had been drinking heavily. Moreover, the world Alfi lives in is deterministic.

This case seems to parallel **Indeterministic Flight**. Whether the crash was caused by a pilot error or a weird quantum event should not affect whether or not Alfi was harmed or benefited, nor whether he would have been harmed or benefited had he failed to catch his flight. Nor should it matter whether he lives in a deterministic or a non-deterministic world. (It shouldn't matter in this way to our actual attributions of harm how the scientific dispute on determinism is finally resolved). Yet our modified versions of CCA imply the following:

> **Benefi** Alfi would have benefited if he had missed the flight.

So, if our modified versions of CCA are correct then we have to treat **Deterministic Flight** and **Indeterministic Flight** very differently. We need to hold that whether or not Alfi would have benefited from missing the flight that kills him depends on the reason for the crash.

Finally, if one of our modified versions of CCA is accepted, we have to give up on either DEGREE OF BENEFIT or COMPARATIVE BENEFIT, or else accept that it can be the case that A) an event E benefits a subject S more than $\neg E$ would have benefited S, and B) $\neg E$ would have benefited S more than E actually benefited S. This is close to contradictory. Yet both DEGREE OF BENEFIT and COMPARATIVE BENEFIT seem beyond reproach. So, biting the bullet on the *Asymmetry Problem* is not an appealing prospect.

So, this leaves us with option 3: reject CCA. We think this is the most plausible option. Firstly, we have considered three separate ways of modifying the semantics for counterfactuals and the *Asymmetry Problem* arises on all of them. So, it is natural to suspect the culprit is CCA. After all, CCA was the common element in each case. It is not clear how CCA can be saved from the *Indeterminacy* or *Asymmetry* problems without more extreme changes to the theory of counterfactuals. This would likely involve giving up the possible worlds framework altogether, meaning that CCA in its current form would not survive. Finally, as we noted in the previous section, CCA is already controversial and faces a number of other problems (such as the apparent inability to separate harms from mere failures to benefit).

However, whilst rejecting CCA is our preferred option, it is worth emphasizing that it comes at a heavy cost: CCA seems to be the only account of harm capable of explaining the harm of death. Hence Feit (2015):

> "The biggest advantage of the counterfactual comparative account is its ability to handle the harm of death in particular, and deprivational or preventive harms more generally."

and Bradley (2012):

> "When we focus on overall extrinsic harm, we are inevitably led to a comparative account, since no non-comparative account offers a way to account for preventive or deprivational harms such as the harm of death."[22]

[22]Immediately emphasizing that by "comparative account" he means CCA, given how poorly other comparative strategies score in this respect.

The basic idea here is that the extent to which prevention and deprivation are harms can only be understood by comparison. And among comparative strategies, CCA seems most promising. For example: if something prevents me from applying to my dream job on time (and I was expected to be appointed), there may be no dramatic change in the mundane way my life goes. We still consider it to be a grave harm, but how can we account for this without comparing what actually happens to what would have happened had I not been prevented from applying? Some form of CCA seems to be called for.[23]

Of course, viable alternatives to CCA may appear in the future. Developing an alternative that preserves the advantages of CCA while steering away from the problems discussed above is much more desirable than biting the bullet. However, it is hard to see how deprivational harms can be accounted for without modal terminology, and hard to speculate how an account spelled out in modal terms will overcome the problems in question. For the time being, the challenge for CCA is a challenge for value theorists in general.

Acknowledgements

We would like to thank Dan Baras, Ran Lanzet, Jens Johansson, Aaron Segal, Robbie Williams, and audience at the New Israeli Philosophy Association conference. The authors were equal contributors to this paper. This research was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 818633)

[23] A different comparative account is temporally-based, crudely: an event is harmful in case, following it, I am worse off. But as the application example shows, this view is inadequate: despite the intuitive harm, my level of well-being is not affected. Noncomparative accounts do not fair much better. Causation based accounts, according to which an event is a harm to the extent that it causes bad states, are unable to account for the harm of death. After all, dying brings about no intrinsically bad states for the subject. According to another non-comparative account, Hanser's (2008) event based account, the harm of an event does not have to do with states, but rather with the loss of basic goods. Death is bad for the one who dies, according to Hanser, since by dying we lose all of our basic goods (for instance, our powers). Nevertheless, as stressed by Purves (2014, p. 96): "death is harmful, not just because [it takes away] the goods we had prior to death, but because it deprives us of the future goods we would have enjoyed had our lives not been cut short." Moreover, it seems hard to explain why losing a power is bad without appealing to the prevention of intrinsically good states (see Bradley (2012) for this objection and further worries).

References

[1] Bassi, A. 2020. Collapse Theories. *Stanford Encyclopedia of Philosophy*https://plato.stanford.edu/entries/qm-collapse/

[2] Bennett, J. 2003. *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.

[3] Bradley, B. 2009. *Well-being and death*. Oxford: Oxford University Press.

[4] Bradley, B. 2012. Doing away with harm. *Philosophy and Phenomenological Research,* 85(2), 390–412.

[5] Broome, J. 1999. *Ethics out of Economics*. Cambridge, U.K.: Cambridge University Press.

[6] Broome, J. 2004. *Weighing Lives*. New York: Oxford University Press.

[7] Cook, R. 2002. Vagueness and Mathematical Precision. *Mind* 111. 225-248.

[8] Dodd, D. 2011. Quasi-Miracles, typicality, and counterfactuals. *Synthese* 179. 351-360.

[9] Dorr, C. 2016. Against Counterfactual Miracles. *Philosophical Review* 125 (2). 241-286.

[10] Edington, D. 1997. Vagueness by Degrees. In Keefe, R. & Smith, p. (eds.) *Vagueness: A Reader*. Cambridge MA: MIT Press. 294-316.

[11] Feit, N. 2002. The Time of Death's Misfortune. *Noûs* 36 (3). 359-383.

[12] Feit, N. 2015. Plural harm. *Philosophy and Phenomenological Research*, 90(2), 361–388.

[13] Feit, N. 2017. Harming by Failing to Benefit. *Ethical Theory and Moral Practice* 22. 809–823.

[14] Feldman, F. 1991. Some Puzzles About the Evil of Death. *The Philosophical Review*, 100: 205–27

[15] Feldman, F. 2004. *Pleasure and the Good Life*. Oxford: Oxford University Press.

[16] Finnis, J. 2011. *Natural Law and Natural Rights*. Oxford: Oxford University Press.

[17] Griffin, J. 1986. *Well-Being: Its Meaning.* Oxford: Clarendon.

[18] Gundersen, L. 2004. Outline of a New Semantics for Counterfactuals. *Pacific Philosophical Quarterly* 85 (1). 1-20.

[19] Hajek, A. MS. *Most Counterfactuals are False.*

[20] Hanna, N. 2016. Harm: Omission, preemption, freedom. *Philosophy and Phenomenological Research,* 91(2), 1–23.

[21] Hanser, M. 2008. The metaphysics of harm. *Philosophy and Phenomenological Research*, 77(2), 421–450.

[22] Harman, E. 2004. Can We Harm and Benefit in Creating? *Philosophical Perspectives.* 18 (1). 89-113.

[23] Hawthorne, J. 2005. Chance and Counterfactuals. *Philosophy and Phenomenological Research* 70 (2). 396-405.

[24] Ichikawa, J. 2008. Quantifiers, Knowledge, and Counterfactuals. *Philosophy and Phenomenological Research* 82 (2). 287-313.

[25] Kamp, J.A. 1975. Two Theories about Adjectives. In Keenan, E.L. (ed.) *Formal Semantics of Natural Language.* Cambridge: Cambridge University Press. 123-155

[26] Klocksiem, J. 2012. A defense of the counterfactual comparative analysis of harm. *American Philosophical Quarterly*, 49(4), 285–300.

[27] Leitgeb, H. 2012a. A Probabilistic Semantics for Counterfactuals, Part A. *Review of Symbolic Logic*, 5 (1). 26-84.

[28] Leitgeb, H. 2012b. A Probabilistic Semantics for Counterfactuals, Part B. *Review of Symbolic Logic*, 5 (1). 85-121.

[29] Lewis, D. 1970. How to Define Theoretical Terms. *The Journal of Philosophy* 67. 427-446.

[30] Lewis, D. 1973. *Counterfactuals.* Blackwell.

[31] Lewis, D. 1986. Postscript to Counterfactual Dependence and Times Arrow. *In Philosophical Papers, Volume 2.* Oxford: Oxford University Press. 32-66.

[32] Lewis, K. 2016. Elusive Counterfactuals. *Noûs* 50 (2). 286-313.

[33] Moss, S. 2013. Subjunctive Credences and Semantic Humility. *Philosophy and Phenomenological Rsearch* 87 (2). 251-278.

[34] Purves, D. 2014. A Counterexample to Two Accounts of Harm. *Southwest Philosophy Review* 30 (1). 243-250.

[35] Purves, D. 2019. Harming as Making Worse Off. *Philosophical Studies* 176 (10):2629-2656

[36] Sandgren, A. & Steele, K. Forthcoming. Levelling Counterfactual Skepticism. *Synthese* 1-21.

[37] Schultz, M. 2014. Counterfactuals and Arbitrariness. *Mind* 123 (429). 1021-1055.

[38] Shiffrin, S. V. 2012. Harm and its moral significance. *Legal Theory,* 18(3), 357–398.

[39] Smith, M. 2016. *Between Probability and Certainty: What Justifies Belief.* Oxford: Oxford University Press.

[40] Stalnaker, R. 1968. A Theory of Conditionals. In N. Rescher (ed.), *Studies in Logical Theory.* Oxford University Press.

[41] Stefansson, O. 2018. Counterfactual Scepticism and Multidimensional Semantics. *Erkenntnis* 83. 875-898.

[42] Williams, R. 2008. Chances, Counterfactuals, and Similarity. *Philosophy and Phenomenological Research*, 77. 385-420.

[43] Williams, R. 2011. Degree Supervaluational Logic. *Review of Symbolic Logic* 4 (1). 130-149.

[44] Williams, R. 2014. Decision-Making Under Indeterminacy. *Philosopher's Imprint* 14 (4). 1-34.