

Counterfactuals, Indeterminacy, and Value: A Puzzle

Andrew Peet & Eli Pitcovski

March 26, 2021

Abstract

According to the Counterfactual Comparative Account of harm and benefit, 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 the Counterfactual Comparative Account (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 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 the Counterfactual Comparative Account in this respect, theorists of harm and benefit face a deadlock.

1 Introduction

What makes an event overall harmful or beneficial for a subject? It is commonly thought that whether or not an event E altogether harms (or a benefits) a subject S is a function of whether or not S would have been better off (or worse off, respectively) had E not occurred. So, in order to evaluate whether

or not an event harms S we have to assess a counterfactual of the form “had E not occurred S would have been better off”. This approach to harm and benefit (better articulated below) is the Counterfactual Comparative Account (henceforth CCA)¹

In this paper we present a trilemma. The options are as follows:

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. Accept that an event can be overall harmful (/beneficial) for a subject, consistently with it’s also being the case that, had this event failed to take place, this would also have been overall harmful (/beneficial) for the subject.

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 would have happened then C would have occurred”, (or, $A \Box \rightarrow C$), 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)).^{2,3}

Here is the problem: 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 only probabilities of locations. This has some surprising consequences: Suppose I let go of a ball. There is, it turns out, a small chance that the ball would remain suspended in the air. Now suppose

¹CCA 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.

²Lewis’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 3.

³If 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)).

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. If physical processes are chancy in the way suggested above, some small proportion of the nearest $\neg E$ worlds will contain highly improbable events (akin to the ball remaining suspended in the air) that radically impact the quality of S 's life (for better or worse). So, it will not straightforwardly be the case that E either harms or benefits S . The most natural way to go here (especially in light of standard formulations of CCA) is to conclude that it is indeterminate whether or not, and to what extent E harms S .⁴ That an event's degree of harmfulness is typically somewhat indeterminate is not especially counterintuitive. The problem is that i) it will rarely determinately be the case that *any* event is harmful (so, even cases of extreme torture will not be judged determinately harmful for their victims), and ii) the harm or benefit of any event will be indeterminate between a huge range of values: from terribly harmful to highly beneficial (so, it will rarely be determinately false that cases of agonizing torture are overall hugely beneficial for their victims). Call this the *Indeterminacy Problem*.

The only way to avoid this problem is to eliminate from consideration worlds at which these a-typical events occur. We outline several ways of doing so and argue that they all succumb to the same problem: suppose that the actual world contains an event E that benefits S . And suppose that S would not have been even better off had E not occurred.⁵ Then, intuitively, we would judge that if E hadn't occurred this would have been overall bad for S (after all, S would

⁴The alternative is to conclude that it is false that E harms S , and false that E benefits S . This would amount to the practical elimination of harm or benefit from the world, not only the elimination of determinate harm or benefit.

⁵Strictly speaking this qualification is not necessary since we are talking about overall

have been deprived of some benefit).⁶ However, if the actual world in which E obtains is itself a-typical we will not always get this result. Rather, CCA will sometimes tell us that, had E not occurred, this would have been overall good for S - potentially the overall benefit for S might be greater than the overall benefit of E in the actual world. This has an air of paradox: E is good for S , S would not have been better off had E not occurred, yet had E not occurred this would have benefited S more than E actually benefited S .

How do we get this result? Well, any version of CCA capable of avoiding the indeterminacy problem will do so by analyzing the harm of an event E at a world w by comparing w to w 's nearest $\neg E$ worlds (or a subset of these worlds), and holding that the relevant set of worlds does not include any at which problematic a-typical events occur. But then in assessing the counterfactual "had E not occurred this would have harmed S " at w we compare these same worlds to *their* nearest E worlds (where this doesn't include a-typical worlds). If w is itself an a-typical world (in a manner relevant to the value of E for S at w) then w will not enter this comparison. It may be the case that whilst w is better for S than w 's nearest $\neg E$ worlds, the nearest E worlds to all these $\neg E$ worlds are far worse for S than the $\neg E$ worlds. Call this the *Asymmetry Problem*.

The paper proceeds as follows: in section 2 we outline CCA and introduce the *Indeterminacy Problem* in more detail. In section 3 we consider responses to the *Indeterminacy Problem* and argue that they all succumb to the *Asymmetry Problem*. In section 4 we consider our options. Rejecting CCA means giving up on seemingly the only account of harm capable of dealing with deprivational harms such as the harm of death. 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 as it poses a burden for practical reasoning. We consider 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.

rather than pro-tanto harm and benefit. It is clear that an event can pro-tanto benefit a subject even if they would have been even better off had it not occurred. But it seems obviously false that an event could overall benefit a subject even though they would have been overall better off had it not occurred. We leave the qualification in place in order to drive home just how counterintuitive this result is.

⁶In the context of this paper we will use 'bad' and 'harmful' interchangeably. Likewise for 'good' and 'beneficial'.

2 Overall Harm: Sources of Indeterminacy

Some events are *intrinsically* good for a subject, or good in and of themselves (like quality time with a friend), other events are intrinsically bad (like a painful accident). Some events are *extrinsically* good for a subject, or good because of something else, often something they prevent or bring about (a shout from the back seat preventing an accident), other events are extrinsically bad (being stuck in traffic and thereby missing quality time with a friend). 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 prevalent approach to overall harm/benefit is the Counterfactual Comparative Account (CCA),⁷ according to which an event E is overall harmful (/beneficial) for a subject S to the extent that S is worse (/better) off than they would be had E not occurred (we will henceforth remove the “beneficial/benefit” qualification and, when not specifically addressing benefit, focus on harm). On a rough first pass formulation,⁸ 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 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.⁹ (For an account of overall benefit, replace ‘harm’ with ‘benefit’ and ‘worse’ with ‘better’).

⁷Among others, adherents include: Feldman (1991), Broome (1999, 2004), Bradley (2009), Feit (2015), Purves (2014, 2019), Klocksiem (2012)

⁸As formulated, CCA is vulnerable to a number of objections. Without going into details: it is often objected that it fails to account for cases of preemption and cases of causal overdetermination (e.g. Bradley (2012)) and that it fails to distinguish harm from failure to benefit (See: Hanser (2008), Shiffrin (2012)). Advocates of the counterfactual comparative approach have made progress both with respect to distinguishing cases of harm from cases of failure to benefit (Purves (2019), Klocksiem (2012)) and with respect to delivering intuitive results in cases of preemption and overdetermination (Feit (2015), Bradley (2004, p.18-21; 2009, 2.2), Hanna (2015)). If we wish to take their insights or future insights into account, a fully worked-out counterfactual comparative criterion will include significant modifications (both qualifications and extensions) of CCA. However, for the sake of presenting the problem below, it will be easiest to focus on CCA.

⁹There are various slightly different formulations of the view, some of which do not account for degrees of harm, and some of which account for overall benefit on top of accounting for overall harm. 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).

Worse and better off are cashed out in terms of the aggregated intrinsic value of the totality of relevant states and their degrees of intrinsic badness/goodness at each relevant world. So, for instance, (using 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. This paper, however, will focus on accounts of overall harm.¹⁰

As previously noted, 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.¹¹

¹⁰Intrinsic 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 takes this axiological neutrality to be a desideratum for an adequate theory of overall harm.

¹¹It might be objected that since determinism is true there will be some determinate fact of the matter concerning whether the coin would have landed heads or tails. If this is correct then it is either determinately true or determinately false that Ruth's death harms her. However, 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, as previously noted, 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

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 perfectly acceptable 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.

At first glance, it appears this indeterminacy conclusion can be readily embraced. Consider a paradigmatically harmful event: death at the age of 20. Typically, one may think, it will still be the case that death at the age of 20 is determinately harmful to the one who dies. After all, we might assume, it will usually be the case that had a subject not died at 20 they would have lived a life worth living. The precise value of their un-lived life may vary from world to world. And so the precise harm of their death will be indeterminate. However, as long as the life they live in each nearby world is worth living, their death will be determinately harmful. Indeed, it will be determinately harmful at least to the degree to which the actual world is worse for the subject than the worst nearby world at which they live. For example, suppose that our subject S dies at the age of 20, and that there are three equally nearby worlds at which they live. Suppose the values of the relevant portion of S 's life at each of these worlds are as follows: $w_1=2$, $w_2=4$, $w_3=6$. If this is the case then it is determinately the case that S 's death has a disvalue of at least -2, and a disvalue between -2 and -6.

Assuming determinism, this response seems promising. However, things start to look a lot worse for CCA when we take into account the kind of physical

Ruth not died, her survival would be due to one particular divergence rather than a some other.

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 atypical 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 such worlds "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. Supposing that S 's life up to the time of death has a value X , 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. Whenever a subject dies there will be, according to the standard modal closeness relation, some small set of nearby worlds at which they live a life not worth living due to some bizarre low probability quantum event. So, death will rarely if ever be determinately harmful for a subject. CCA's greatest advantage (the ability to account for the harm of death) is in jeopardy. Yet the problem goes deeper, as it does not only pertain to the harm of death: 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*.¹²

3 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 to the extent that things would have been better for S had E not occurred. That is, E is harmful for S iff the nearest worlds at which E does not occur are better for E than the actual world. The nearest worlds at which an event E occurs will be those that match the actual world until (shortly before) the occurrence of E , and which match the actual world in physical law (but not necessarily in matters of particular fact) after the occurrence of E . 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 occurs. 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 the introduction, 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 offer 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: if an event E is overall beneficial for a subject, then it should be the case that its failure to occur would have overall harmed the subject. After all, had E not occurred our subject would have been deprived of some benefit.^{13,14} Yet, if we modify CCA in such a way that the indeterminacy problem is avoided we get violations of this principle.

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

¹²It is worth noting that, strictly speaking, the *Indeterminacy Problem* doesn't require that the world actually is chancy in the manner described above. Regardless of whether or not the actual world is chancy in this way, presumably determinately harmful events *could* occur even if the world was chancy in the manner described here. This seems to be ruled out by the considerations aduced above.

¹³Indeed, we would naturally expect the absolute value of an event E ($|V_{E_S}|$) to be equal to the absolute value of its counterfactual failure to occur ($|V_{\neg E_S}|$).

¹⁴Recall that if an event is overall beneficial to a subject (rather than just pro-tanto beneficial) it cannot be the case that the subject would have been even better off had the event not occurred.

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 $A \square \rightarrow C$ 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 $A \square \rightarrow C$ 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 which 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 the “modified nearness” approach.^{15,16}

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

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 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.

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

¹⁵Although 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

¹⁶There is one species of approach we have not mentioned here: contextualism (e.g. Ichikawa (2008) & Lewis (2016)). Contextualists hold that a counterfactual $A \square \rightarrow C$ 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. Something like typicality. 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 trying to find the closeness ordering relation at play in contexts where we are discussing harm). The problems we raise for the above approaches thus carry over straightforwardly to contextualism.

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.

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 can be formulated in two ways. Firstly, the MODIFIED NEARNESS theorist can 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 avoids our problem in much the same way as the most worlds approach. Alternatively, the MODIFIED NEARNESS theorist can stick with the original formulation of CCA repeated above (i.e. in terms of the nearest $\neg E$ world rather than the nearest $\neg E$ 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. However, in most cases there will no longer be any nearby worlds at which they live a life worse than death.

So, we have three responses to our problem. Each response has problems of its own. However, we want to focus on a problem faced by all three responses. Indeed, it is a problem that seems to be faced by any attempt to save CCA from the problem of indeterminacy. Consider the following principle:

Counterfactual Symmetry If an event E is overall beneficial for a subject S (S would have been worse off had E failed to occur), then had E failed to occur this would have harmed S (swapping “beneficial” with “harmful” and “worse” with “better” yields the equivalent principle for counterfactual benefit).

COUNTERFACTUAL SYMMETRY seems obviously true. Hence Broome (2004): “When I asked what your life is worth to you, I meant, more precisely: how good is it for you to continue living? Conversely, what harm would be done to you by not continuing to live? What would you lose by dying?”. However,

if any of the above modifications of CCA are correct then COUNTERFACTUAL SYMMETRY is false. 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 very good for Buster that he was saved. Things went extremely well for him in the actual world. Would it have been overall bad for him not to have been saved? Intuitively, yes. This is also what COUNTERFACTUAL SYMMETRY tells us. 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 benefitted from not being saved. This is the *Asymmetry Problem*. Let us demonstrate the inevitability of this problem for modified versions of CCA by considering each approach in turn.

We are evaluating the harm that would have resulted from Buster's not being saved. 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 worse 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 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.¹⁷

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

¹⁷To put this another way, suppose that the nearest worlds to actuality at which Buster is not saved are $w_{\alpha 1}, w_{\alpha 2} \dots w_{\alpha n}$. Suppose that the nearest worlds to $w_{\alpha 1}$ at which he is saved are $w_{1_1}, w_{1_2} \dots w_{1_n}$, and that the nearest worlds to $w_{\alpha 2}$ at which he is saved are $w_{2_1}, w_{2_2} \dots w_{2_n}$ etc. In order to judge whether Buster's death would have benefited him we need to compare $w_{\alpha 1}$ to $w_{1_1}, w_{1_2} \dots w_{1_n}$, and $w_{\alpha 2}$ to $w_{2_1}, w_{2_2} \dots w_{2_n}$ etc. According to MOST WORLDS CCA Buster's 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 ($w_{\alpha x_1}, w_{\alpha x_2} \dots w_{\alpha x_n}$). 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.

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.¹⁸

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). That is, any modification of CCA capable of dealing with the indeterminacy problem will do so by eliminating quantum weirdness worlds from consideration. In doing so they open themselves up to the value asymmetry problem. This is because any such approach will analyze the harm of an event E at a world w by comparing w to w ’s nearest $\neg E$ worlds (or a subset of these worlds), and holding that the relevant set of worlds does not include quantum weirdness worlds. But then in assessing the counterfactual “had E not occurred, this would have harmed S ” at w we compare these same worlds to *their* nearest E worlds (where this doesn’t include quantum weirdness worlds). If w is itself a quantum weirdness world (in a manner relevant to the value of E for S at w) then there will be a value asymmetry. The only way to avoid this would be to deny that we can derive the truth conditions of the counterfactual “if E had occurred it would have been bad for S ” from our theory of counterfactuals together with our theory of harm. This should be avoided at all costs.¹⁹

¹⁸The idea here is very similar to what we saw with MOST WORLDS CCA: Suppose that the nearest worlds to actuality at which Buster is not saved are $w_{\alpha 1}, w_{\alpha 2} \dots w_{\alpha n}$. Suppose that the nearest worlds to $w_{\alpha 1}$ at which he is saved are $w_{11}, w_{12} \dots w_{1n}$, and that the nearest worlds to $w_{\alpha 2}$ at which he is saved are $w_{21}, w_{22} \dots w_{2n}$ etc. In order to judge whether Buster’s death would have benefited him we need to compare $w_{\alpha 1}$ to $w_{11}, w_{12} \dots w_{1n}$, and $w_{\alpha 2}$ to $w_{21}, w_{22} \dots w_{2n}$ etc. 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 ($w_{\alpha x 1}, w_{\alpha x 2} \dots w_{\alpha x n}$). 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.

¹⁹It is worth noting that the asymmetry problem arises for some comparative accounts other

4 Evaluating the Options

At this point we have three options:

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

We will not take a stand here on which option should be taken. But we will close by briefly highlighting the costs of each option. Option one is unattractive because 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.”²⁰

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

than CCA. Suppose we say that E is harmful to S if the value of the actual world is lower for S than the average value of the nearest $\neg E$ worlds. This is similar to holding that E is harmful to S if the expected utility of E is low for S (Hare (Forthcoming) considers a view along these lines when discussing a problem similar to the *Indeterminacy Problem* for consequentialism). This view will also face the *Asymmetry Problem*: The average value of the worlds nearest to actuality at which Buster dies is lower than that of the actual world. So, his being saved benefits him. Yet, at the same time, the value of each such world is higher than the average value of each of *their* nearest “saved” worlds. So, it is also true that, had Buster died, this would have benefited him.

²⁰immediately emphasizing that by “comparative account” he means CCA, given how poorly other comparative strategies score in this respect.

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.²¹

Of course, viable alternatives to CCA may appear in the future. But, it is hard to see how such an account could be given without appeal to modal notions. And any account formulated by explicit or implicit appeal to modal notions will have to carefully steer clear of the kinds of worries we have raised here. So, for the time being, it seems that unless we follow Epicurus and deny the harm of death altogether, a challenge for CCA is a challenge for harm theorists in general.

Perhaps, then, we should 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

²¹In detail, let us consider two sorts of noncomparative accounts. 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. Another non-comparative account, Hanser’s (2008) event based account, fares slightly better. According to Hanser, 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). Accounting for deprivational harm seems to require some form of comparison. CCA is not the only comparative account. There is also the temporal comparative account, according to which an event harms a subject iff due to this event that subject is worse off (i.e. has a lower degree of well-being) after the event that they were before it. This view also struggles with the harm of death: if one fails to have *any* degree of well-being once one ceases to exist (see Johansson (2012, p.264-266) and Luper (2012, p.320-321) for discussion), the comparison cannot go through. Moreover, the temporal comparison view is also extensionally inadequate: it delivers that death during a short-term extremely limiting and unpleasant disease is better than death in full health at seventy. For further criticism of this view, see Norcross (2005, p.149-50) and Holtug (2002, p.368).

$\neg E$ 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 than actuality (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.^{22,23}

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.²⁴ The problem is 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

²²In 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.

²³See Lewis (1970), Edgington (1997), Kamp (1975), Cook (2002), and Williams (2011) for discussion of supervolational degrees of truth.

²⁴This is not quite what our best theories of decision making under indeterminacy tells 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 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 .

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)²⁵. 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. This is simply the *Asymmetry Problem* again.

So, what exactly is wrong with asymmetry? It is highly counter intuitive, but it seems we are forced in this direction unless we give up on CCA. Do the problems go beyond weird intuitions about counterfactual harms 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 in actuality.²⁶ 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”

²⁵A 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.

²⁶Indeed, 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))

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 highly counterintuitive, it also raises difficult questions about practical reasoning: judgements of the form “ ϕ ing would be more beneficial than ψ ing” play important roles in our practical reasoning. If we judge that ϕ ing would be more beneficial than ψ ing 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 ϕ ing would be more beneficial than ψ ing even when ψ ing 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 “ ϕ ing would be more beneficial than ψ ing” we will sometimes end up performing the action that yields the least optimal outcome.

5 Concluding Remarks

So, we have a predicament. We must either reject CCA, accept indeterminacy, or accept asymmetry. Each option is unattractive. The rejection of CCA is unattractive because CCA is seemingly the only approach to value capable of dealing with deprevaluational harms such as the harm of death. Rejecting CCA would either involve denying that death is harmful to the one who dies, or producing a new approach of value capable of dealing with deprivational harms without giving rise to problematic indeterminacy or value asymmetries. Indeterminacy is both counterintuitive and, in its least extreme form (allowing for degrees of determinacy) leads back to a version of the asymmetry problem. And asymmetry is not easily accepted since it A) is not limited to marginal unrealistic cases, and B) raises difficult questions about the role of counterfactual harm

and benefit judgements in practical reasoning.

References

- Bennett, J. 2003. *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.
- Bradley, B. 2004. When is death bad for the one who dies? *Noûs*, 38(1), 1–28.
- Bradley, B. 2009. *Well-being and death*. Oxford: Oxford University Press.
- Bradley, B. 2012. Doing away with harm. *Philosophy and Phenomenological Research*, 85(2), 390–412.
- Broome, J. 1999. *Ethics out of Economics*. Cambridge, U.K.: Cambridge University Press.
- Broome, J. 2004. *Weighing Lives*. New York: Oxford University Press.
- Cook, R. 2002. Vagueness and Mathematical Precision. *Mind* 111. 225-248.
- Dodd, D. 2011. Quasi-Miracles, typicality, and counterfactuals. *Synthese* 179. 351-360.
- Dorr, C. 2016. Against Counterfactual Miracles. *Philosophical Review* 125 (2). 241-286.
- Edington, D. 1997. Vagueness by Degrees. In Keefe, R. & Smith, p. (eds.) *Vagueness: A Reader*. Cambridge MA: MIT Press. 294-316.
- Feit, N. 2002. The Time of Death's Misfortune. *Noûs* 36 (3). 359-383.
- Feit, N. 2015. Plural harm. *Philosophy and Phenomenological Research*, 90(2), 361–388.
- Feldman, F. 1991. Some Puzzles About the Evil of Death. *The Philosophical Review*, 100: 205–27
- Feldman, F. 2004. *Pleasure and the Good Life*. Oxford: Oxford University Press.
- Finnis, J. 2011. *Natural Law and Natural Rights*. Oxford: Oxford University Press.
- Griffin, J. 1986. *Well-Being: Its Meaning*. Oxford: Clarendon.
- Gundersen, L. 2004. Outline of a New Semantics for Counterfactuals. *Pacific Philosophical Quarterly* 85 (1). 1-20.
- Hajek, A. MS. *Most Counterfactuals are False*.

- Hanna, N. 2015. Harm: Omission, preemption, freedom. *Philosophy and Phenomenological Research*, 91(2), 1–23.
- Hanser, M. 2008. The metaphysics of harm. *Philosophy and Phenomenological Research*, 77(2), 421–450.
- Hare, C. Forthcoming. Consequentialism and Indeterminacy. In Portmore D. (ed.) *The Oxford Handbook of Consequentialism*. Oxford University Press.
- Harman, E. 2004. Can We Harm and Benefit in Creating? *Philosophical Perspectives*. 18 (1). 89–113.
- Hawthorne, J. 2005. Chance and Counterfactuals. *Philosophy and Phenomenological Research* 70 (2). 396–405.
- Heathwood, C. 2014. Desire-Fulfillment Theory. In Eggleston, B. & Miller, D. E. (eds.), *The Cambridge Companion to Utilitarianism*. Cambridge: Cambridge University Press.
- Heathwood, C. 2019. Which Desires Are Relevant to Well-Being? *Nous* 53 (3). 664–668.
- Holtug, N. 2002. The Harm Principle. *Ethical Theory and Moral Practice* 5. 357–389.
- Ichikawa, J. 2008. Quantifiers, Knowledge, and Counterfactuals. *Philosophy and Phenomenological Research* 82 (2). 287–313.
- Johansson, J. 2012. The Timing Problem. In Bradley, B., Feldman, F. & Johansson, J. (eds.) *The Oxford Handbook of Philosophy of Death* New York: Oxford University Press. 225–273.
- Kamp, J.A. 1975. Two Theories about Adjectives. In Keenan, E.L. (ed.) *Formal Semantics of Natural Language*. Cambridge: Cambridge University Press. 123–155
- Klocksien, J. 2012. A defense of the counterfactual comparative analysis of harm. *American Philosophical Quarterly*, 49(4), 285–300.
- Leitgeb, H. 2012a. A Probabilistic Semantics for Counterfactuals, Part A. *Review of Symbolic Logic*, 5 (1). 26–84.
- Leitgeb, H. 2012b. A Probabilistic Semantics for Counterfactuals, Part B. *Review of Symbolic Logic*, 5 (1). 85–121.
- Lewis, D. 1970. How to Define Theoretical Terms. *The Journal of Philosophy* 67. 427–446.

- Lewis, D. 1973. *Counterfactuals*. Blackwell.
- Lewis, D. 1986. Postscript to Counterfactual Dependence and Times Arrow. In *Philosophical Papers, Volume 2*. Oxford: Oxford University Press. 32-66.
- Lewis, K. 2016. Elusive Counterfactuals. *Noûs* 50 (2). 286-313.
- Luper, S. 2012. Retroactive Harms and Wrongs. In Bradley, B., Feldman, F. & Johansson, J. (eds.) *The Oxford Handbook of Philosophy of Death* New York: Oxford University Press. 317-335.
- Moss, S. 2013. Subjunctive Credences and Semantic Humility. *Philosophy and Phenomenological Research* 87 (2). 251-278.
- Norocross, A. 2005. Harming in Context. *Philosophical Studies* 123. 149-173.
- Purves, D. 2014. A Counterexample to Two Accounts of Harm. *Southwest Philosophy Review* 30 (1). 243-250.
- Purves, D. 2016. Accounting for the Harm of Death. *Pacific Philosophical Quarterly* 97 (1). 89-112.
- Purves, D. 2019. Harming as Making Worse Off. *Philosophical Studies* 176 (10):2629-2656
- Schultz, M. 2014. Counterfactuals and Arbitrariness. *Mind* 123 (429). 1021-1055.
- Shiffrin, S. V. 2012. Harm and its moral significance. *Legal Theory*, 18(3), 357-398.
- Smith, M. 2016. *Between Probability and Certainty: What Justifies Belief*. Oxford: Oxford University Press.
- Stalnaker, R. 1968. A Theory of Conditionals. In N. Rescher (ed.), *Studies in Logical Theory*. Oxford University Press.
- Stefansson, O. 2018. Counterfactual Scepticism and Multidimensional Semantics. *Erkenntnis* 83. 875-898.
- Williams, R. 2008. Chances, Counterfactuals, and Similarity. *Philosophy and Phenomenological Research*, 77. 385-420.
- Williams, R. 2011. Degree Supervaluational Logic. *Review of Symbolic Logic* 4 (1). 130-149.
- Williams, R. 2014. Decision-Making Under Indeterminacy. *Philosopher's Imprint* 14 (4). 1-34.
- Woodward, C. 2013. Classifying Theories of Welfare. *Philosophical Studies*, 165, 787-803.