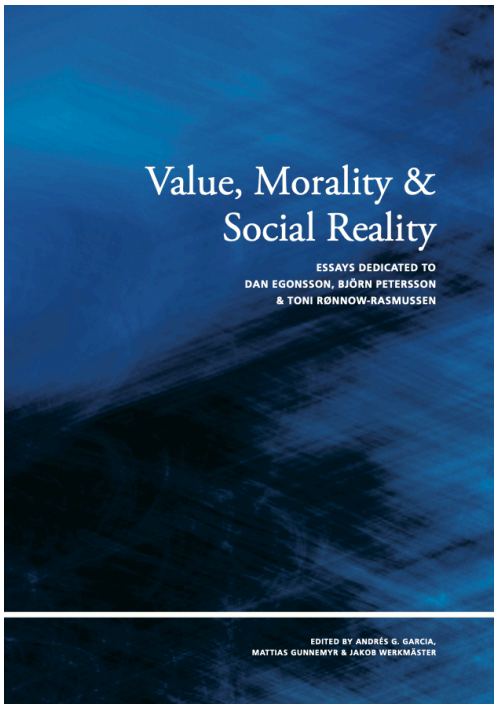
An Account of Instrumental Value

Erik Carlson

In: Garcia, A., Gunnemyr, M. & Werkmäster, J. (2023) *Value, Morality & Social Reality: Essays dedicated to Dan Egonsson, Björn Petersson & Toni Rønnow-Rasmussen.* Lund: Department of Philosophy, Lund University. DOI: https://doi.org/10.37852/oblu.189

ISBN: 978-91-89415-65-2 (print), 978-91-89415-66-9 (digital)



Published by the Department of Philosophy, Lund University.
Edited by: Andrés Garcia, Mattias Gunnemyr, and Jakob Werkmäster
Cover image by Fabian Jones. Cover layout by Gunilla Albertén.

DOI: https://doi.org/10.37852/oblu.189.c516

An Account of Instrumental Value

Erik Carlson[1]

Abstract. In this paper, I tentatively suggest an account of how the instrumental value of a state of affairs derives from the intrinsic value of other states. According to this account, a state's instrumental value depends on how its outcome compares to the outcomes of its best and its worst alternative. Further, I briefly discuss similar accounts of personal instrumental value, and of harm and benefit.

1. Introduction

Some things are good or bad for their own sake. Such things have *intrinsic* value. Other things are good or bad because they lead to or prevent something that has intrinsic value. These things have *instrumental* value. In this paper I shall tentatively suggest an account of how a thing's instrumental value derives from the intrinsic value of other things. [2]

To make this task somewhat more tractable, I will make a number of simplifying assumptions. I will assume that the bearers of value are contingent states of affairs,

[1] I dedicate this paper to Toni Rønnow-Rasmussen, although I am afraid it lacks much of the philosophical subtlety and sophistication characteristic of Toni's work in value theory.

[2] My usage of the terms 'intrinsic value' and 'instrumental value' is to some extent stipulative. Some authors prefer 'final value', to denote value for a thing's own sake, and 'instrumental value' is sometimes used in both broader and narrower senses than mine. Often, a main distinction is drawn between intrinsic and *extrinsic* value. Some philosophers equate extrinsic value with instrumental value, whereas others regard extrinsic value as a broader category. There are many suggestions about how to sharpen and elaborate on these and related distinctions. Rønnow-Rasmussen (2002, 2015) and Zimmerman & Bradley (2019) contain excellent discussions and overviews of the literature.

which may be either atomic or conjunctive, and that a possible world is a maximal consistent conjunctive state. As concerns instrumental value in particular, it is often natural to regard events, including actions, as value bearers. To accommodate this possibility, I shall view events as a species of states of affairs. Alternatively, one could assume that instrumental value is borne by the state of affairs that a certain event occurs, rather than by the event itself.

Further, I will assume that intrinsic value can be measured on a real-valued ratio scale, such that the value of an intrinsically good (bad) state of affairs is represented by a positive (negative) number and the value of an intrinsically neutral state by zero. The intrinsic value of a conjunctive state of affairs S is, I will suppose, the sum of the basic intrinsic values of its atomic or conjunctive parts, including S itself.[3] Finally, I will assume that for any contingent state of affairs, there is a possible world that would be actual if this state were to obtain, and a possible world that would be actual if it were not to obtain.[4] Some of these assumptions may not be very realistic, but they allow us to avoid a number of difficulties that are not directly relevant to the main issues.

2. The Simple Account

It might be suggested that the instrumental value of a state of affairs is simply the intrinsic value there would be in the universe if the state were to obtain. Thus, let us start by considering the following account, letting W_S denote the possible world that would be actual if state of affairs S were to obtain:

> *The Simple Account*. The instrumental value of a state of affairs S is equal to the intrinsic value of W_S minus the intrinsic value of S (which may be zero).

Although appealingly simple, the Simple Account will not do. It implies that all states with the same intrinsic value that obtain in a given possible world have the same instrumental value in that world. (Note that in a world where states S and S^* both obtain, W_S and W_{S^*} are identical.) Further, in an intrinsically good (bad) world all obtaining intrinsically neutral states are instrumentally good (bad). This is very implausible. Surely, states with equal intrinsic value may differ in instrumental value, and an intrinsically good or bad world may contain both instrumentally good and instrumentally bad states, which are intrinsically neutral.

These objections indicate that the connection between intrinsic and instrumental value posited by the Simple Account is too tenuous. The mere fact a certain

[3] Intuitively, a thing's basic intrinsic value is that part of its intrinsic value that does not derive from any of its proper parts. See, e.g., Feldman (2000) and Zimmerman (2001, chapter 5).

[4] The last assumption will be relaxed in section 6.

intrinsically good or bad state would obtain if a state S were to obtain is insufficient to confer positive or negative instrumental value on S. At the very least, what would be the case were S *not* to obtain also seems relevant as regards the instrumental value of S.

3. The Revised Simple Account

This suggests the following account:

> *The Revised Simple Account.* The instrumental value of a state of affairs S is equal to the intrinsic value of the conjunction of the states of affairs $S^* \neq S$, such that S^* would obtain if and only if S were to obtain.

By including the "only if" clause, this account avoids the most obvious flaws of the Simple Account. It allows that states with the same intrinsic value differ in instrumental value, and that intrinsically good (bad) worlds contain states that are intrinsically neutral and instrumentally bad (good).

However, the Revised Simple Account faces other serious problems. Suppose the world would have contained no intrinsically good or bad states of affairs had there not been life on Earth, and let S be the state that an asteroid hits the Earth early in its history, preventing life from ever evolving. Suppose also that the actual world is intrinsically very good. Intuitively, S is then instrumentally bad. According to the Revised Simple Account, however, it is instrumentally neutral.

Moreover, the Revised Simple Account also yields implausible results concerning the relative ranking of states of affairs in terms of instrumental value. Consider the following case:

> *Levers.* God offers you to pull one of three levers, labelled L_1 to L_3. You cannot refuse God's offer. Pulling a lever has no intrinsic value. If you pull L_i possible world W_i will be actual. W_1 and W_2 are very good worlds, containing many and exactly the same intrinsically good states of affairs, and no intrinsically bad ones. W_3 is not nearly as good, containing no intrinsically bad states, but only one intrinsically slightly good state. This state is not included in W_1 or W_2, and its intrinsic value is 1. Suppose also that you pull L_1, and that you would have pulled L_2, had you not pulled L_1.

In this case there is no obtaining intrinsically good or bad state that would not have obtained if you had not pulled L_1. Hence, the Revised Simple Account implies that the instrumental value of pulling L_1 is zero. The instrumental value of pulling L_3, on the other hand, is 1, since the only intrinsically good state in W_3 would obtain just in case you were to pull L_3. But the conclusion that pulling L_3 is instrumentally better than pulling L_1 is surely false.

4. The Counterfactual Comparative Account

The Revised Simple Account is insensitive to the fact that the asteroid's hitting the Earth, or your pulling L_3 in *Levers*, would *prevent* many intrinsically good states from obtaining. The remedy, it may be thought, is the following further revision:

> *The Re-Revised Simple Account.* The instrumental value of a state of affairs S is equal to the intrinsic value of the conjunction of the states $S^* \neq S$, such that S^* would obtain if and only if S were to obtain, minus the intrinsic value of the conjunction of the states S^{**}, such that S^{**} would obtain if and only if S were *not* to obtain.

This revision lets us take into account the intrinsically good or bad states that a state S prevents, when calculating the instrumental value of S. Thus revised, the account still implies that pulling L_1 in *Levers* has zero instrumental value. But the new revision implies that pulling L_3 has negative instrumental value. Relative to W_3, the nearest world where you do not pull L_3 is either W_1 or W_2. The Re-Revised Simple Account hence implies that the intrinsic value of W_1 or W_2 (which is the same) should be subtracted from the intrinsic value of W_3, which is 1, in order to arrive at the instrumental value of pulling L_3. Thus, the Re-Revised Simple Account yields the intuitively correct verdict that pulling L_1 is instrumentally better than pulling L_3.[5] (One might still object, of course, to the conclusion that pulling L_1 is instrumentally neutral, rather than instrumentally good.)

Given the assumption that the intrinsic value of a possible world is the sum of the basic intrinsic values of its parts, the Re-Revised Simple Account can be stated in a simpler way, letting $W_{\neg S}$ denote the possible world that would be actual were state S not to obtain:[6]

> *The Counterfactual Comparative Account.* The instrumental value of a state of affairs S is the difference between the intrinsic value of Ws and that of $W_{\neg S}$, minus the intrinsic value of S.

I have renamed the account in order to highlight its close similarity to the much-discussed Counterfactual Comparative Account in the literature on harm and personal value.[7]

I believe, however, that this account also faces fatal counterexamples. This is one:

---

[5] Like the Revised Simple Account, this account also avoids the above-mentioned problems for the Simple Account.

[6] To clarify, $W_{\neg S}$ is assumed to be the non-S-world that is nearest to W_S, rather than the non-S-world that is nearest to the actual world. These two worlds may be different, if the actual world is a non-S-world.

[7] See section 8.

> *Buttons.* God offers you to push one of four buttons, labelled B_1 to B_4. You cannot refuse God's offer. Pushing a button has no intrinsic value. If you push B_i possible world W_i will be actual. W_1 is an extremely good world, and W_2 is almost as good. W_3 is an extremely bad world, and W_4 is even worse. In the nearest possible world where you push B_2 it is true that if you were not to do so, you would push B_1. Further, in the nearest possible world where you push B_3 it is true that if you were not to do so, you would push B_4.[8]

The Counterfactual Comparative Account implies that pushing B_2 is instrumentally bad, while pushing B_3 is instrumentally good. This conjunction of claims is highly implausible in itself, and it has the even more implausible implication that pushing B_3 is instrumentally *better* than pushing B_2. This follows if we assume the principle, which I take to be a conceptual truth, that any good bearer of a certain kind of value is better, as regards this kind of value, than any bad bearer of the same kind of value.

5. Contextualism and Contrastivism

Ben Bradley has suggested a contextualist version of the Counterfactual Comparative Account.[9] On this account, different conversational contexts pick out different similarity relations between possible worlds.[10] It is hence context-dependent what the nearest non-*S*-world is, for a given state *S*. Therefore, Bradley's account does not imply, in *Buttons*, that pushing B_2 is instrumentally bad, or that pushing B_3 is instrumentally good *simpliciter*. Rather, pushing B_2 is instrumentally good relative to contexts where it is true that you would otherwise push B_3 or B_4, and instrumentally bad relative to contexts where it is true that you would otherwise push B_1. Similarly, pushing B_3 is instrumentally good relative to contexts where it is true that you would otherwise push B_4, and instrumentally bad relative to contexts where it is true that you would otherwise push B_1 or B_2.

This contextualist element does not save Bradley's account from trouble in *Buttons*. In the stipulated context, call it *C*, it is true in the nearest world where you push B_2 that you would otherwise push B_1, and also true in the nearest world where you push B_3 that you would otherwise push B_4. Hence, Bradley's account implies that pushing B_3 is instrumentally good and that pushing B_2 is instrumentally bad, relative to *C*. It follows that pushing B_3 is instrumentally better than pushing B_2,

---

[8] Essentially this example is given in Carlson (2020: 409), as part of an argument against the Counterfactual Comparative Account of harm and benefit. See also Carlson, Johansson & Risberg (2021, forthcoming).

[9] Bradley (1998). He intends his account to cover extrinsic value in general, considered as a broader category than instrumental value (see footnote 2). In his (2009: 50-52), Bradley proposes a similar account for personal extrinsic value.

[10] Bradley (1998: 116); cf. Bradley (2009: 50).

relative to *C*. But, it seems to me, pushing B_3 is not instrumentally better than pushing B_2 relative to *any* context.

Bradley might object that *C* is for some reason an unrealistic context. But this does not seem to be the case. To make the stipulated counterfactuals plausible, suppose, for instance, that you can reach B_1 and B_2 most easily with your left hand, while B_3 and B_4 are most easily reached with your right hand. Suppose you just pick a button, say B_2. (Maybe you are unaware of the effects of pushing the buttons.) Had you not pushed B_2, you would still have used your left hand and pushed B_1. Had you pushed B_3, on the other hand, it would have been true that if you had not done so, you would still have used your right hand and pushed B_4.

An idea in the vicinity of Bradley's contextualism is to formulate the Counterfactual Comparative Account as a contrastivist account.[11] According to such an account, a state's instrumental value is relativized to a relevant contrast state. Thus, a state *S* may be instrumentally good relative to state *S** (if W_S is intrinsically better than W_{S*}), but instrumentally bad relative to state *S*** (if W_{S**} is intrinsically better than W_S). Another way to express these contrastive evaluations is to say that it is instrumentally good that *S* obtains rather than *S**, but instrumentally bad that *S* obtains rather than *S***.

Applied to *Buttons*, this account avoids the implausible result that pushing B_2 is instrumentally bad and pushing B_3 is instrumentally good. Hence, we cannot draw the even more implausible conclusion that pushing B_3 is instrumentally better than pushing B_2. What the contrastive account implies is that pushing B_2 rather than B_1 is instrumentally bad, that pushing B_2 rather than B_3 or B_4 is instrumentally good, that pushing B_3 rather than B_1 or B_2 is instrumentally bad, and that pushing B_3 rather than B_4 is instrumentally good.

My main objection to this account is that it is too uninformative. Suppose we are asking whether pushing B_2 is instrumentally good or bad. The reply that pushing B_2 rather than B_3 or B_4 is instrumentally good, whereas pushing B_2 rather than B_1 is instrumentally bad, does not really seem to answer our question. A possible response to this objection would be to claim that for any state of affairs, there is only one relevant contrast state. This would preclude that a state is instrumentally good relative to one contrast state and instrumentally bad relative to another, but it would make the account even less informative. Given the counterfactuals stipulated in *Buttons*, the contrast state to pushing B_2 would have to be pushing B_1, and the contrast state to pushing B_3 would have to be pushing B_4. All we would be able to say about the instrumental value of pushing B_2, then, would be that pushing B_2 rather than B_1 is instrumentally bad. Similarly, all we could say about pushing B_3 would be that pushing B_3 rather than B_4 would be instrumentally good. No comparison could be made between the instrumental value of pushing B_2 and that of pushing B_3.

[11] Comments by an anonymous reviewer prompted me to discuss this possibility. Alastair Norcross (2005) has suggested a contrastive version of the Counterfactual Comparative Account of harm and benefit.

This seems unsatisfactory. (It might be suggested that if a state is instrumentally good relative to its contrast state, then it is instrumentally good *simpliciter*. But this move would take us back to the standard Counterfactual Comparative Account.)

6. The Midpoint Account

A potential lesson to draw from the failure of the Counterfactual Comparative Account is that the relevant comparison, for determining the instrumental value of a state *S*, is not what *would* be the case if *S* were not to obtain, but rather what *could* be the case. Thus, in *Buttons* it seems that we should compare the outcome of pushing a certain button with the respective outcomes of pushing the other buttons, and not just with that of not pushing the button in question.[12] More generally, we should compare a given state *S* to the states that are, in some sense, alternatives to *S*.

In order to capture this idea, let us assume that for any state *S*, there is a finite set of mutually exclusive states that contains *S* and its alternatives. Call such a set an *alternative-set*. The alternatives to *S* are the states that might obtain instead of *S*. (Somewhat more will be said about this assumption below.) Let $A_S = \{S, S^*, \ldots, S^{**}\}$ be the alternative-set to which *S* belongs, and let $A_{WS} = \{W_S, W_{S^*}, \ldots, W_{S^{**}}\}$ be the corresponding set of possible worlds. A straightforward suggestion is that the instrumental value of *S* is determined by comparing the intrinsic value of W_S to the intrinsic value of the best and the worst world in A_{WS}. Thus, add the intrinsic values of these two worlds, and divide this sum by 2.[13] Call the result the *midpoint* of A_{WS}. We can now consider:

> *The Midpoint Account.* The instrumental value of a state of affairs *S* is the difference between the intrinsic value of *Ws* and the midpoint of A_{WS}, minus the intrinsic value of *S*.[14]

---

[12] By the "outcome" of a state of affairs I mean the possible world that would be actual were the state to obtain.

[13] If two or more worlds are tied for best (worst) in A_{WS}, choose any of the best (worst) worlds.

[14] Why not instead choose the *average* intrinsic value of the worlds in A_{WS} as the baseline, and define the instrumental value of *S* as the difference between the intrinsic value of W_S and this average, minus the intrinsic value of *S*? A drawback of this account is that it makes instrumental value depend on the number of alternatives, in an arguably implausible way. Consider a situation in which states S_1 and S_2, which both have zero intrinsic value, are the only alternatives, and assume that the intrinsic values of W_{S1} and W_{S2} are 10 and −10, respectively. Choosing the average as the baseline yields the result that the instrumental values of S_1 and S_2 are, respectively, 10 and −10. Now suppose that the alternative-set is expanded with S_3, S_4 and S_5, and that the intrinsic values of W_{S3}, W_{S4} and W_{S5} are all −10. In this second situation, the instrumental values of S_1 and S_2 are 16 and −4, respectively. It seems, however, that the instrumental values of S_1 and S_2 should not vary, solely depending on whether S_3, S_4 and S_5 are included in the alternative-set.

This account yields plausible results in the cases we have discussed so far. In *Levers*, it implies that pulling L_1 and pulling L_2 are instrumentally good, while pulling L_3 is instrumentally bad. In *Buttons*, the implications are that pushing B_1 and pushing B_2 are instrumentally good, whereas pushing B_3 and pushing B_4 are instrumentally bad.

As compared to the Counterfactual Comparative Account, a further advantage of the Midpoint Account is that it does not require the questionable assumption that there is, for any state of affairs, a possible world that *would* be actual if this state were not to obtain. The set A_{WS} can be taken to include W_S and the set of worlds that *might* be actual, were S not to obtain. The alternatives to S are then the set of states that might obtain, instead of S, were S not to obtain. We need not assume that one of these states is such that it *would* obtain, in the absence of S. If S is an action, the alternatives to S are plausibly taken to be the other actions, incompatible with S and with each other, that are available to the agent in the situation. If S is an event but not an action, its alternatives might be the set of events, incompatible with S and with each other, whose occurrence at the same time and place is consistent with the past and the laws of nature of W_S.[15]

Concerning states of affairs other than events, it may often be unclear what states should be included in an alternative-set. Consider, for example, the state that Joe Biden is the present President of the United States. Who might have been President now instead of Biden? It is natural to include Donald Trump among the alternatives, and to exclude Abraham Lincoln. But what about Sarah Palin, say? Whether or not she should be included is arguably a context-dependent matter. We might want to consider only persons who actually ran for President in 2020, or we might be willing to consider a larger group of persons. It seems difficult to argue that one choice is objectively more correct than the other. The most feasible fully general version of the Midpoint Account may therefore be one that does not assign instrumental value to states of affairs *simpliciter*, but rather to states relative to an alternative-set, determined by a context of utterance. This allows for the possibility that a state is instrumentally good relative to one alternative-set and instrumentally bad relative to another.

7. Two Objections to the Midpoint Account

To be sure, the Midpoint Account is not unassailable. It has somewhat counterintuitive implications in cases like the following:

> *Knobs.* God offers you to turn one of three knobs, labelled K_1 to K_3. You cannot refuse God's offer. Turning a knob has no intrinsic value. If you turn K_i possible

[15] If physical determinism is true this condition has to be relaxed, in order to avoid the conclusion that no event has any alternatives.

> world W_i will be actual. W_1 is a very good world, having an intrinsic value of 60. W_2 is a very bad world, having an intrinsic value of −110. W_3, finally, is an extremely bad world, having an intrinsic value of −300.

The midpoint of $\{W_1, W_2, W_3\}$ is −120. Hence, the Midpoint Account implies that turning K_2 has an instrumental value of 10, thereby being instrumentally good. But it may seem that turning K_1 is the only instrumentally good alternative in *Knobs*, and that turning K_2 and turning K_3 are both instrumentally bad.

I think, however, that it is defensible to claim that turning K_2 is instrumentally good. After all, it prevents an extremely bad world from being actual. Of course, it also prevents a very good world from being actual. But since the difference in intrinsic value between W_2 and W_3 is greater than that between W_1 and W_2, the former, good aspect of turning K_2 arguably outweighs the latter, bad aspect.

In general terms, the Midpoint Account implies that no matter how bad the outcome of a state S is, and no matter how good alternative outcomes there are, S can be instrumentally good, provided that there is an alternative with an outcome bad enough to lower the midpoint below the intrinsic value of W_S. Conversely, a state with an extremely good outcome, and some extremely bad alternative outcomes, can still be instrumentally bad, if there is an alternative with an enormously good outcome that raises the midpoint high enough.

I am not sure that these implications are unacceptable. In any case, it is worth noting that the Counterfactual Comparative Account faces a similar problem. According to that account, too, a state S with an extremely bad (good) outcome can be instrumentally good (bad), if $W_{\neg S}$ is intrinsically even worse (better) than W_S.

Another objection to the Midpoint Account is that it fails to reflect the importance of *causation*, as regards instrumental value.[16] In one situation, let us suppose, actions a and b are your only alternatives. Both actions would cause a state of affairs S with intrinsic value 10 to obtain, and have no other intrinsically good or bad states in their outcomes. In another possible situation, actions c and d are your only alternatives. They would both cause a state S^* with intrinsic value −10 to obtain, and have no other intrinsically good or bad states in their outcomes. The Midpoint Account implies that a, b, c and d are all instrumentally neutral. But, the objection goes, a and b are in fact instrumentally good, since they would cause an intrinsically good outcome to obtain, and c and d are in fact instrumentally bad, since they would cause an intrinsically bad outcome to obtain.

This objection presupposes controversial claims about causation. Since S is unavoidable in the first situation, the assumption that a and b would each cause S to obtain seems difficult to square with theories of causation honouring the slogan that "causation is difference-making". And likewise regarding c, d and S^* in the second situation. But suppose, for the sake of argument, that the causal claims involved are consistent. Then my inclination is to conclude that causation is less relevant for

[16] This objection stems from comments by Olle Risberg.

instrumental value than one might think. If exactly the same intrinsically good or bad states of affairs will obtain whatever you do in a situation, I find it plausible to conclude that all your alternatives have neutral instrumental value. Whatever is true of causation, it would seem that instrumental goodness and badness require difference-making.

8. Personal Instrumental Value, Harm and Benefit

Several philosophers have proposed the Counterfactual Comparative Account as an account of *personal* instrumental value.[17] In our framework, this proposal can be put as follows:

> *The Counterfactual Comparative Account of personal instrumental value.* The instrumental value for a person *P* of a state of affairs *S* is the difference between the intrinsic value for *P* of *Ws* and that of $W_{\neg S}$, minus the intrinsic value for *P* of *S*.[18]

It is easy to see that this account is vulnerable to a variant of *Buttons*, in which pushing the buttons affects your, or someone else's, personal intrinsic value. As in the case of impersonal instrumental value, the Midpoint Account fares better (although the objections discussed in section 7 are relevant). Define the set A_{WS} as in section 6, and add the intrinsic values for *P* of the best and the worst world for *P* in A_{WS}. Let the *midpoint for P* of A_{WS} be this sum divided by 2. We can now state:

> *The Midpoint Account of personal instrumental value.* The instrumental value for a person *P* of a state of affairs *S* is the difference between the intrinsic value for *P* of *Ws* and the midpoint for *P* of A_{WS}, minus the intrinsic value for *P* of *S*.

As far as I can see, this account is equally plausible for personal as for impersonal instrumental value.

The Counterfactual Comparative Account is even more popular as an account of *harm* and *benefit*:

> *The Counterfactual Comparative Account of harm and benefit.* A state of affairs *S* harms (benefits) a person *P* if and only if the intrinsic value for *P* of *Ws* is lower (higher) than the intrinsic value for *P* of $W_{\neg S}$.[19]

---

[17] See Bradley (2009: 50); Feit (2016: 138f); Feldman (1991: 214f, 1992).

[18] Personal intrinsic value is often equated with welfare.

[19] For defences of this account, see, e.g., Boonin (2014); Bradley (2009); Jedenheim Edling (2021); Feit (2015, 2016, 2019); Klocksiem (2012, 2019); Parfit (1984: 69); Petersson (2018); Purshouse (2016); Timmerman (2019). Not all of these authors give an explicit account of benefit, but in most

One of several problems with this account is that it is vulnerable to variants of *Buttons*. If we take the value assumptions in that case to concern your personal intrinsic value, the Counterfactual Comparative Account implies that pushing B_2 would harm you, whereas pushing B_3 would benefit you. This runs afoul of a very plausible principle, stating that if *a* and *a** are alternative actions open to you in a situation, and doing *a* would benefit you while doing *a** would harm you, then you have a prudential reason to do *a* rather than *a**. In *Buttons*, there seems to be absolutely no reason for you to push B_3 rather than B_2. Moreover, the account also violates another very plausible principle, to the effect that if states *S* and *S** belong to the same alternative-set and the intrinsic value for *P* of W_S is much higher than that of W_{S^*}, then *S* would harm *P* only if *S** would, and *S** would benefit *P* only if *S* would.[20]

Again, the Midpoint Account seems more promising:

> *The Midpoint Account of harm and benefit.* A state of affairs *S* harms (benefits) a person *P* if and only if the intrinsic value for *P* of *Ws* is lower (higher) than the midpoint for *P* of A_{WS}.

Assuming that it is your personal intrinsic value that is at stake in the cases we have considered, this account implies that pulling L_1 or L_2 would benefit you in *Levers*, while pulling L_3 would harm you. In *Buttons*, pushing B_1 or B_2 would benefit you, whereas pushing B_3 or B_4 would harm you. In *Knobs*, finally, turning K_1 or K_2 would benefit you, while turning K_3 would harm you. Of these results, the only one that is not intuitively quite plausible is that turning K_2 would benefit you. (Obviously, this is closely connected to the first objection discussed in section 7.)

I am, nevertheless, unsure whether the Midpoint Account is acceptable as an account of harm and benefit.[21] Its plausibility will largely depend on how well it handles variants of much-discussed difficulties for the Counterfactual Comparative Account; in particular the "preemption" and "failure to benefit" problems.[22] Pursuing these matters here would, however, take us too far afield.

---

cases it is clear that they take benefit to be analogous to harm. The Counterfactual Comparative Account is typically taken to be an account of *overall*, rather than *pro tanto*, and *extrinsic*, rather than *intrinsic*, harm and benefit. A state of affairs is intrinsically (extrinsically) harmful or beneficial to the extent that it is harmful or beneficial because of its intrinsic (extrinsic) properties.

[20] These criticisms are developed in Carlson (2019, 2020) and in Carlson, Johansson & Risberg (2021).

[21] A general argument against "well-being counterfactualist" accounts of harm and benefit, to which category the Midpoint Account belongs, is stated in Carlson, Johansson & Risberg (2021: 171-73).

[22] For a thorough discussion of the preemption problem, see Johansson & Risberg (2019). The failure to benefit problem is discussed in, e.g., Feit (2019); Purves (2019); Johansson & Risberg (2020); Klocksiem (2022).

9. Concluding Remarks

I have tentatively suggested the Midpoint Account as an account of impersonal and personal instrumental value, and also floated it as a possible account of harm and benefit. Even if these accounts should ultimately be rejected, there may be some weaker positive results to be salvaged. The central idea behind the suggested accounts is that the instrumental value of a state of affairs depends on how its outcome compares to those of alternative states, in terms of intrinsic value. If this idea is sound, we may at least have arrived at a partial account of instrumental *betterness*. According to this partial account, a state S is impersonally instrumentally better than an alternative state S^* if and only if W_S is intrinsically better than W_{S^*}. And analogously for personal instrumental value. To this partial account, the Midpoint Account adds a zero point or baseline, categorizing states as instrumentally good, bad or neutral, and allowing for comparisons of instrumental value across alternative-sets. Clearly, the partial betterness account may be correct also if the Midpoint Account mislocates the baseline. Similarly, even if the Midpoint Account of harm and benefit is wrong about the baseline separating beneficial states from harmful ones, it may nevertheless be true that a state S is less harmful or more beneficial than an alternative state S^*, for a person P, just in case W_S is intrinsically better for P than W_{S^*}. If so, we have at least obtained a partial account of the relation "less harmful or more beneficial than".

A possible and somewhat skeptical position is that these partial accounts of instrumental betterness and relative harmfulness are accurate, but that there is no general way to correctly locate the baseline. The factors relevant for determining the baseline may be different, or have different relative weights, for different alternative-sets.[23]

References

Boonin, David (2014) *The Non-Identity Problem and the Ethics of Future People*. New York: Oxford University Press.

Bradley, Ben (1998) "Extrinsic Value". *Philosophical Studies*, 91(2): 109-26.

Bradley, Ben (2009) *Well-Being and Death.* New York: Oxford University Press.

Bradley, Ben (2012) "Doing Away with Harm". *Philosophy and Phenomenological Research*, 85(2): 390-412.

[23] Jens Johansson, Olle Risberg, and two anonymous reviewers gave very helpful comments on earlier versions of this paper. I also wish to acknowledge financial support from Riksbankens Jubileumsfond, Grant P21-0462, and Vetenskapsrådet, Grant 2018-01361.

Carlson, Erik (2019) "More Problems for the Counterfactual Comparative Account of Harm and Benefit". *Ethical Theory and Moral Practice*, 22(4): 795-807.

Carlson, Erik (2020) "Reply to Klocksiem on the Counterfactual Comparative Account of Harm." *Ethical Theory and Moral Practice*, 23(2): 407-13.

Carlson, Erik, Jens Johansson & Olle Risberg (2021) "Well-Being Counterfactualist Accounts of Harm and Benefit". *Australasian Journal of Philosophy*, 99(1): 164-74.

Carlson, Erik, Jens Johansson & Olle Risberg (forthcoming) "Benefits Are Better than Harms: A Reply to Feit". *Australasian Journal of Philosophy*.

Feit, Neil (2015) "Plural Harm". *Philosophy and Phenomenological Research*, 90(2): 361–88.

Feit, Neil (2016) "Comparative Harm, Creation and Death". *Utilitas*, 28(2): 136–63.

Feit, Neil (2019) "Harming by Failing to Benefit". *Ethical Theory and Moral Practice*, 22(4): 809–23.

Feldman, Fred (1991) "Some Puzzles About the Evil of Death". *The Philosophical Review*, 100(2): 205–27.

Feldman, Fred (1992) *Confrontations with the Reaper*. New York: Oxford University Press.

Feldman, Fred. (2000) "Basic Intrinsic Value". *Philosophical Studies*, 99(3): 319-46.

Jedenheim Edling, Magnus (2022) "A New Principle of Plural Harm". *Philosophical Studies*, 179(6): 1853-72.

Johansson, Jens & Olle Risberg (2019) "The Preemption Problem". *Philosophical Studies*, 176(2): 351-65.

Johansson, Jens & Olle Risberg (2020) "Harming and Failing to Benefit: A Reply to Purves". *Philosophical Studies*, 177(6): 1539-48.

Klocksiem, Justin (2012) "A Defense of the Counterfactual Comparative Account of Harm". *American Philosophical Quarterly*, 49(4): 285–300.

Klocksiem, Justin (2019) "The Counterfactual Comparative Account of Harm and Reasons for Action and Preference: Reply to Carlson". *Ethical Theory and Moral Practice*, 22(3): 673-77.

Klocksiem, Justin (2022) "Harm, Failing to Benefit, and the Counterfactual Comparative Account". *Utilitas*, 34(4): 428-44.

Norcross, Alastair (2005) "Harming in Context". *Philosophical Studies*, 123(1/2): 149-73.

Parfit, Derek (1984) *Reasons and Persons*. Oxford: Oxford University Press.

Petersson, Björn (2018) "Over-Determined Harms and Harmless Pluralities". *Ethical Theory and Moral Practice*, 21(4): 841–50.

Purshouse, Craig (2016) "A Defence of the Counterfactual Account of Harm". *Bioethics*, 30(4): 251-59.

Purves, Duncan (2019) "Harming as Making Worse Off". *Philosophical Studies*, 176(10): 2629–56.

Rønnow-Rasmussen, Toni (2002) "Instrumental Values – Strong and Weak". *Ethical Theory and Moral Practice*, 5(1): 23-43.

Rønnow-Rasmussen, Toni (2015) "Intrinsic and Extrinsic Value" in I. Hirose & J. Olson (Eds.) *The Oxford Handbook of Value Theory* (29-43). New York: Oxford University Press.

Timmerman, Travis (2019) "A Dilemma for Epicureanism". *Philosophical Studies*, 176(1): 241–57.

Zimmerman, Michael J. (2001) *The Nature of Intrinsic Value*. Lanham, MD: Rowman and Littlefield.

Zimmerman, Michael J. & Ben Bradley (2019) "Intrinsic vs. Extrinsic Value" in Edward N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.