

Against Negative Utilitarianism

Johan E. Gustafsson*

Draft: October 21, 2023 at 9:58 p.m.

ABSTRACT. According to Negative Utilitarianism, you ought to minimize the sum total of pain, whereas, according to (the more standard) Classical Utilitarianism, you ought to maximize the sum total of pleasure minus pain. There are several well-known counter-examples to Negative Utilitarianism. Yet, for many of them, there are analogous counter-examples to Classical Utilitarianism. So these objections have little force when we assess the relative merits of Classical and Negative Utilitarianism. Some further counter-examples to Negative Utilitarianism may, arguably, be resisted if we cling to the intuition that evil and suffering have greater moral import than goodness and happiness. And, some of these counter-examples may be blocked if we modify Negative Utilitarianism so that the sum total of pleasure breaks ties between outcomes which have the same sum total of pain. I present a new counter-example to Negative Utilitarianism which avoids these drawbacks. In addition, I also present counter-examples to suffering-focused variations of Negative Utilitarianism.

Classical Utilitarianism has a suffering-focused twin. Whereas *Classical Utilitarianism* is the view that you ought to maximize the sum total of pleasure minus pain, *Negative Utilitarianism* is the view that you ought to minimize the sum total of pain.¹ There are several well known counter-examples to Negative Utilitarianism.² Yet, for many of them, there are analogous counter-examples to Classical Utilitarianism. This

* I would be grateful for any thoughts or comments on this paper, which can be sent to me at johan.eric.gustafsson@gmail.com.

¹ The label 'Negative Utilitarianism' come from R. N. Smart 1958, p. 542. (Smart used the label 'Positive Utilitarianism' for Classical Utilitarianism, but that suggests the view that we should just maximize the sum total of pleasure.) Popper (1945, pp. 205n6, 241–242n2) suggests a view along these lines, but he (1962, p. 386) does not accept it as a criterion of rightness. For more contemporary defences of Negative Utilitarianism, see Wolf 1996, p. 273, Metzinger 2003, p. 570, and Knutsson 2021, pp. 1004–1014.

² Why should we be interested in counter-examples to Negative Utilitarianism? For most non-utilitarian moral theories there are more convincing objections that do not rely on intuitions about cases. Many moral theories are open to structural objections such as value-pump arguments. By contrast, Negative Utilitarianism has the same overall structure as Classical Utilitarianism, which is not open to value-pump arguments. So we cannot rule out Negative Utilitarianism on those grounds. (Note, as we shall see

weakens their force when we assess the relative merits of Classical and Negative Utilitarianism. And the counter-examples that lack classical analogies can, as we shall see, be resisted in other ways.

The Elimination Argument is the argument that Negative Utilitarianism may prescribe eliminating everyone to avoid future suffering.³ Nevertheless, Classical Utilitarianism may also prescribe eliminating everyone in some circumstances, and it is open to the *Replacement Argument* — the similar charge that Classical Utilitarianism may prescribe that we kill everyone if that would let us replace everyone with happier people.⁴

Pinpricks versus Torture is a choice between (i) torture for one person and (ii) a pinprick for each of a large number of people. Given that sufficiently many people would get a pinprick, Negative Utilitarianism prescribes the torture.⁵ But so does Classical Utilitarianism.⁶

The Pinprick Argument is the argument that Negative Utilitarianism may prescribe choosing an outcome with no pleasure and no pain for anyone rather than an outcome where everyone enjoys an enormous amount of pleasure and someone gets a pinprick.⁷ While the Pinprick Argument is compelling, it may (arguably) be resisted if we cling to the intuition that evil and suffering have greater moral import than goodness and happiness.⁸

The Indifference Argument is the argument that Negative Utilitarianism permits refraining from bringing about more pleasure even if the sum total of pain would be unchanged.⁹ The Indifference Argument could be

in the appendix, that this need not be true for more complicated variants of Negative Utilitarianism which do deviate from the overall structure of Classical Utilitarianism.) Furthermore, since Negative and Classical Utilitarianism share the same overall structure, structural arguments are of little help when we compare their relative merits.

³ Hedenius 1955, pp. 104–105, R. N. Smart 1958, p. 542, and Bergström 1978. The name comes from Arrhenius and Bykvist 1995, p. 31.

⁴ See Knutsson 2021, pp. 1005–1009 for these analogues, and Hurka 1982, pp. 66–67 and Hiž 1992, pp. 425–426 for some examples. For the Replacement Argument, see Singer 1979, pp. 100–103 and Hart 1980, pp. 28–31.

⁵ This kind of counter-example has been levelled against Negative Utilitarianism in Arrhenius and Bykvist 1995, p. 31.

⁶ The same equivalence between Classical and Negative Utilitarianism also holds for J. J. C. Smart's (1989, p. 40) variation where an innocent person is severely punished to avoid a large number of people each getting a minor pain.

⁷ Arrhenius and Bykvist 1995, pp. 40–41 and Gandjour and Lauterbach 2003, p. 241. The name comes from Pearce 2005.

⁸ See, for example, Santayana 2011, pp. 144–145.

⁹ Arrhenius and Bykvist 1995, pp. 36–37. The name comes from Ord 2013.

blocked, however, by making Negative Utilitarianism lexical so that the sum total of pleasure breaks ties between outcomes which have the same sum total of pain.¹⁰

Hence these counter-examples could (arguably) be resisted with at least some plausibility. Nevertheless, by combining different elements from these counter-examples, we can construct a more challenging counter-example. The point of this new counter-example to Negative Utilitarianism is that (i) it is compelling, (ii) it cannot, plausibly, be blocked by clinging to the intuition that evil and suffering have greater moral import than goodness and happiness, (iii) it cannot plausibly be blocked by a lexical tie-breaker rule, (iv) it has no analogue for Classical Utilitarianism, and (v) it is applicable to fixed populations. Earlier counter-examples have, as I have argued, failed on at least one of these scores.¹¹ The following is a better counter-example.

Consider

Bliss versus Torture You have a choice between the following outcomes, where the same people live for the same duration:

- A Everyone gets a century of pure bliss followed by a pinprick.
- B Someone gets a century of torture, and everyone else gets a century of no pleasure and no pain.

Given a large enough population, you ought to choose *B* over *A* according to Negative Utilitarianism. Yet *A* seems more choice-worthy than *B* on, basically, any plausible moral metric: *A* is overwhelmingly in everyone's subjective interest (given, as seems plausible, that everyone strongly

¹⁰ Arrhenius and Bykvist 1995, p. 40.

¹¹ A potential exception may be an example that McMahan (2013, pp. 22–23) put forward, in a different context. It is a single-person case with a choice between (i) creating a person who lives a 2 year long life of unremitting pain and (ii) creating another person who lives a typical life for at least 80 years where the benefits would greatly exceed the harms yet where the sum total of pain would be greater than in the 2-year life. McMahan's case could serve as a compelling counter-example to Negative Utilitarianism. But, since it involves variable population, it does not (as it stands) work against a version of Negative Utilitarianism which is restricted to fixed populations.

prefers ending up in *A* to ending up in *B*).^{12,13} *A* is more equal than *B*. The worse-off are better off in *A* than in *B*. There is less torture in *A* than in *B*. And so on.

Furthermore, it seems that Bliss versus Torture lacks a clear analogue for Classical Utilitarianism.¹⁴ (But it is, admittedly, somewhat slippery

¹² Note that this could be true even if there is no personal identity over a full century. The current people could still be related by what matters in survival to the later persons; see Parfit 1971. Note, furthermore, that the claim that something is in an individual's subjective interest is not the substantive evaluative claim that it would be better for them. It is the claim that the individual judges, by their own lights, that they prefer it for their own sake.

¹³ In this respect, the example is similar to Ord's (2013) *Worse-for-Everyone Argument*, which is based on a case where he claims (plausibly, in my opinion) that following Negative Utilitarianism makes everyone worse off. Ord's argument, however, may (at least in part) seem to assume the point at issue against versions of Negative Utilitarianism without trade-offs between pleasure and pain, because it assumes that, for personal value, pleasure can outweigh pain. To avoid this, I'm only claiming that everyone, plausibly, has a strong *preference* for ending up in *A* rather than in *B*. Broome (1991, pp. 152–154) argues against the Pareto Principle, which says that a prospect *X* is better than a prospect *Y* if everyone prefers *X* to *Y*. Instead of the Pareto Principle, Broome (1991, p. 155) favours the Principle of Personal Good, which says that a prospect *X* is better than a prospect *Y* if *X* is better for everyone than *Y*. Broome's argument against the Pareto Principle, however, doesn't work in case the prospects do not involve risk. So it doesn't apply to the use of the Pareto Principle in Bliss versus Torture, to conclude that *A* is better than *B*.

¹⁴ Magnus Vinding suggests a possible analogue, where, in *A'*, everyone gets a century of hedonic neutrality filled with a very large amount of non-hedonic goods followed by a pinprick and, in *B'*, someone gets a century of torture and everyone else gets a century of hedonic neutrality followed by two pinpricks and four micro pleasures (pleasures corresponding in intensity to a pinprick). Given a sufficiently large population, you ought to choose *B'* according to Classical Utilitarianism. But this counter-example is disanalogous. It introduces, in addition to pleasures and pains, a third element, namely non-hedonic goods. And, if those goods are good for people, it would merely motivate a switch to a version of Classical Utilitarianism where these non-hedonic goods also contribute to well-being rather than a switch to Negative Utilitarianism. Vinding also suggest a possible analogue where, in *A''*, everyone gets a century of hedonic neutrality followed by a single pinprick and, in *B''*, someone gets a century of torture and everyone else gets 50 years of pinpricks and 50 years of micro pleasures followed by two micro pleasures. Once again, given a sufficiently large population, you ought to choose *B''* according to Classical Utilitarianism. But, if the micro pleasures really have the corresponding intensity to the pinpricks, they should outweigh the pinpricks in *B''* (otherwise they wouldn't have the corresponding intensity according to Classical Utilitarianism). If you don't find that intuitive, you may be imagining micro pleasures of too low intensity. Once the intensity is imagined correctly, *B''* should then be in most people's subjective interest (or, at least, *A''* would not be overwhelmingly in their interest).

what counts as analogous to what.) Furthermore, the suffering in *B* seems morally worse, and a greater evil, than the suffering in *A*. Finally, this counter-example cannot be blocked by letting the sum total of pleasure break ties between outcomes with the same sum total of pain. Hence this counter-example to Negative Utilitarianism should be more challenging than the earlier counter-examples.

It may be objected that we could deny the possibility of any disanalogy between Classical and Negative Utilitarianism by claiming that there is no such thing as positive well-being.¹⁵ If there is no such thing as positive well-being, the bliss in *A* would count for nothing. And, if so, even Classical Utilitarianism would prescribe *B*. But note that the claim that bliss does not provide positive well-being (that is, that bliss is not good for the person) is a substantive evaluative claim. This claim, which Classical Utilitarianism rules out, is part of the package of normative and evaluative ideas that the counter-example targets. Bliss versus Torture does not refer to positive and negative well-being; it only refers to bliss, pinpricks, and torture. Bliss, pinpricks, and torture do exist.

Still, one may defend the rejection of positive well-being with the claim that there are no experiences with a positive hedonic tone.¹⁶ But this conflicts with what I suspect is most people's introspective judgement that pleasurable experiences from the best things in life have a positive hedonic tone.¹⁷

Furthermore, it seems that our reasons for thinking that there is negative well-being are analogous to our reasons for thinking that there is positive well-being. If we are wrong about the existence of positive well-being, we should also be sceptical about the existence of negative well-being.

It may next be objected that a version of Negative Utilitarianism that focused, not on pain, but on negative lifetime well-being would avoid this counter-example. According to *Lifetime Negative Utilitarianism*, pleasure

¹⁵ Knutsson 2021, p. 1013. See also Fehige's (1998, p. 518) view that satisfied preferences are not good but frustrated preferences are bad. Moreover, Schopenhauer (*WWR2*, §46; 2018, p. 590) takes pleasure to be the absence of pain.

¹⁶ Knutsson 2021, p. 1013.

¹⁷ Note, however, that the rejection of experiences with a positive hedonic tone need not conflict with the existence of experiences that are not painful (hence not negative) yet which get more pleasant. One might think that, when such non-negative experiences get more pleasant (so that their hedonic tone improves), they should result in experiences with a positive hedonic tone. But this ignores the possibility that there is a range in the hedonic spectrum that is not positive (good), not negative (bad), not zero (neutral); see Gustafsson 2019, p. 101.

compensates pain within a life just like in Classical Utilitarianism, but you ought minimize the sum total of negative lifetime well-being.¹⁸ This view avoids my counter-example, since the pinpricks are compensated by the bliss so that there is no negative lifetime well-being in *A*. But it gives less weight to torture for a person who still ends up with non-negative lifetime well-being than a pinprick for a person who ends up with negative lifetime well-being, which is implausible.¹⁹

The implication that *B* ought to be chosen in Bliss versus Torture would not follow on a lexical variant of Negative Utilitarianism, where the pains that are more severe than a certain critical level of severity are lexically more important than other pains.²⁰ So, on this lexical view, we ought to minimize the amount of pain that is more severe than the critical level. This view, however, is open to a similar counter-example. Let us call pains that are more severe than the critical level *severe* pains. Consider

Bliss and Severe Pain versus Almost-Severe Pains You have a choice between the following outcomes, where the same people live for the same duration:

- C Everyone gets a century of pure bliss followed, for someone, by the briefest, least severe pain that counts as severe.
- D Everyone gets a century of pain that is just slightly less severe than the critical level in severity.

On the lexical view, we ought to choose *D*, no matter how large the population is. Yet *D* is worse than *C* on almost any plausible moral metric. There is a lot more suffering in *D* than in *C*. While the pain in *C* is worse in severity, it is only slightly worse in severity than the pains in *D* which last much longer and afflict an arbitrarily large number of people. The

¹⁸ See Arrhenius and Bykvist 1995, p. 33.

¹⁹ Arrhenius and Bykvist (1995, pp. 34–25) object that Lifetime Negative Utilitarianism may permit bringing about more pain other things being equal, since the resulting lifetime well-being still ends up non-negative. But this objection can be blocked by modifying Lifetime Negative Utilitarianism so that the sum total of positive lifetime well-being break ties between outcomes which have the same sum total of negative lifetime well-being.

²⁰ See Hedenius 1955, p. 100; Locke 1987, pp. 147–148 and Arrhenius and Bykvist 1995, p. 35. See Sidgwick 1907, pp. 123–124n for a pessimistic assessment of this kind of view. See also Klocksiesm 2016 for a general defence of lexicality in the value of hedonic episodes.

difference in severity between the pain in *C* and the pains in *D* is, we can assume, barely perceptible.²¹

Moreover, *C* is in everyone's subjective interest, given (as seems plausible) that everyone prefers ending up in *C* to ending up in *D*. Obviously, the people who wouldn't get any pain in *C* prefer *C* to *D*. But consider the person who would suffer the severe pain in *C*. In this person's subjective interest, the barely perceptible additional severity of the pain in *C* is plausibly mitigated by a century of bliss and outweighed by the much longer duration of the almost as severe pain in *D*.

It may seem that *D* is better in one respect, namely, it is more equal. But we could easily change the example so that some small portion of the people in *D* would not get any pain at all. Then *D* would be unequal too.

It may be objected that none of the counter-examples we have discussed so far applies to a variant of utilitarianism that gives weight to pleasure but that gives more weight to pain.²² According to *Weak Negative Utilitarianism*, you should maximize the sum total of pleasure and pain with the pain multiplied by a certain constant greater than 1. This variant is a much less drastic departure from Classical Utilitarianism than Negative Utilitarianism. It is similar in many respects to *Prioritarianism*, a variant of utilitarian where benefiting people matters more the worse off they are.²³ And it is open to some of the same objections. A compelling account of the units on the scale of well-being is that these units measure how much these differences contributes to moral value. But, if units of well-being count differently for moral value depending on where they are on the scale of well-being, then we need another account of the units on the scale of well-being.²⁴ Another compelling account of the units on the scale of well-being is that they measure differences in personal value

²¹ This assumes that there aren't any large gaps in the range of possible severities of pain such that, for one level of pain severity, all possible levels that are less severe are a lot less severe. This kind of gap strikes me as implausible. Even if humans couldn't experience pains with a severity in a certain range, it seems that some possible beings should be able to do so.

²² See Walker 1974, p. 426 and Arrhenius and Bykvist 1995, pp. 43–46.

²³ Atkinson and Stiglitz 1980, p. 340 and Parfit 1995, p. 19.

²⁴ Broome (1991, p. 217) levels this sort of objection against Prioritarianism, whereas Ord (2013) levels it against Weak Negative Utilitarianism. Mathiason (2018, pp. 14–15) stipulates a hedonic scale where the difference between the most severe pain imaginable and no feeling is equal to the difference between the most pleasant pleasure imaginable and no feeling. But it's hard to see why we should attach any significance to the units of this scale.

(that is, how good or bad things are for the person), where the amount of personal value is understood in a way that is independent from its contribution to moral value. But, given this account, Weak Negative Utilitarianism will in some cases prescribe outcomes that are worse for everyone, because trade-offs that are good for each person need not be morally good.²⁵ Consider the following case:

Pain plus More Pleasure You have a choice between the following outcomes, where the same people live for the same duration:

- E* Everyone gets a painful experience, amounting to 1 unit of negative well-being, and a pleasurable experience, amounting to $1 + \epsilon$ units of positive well-being (where ϵ is positive).
- F* No one gets any painful or pleasurable experience, and hence no one gets any positive or negative well-being.

Given that ϵ is sufficiently small, you ought to choose *F* according to Weak Negative Utilitarianism. Yet it would be better for everyone if you chose *E*, because each person gets a net gain of ϵ units of well-being from *E*.

It may be objected that we can avoid this counter-example if we allow that experiences of positive and negative well-being are given the same weight within a life, but still claim that negative lifetime well-being counts more than positive lifetime well-being for moral value. According to *Weak Lifetime Negative Utilitarianism*, you should maximize the sum total of positive and negative lifetime well-being with the negative lifetime well-being multiplied by a certain constant greater than 1.²⁶

²⁵ Ord 2013.

²⁶ MacAskill (2019, p. 236) argues that under moral uncertainty we should treat avoiding pain as more important than promoting pleasure, because (i) there are some plausible moral views where avoiding pain is more important than promoting pleasure and (ii) there are no plausible moral views where promoting pleasure is more important than avoiding pain. But, as I hope our discussion has shown, premise (i) is dubious. And, if we lower the standards for plausibility so that (i) is credible, then (ii) is dubious. Parfit (1986, pp. 161–164), for example, finds a version of perfectionism plausible which the best things in life are lexically more important than anything else (although he does leave open how perfectionists should handle great suffering). And, while it is hard to justify, it doesn't seem entirely unreasonable that we shouldn't reduce suffering among non-human animals if it would rob humans of the best things in life.

But note that this no longer fits with the intuition that evil and suffering have greater moral import than goodness and happiness. If two people will both end up with positive (or both with negative) well-being in any case, then Weak Lifetime Negative Utilitarianism will not favour avoiding pain over promoting pleasure, other things being equal. Consider the following case:

Pain versus More Pleasure You have a choice between the following outcomes, where the same two people — Alice and Bob — live for the same duration:

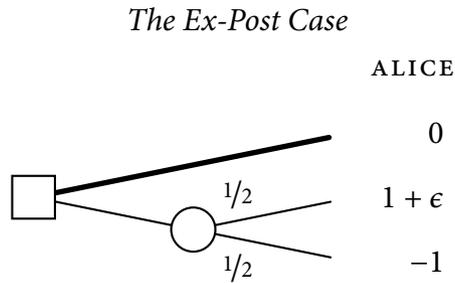
- G* Alice gets a painful experience, amounting to 1 unit of negative well-being, and Bob gets a pleasurable experience, amounting to $1 + \delta$ units of positive well-being.
- H* Neither Alice nor Bob gets any painful or pleasurable experience, and hence neither Alice nor Bob gets any further positive or negative well-being.

Given that Alice and Bob have so far had so much pleasure in their lives that they will end up with a positive lifetime well-being in any case (or so much pain that they will end up a negative lifetime well-being in any case), you ought to choose *H* according to Weak Lifetime Negative Utilitarianism if δ is a positive constant however small. Likewise, if δ is zero, then it does not matter whether you choose *G* or *H*, according to Weak Lifetime Negative Utilitarianism. This does not fit with the basic intuition behind Negative Utilitarianism, namely, the intuition that reducing suffering is more important than promoting pleasure other things being equal.

There are also some more technical problems once we extend Weak Lifetime Negative Utilitarianism to handle risky prospects. There are two main ways of extending Weak Lifetime Negative Utilitarianism so that it can handle risky prospects. According to *Ex-Post Weak Lifetime Negative Utilitarianism*, you should maximize the expectation of the sum total of positive and negative lifetime well-being with the negative lifetime well-being multiplied by a certain constant greater than 1. And, according to *Ex-Ante Weak Lifetime Negative Utilitarianism*, you should maximize the sum total of positive and negative expectation of lifetime well-being with the negative expectation of lifetime well-being multiplied by certain constant greater than 1. Both of these views may prescribe choices that lowers everyone's expectation of lifetime well-being relative to their expectations from some alternative choices. On the *ex-post* approach, this can happen

with a single choice for a single person. On the *ex-ante* approach, this can happen with a sequence of choices.²⁷

To see how *Ex-Post* Weak Lifetime Negative Utilitarianism can lower the expectations of lifetime well-being for everyone, consider the following case, where you have single choice which only affects Alice:



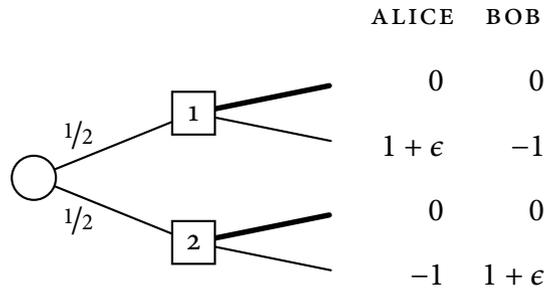
Here, the square represents a choice node where you have a choice between going up or down. If you go up, then Alice gets a lifetime well-being of 0. If you go down, we reach the chance node represented by the circle, where chance goes up or down depending on a fair coin toss. If chance goes up at the chance node, Alice gets a lifetime well-being of $1 + \epsilon$, where ϵ is a small positive constant. If chance goes down at the chance node, Alice gets a lifetime well-being of -1 .

Given that ϵ is sufficiently small, the adjusted moral value of Alice getting a lifetime well-being of -1 will outweigh the moral value of Alice getting a lifetime well-being of $1 + \epsilon$. So the expectation of an equal lottery between these outcomes is then negative and therefore worse than the expectation of going up, which is zero. So, according to *Ex-Post* Weak Lifetime Negative Utilitarianism, you should go up at the choice node (this prescription is represented by the thicker line). But this lowers Alice's expectation of lifetime well-being, because her expectation from going up is 0 but her expectation from going down is $\epsilon/2$.

Ex-Ante Weak Lifetime Negative Utilitarianism avoids this result in the *Ex-Post* Case, because, on this view, we should maximize Alice's expectation of lifetime well-being by going down. But, to see how *Ex-Ante* Weak Lifetime Negative Utilitarianism can lower the expectations of lifetime well-being for everyone, consider the following sequential case that only affects Alice and Bob:

²⁷ There is an analogous problem for *Ex-Post* and *Ex-Ante* Prioritarianism; see Gustafsson 2022a.

The Ex-Ante Case



In this case, an initial chance node determines, depending on a fair coin, whether you will face the choice at the choice node 1 or the choice at the choice node 2. If you go up at either of the choice nodes, everyone gets a lifetime well-being of 0. If you go down at choice node 1, Alice gets a lifetime well-being of $1 + \epsilon$ (once again, ϵ is a small positive constant) and Bob gets a lifetime well-being of -1 . If you go down at choice node 2, Alice gets a lifetime well-being of -1 and Bob gets a lifetime well-being of $1 + \epsilon$.

Given that ϵ is sufficiently small, the adjusted moral value of one of Alice and Bob getting a lifetime well-being of -1 will outweigh the moral value of the other getting a lifetime well-being of $1 + \epsilon$. So, at either choice node, the sum total of adjusted expectation for the outcome of going down will be negative but the sum total of adjusted expectation for the outcome of going up will be zero. Hence, according to *Ex-Ante* Weak Lifetime Negative Utilitarianism, you should go up at each choice node (these prescriptions are represented by the thicker lines).

These choices (that is, going up at each choice node), which you would make if you follow *Ex-Ante* Weak Lifetime Negative Utilitarianism would, at the initial chance node, give everyone an expected lifetime well-being of 0. But consider the opposite choices — that is, going down at each choice node. These opposite choices would, at the initial chance node, give everyone an expected lifetime well-being of $\epsilon/2$. Hence following *Ex-Ante* Weak Lifetime Negative Utilitarianism in the *Ex-Ante* Case lowers everyone's expectation of lifetime well-being.

It may be objected that we could avoid this result with a resolute variant of *Ex-Ante* Weak Lifetime Negative Utilitarianism which would always calculate peoples expectations from the perspective of a privileged node, following the model of resolute choice in decision theory.²⁸ If we let

²⁸ See McClellan 1990, p. 13. And see Steele 2018, p. 662 and Gustafsson 2022b for

the privileged node be the initial chance node in the *Ex-Ante* Case, this resolute variant would prescribe going down at each choice node.

But, to calculate expectations at the choice nodes in this resolute manner from the perspective of a node in the past, seems to depart from the consequentialist part of Negative Utilitarianism. A compelling tenet of consequentialism is that only the consequences of the currently available options matter. The consequences that we could have realized only if earlier chance and choice events had resolved differently lack moral relevance. Furthermore, it is far from clear what node should serve as the privileged node. In the decision tree for a toy example like the *Ex-Ante* Case, it may seem that there is an obvious candidate: the initial node of the tree. But, in a real life case, there are lots of different decision trees that could have been used to model the situation with different initial nodes. Moreover, some current people who you can affect may not have come into existence at the time of the privileged node in the past. And, if so, they need not have any expectation of lifetime well-being from the perspective of the privileged node unless existence can be compared with non-existence, which is doubtful.²⁹

I wish to thank Gustav Alexandrie, Krister Bykvist, Roger Crisp, Anthony DiGiovanni, Tomi Francis, Simon Knutsson, Andreas Mogensen, Toby Ord, Wlodek Rabinowicz, Dean Spears, Brian Tomasik, Philip Trammell, and Magnus Vinding for valuable comments.

some objections to resolute choice.

²⁹ Gustafsson 2018, p. 602.

References

- Arrhenius, Gustaf and Krister Bykvist (1995) *Future Generations and Interpersonal Compensations: Moral Aspects of Energy Use*, Uppsala: Uppsala University.
- Atkinson, Anthony B. and Joseph E. Stiglitz (1980) *Lectures on Public Economics*, New York: McGraw-Hill.
- Bergström, Lars (1978) 'Pessimismens konsekvenser', in Lars Bergström, Harald Ofstad, and Dag Prawitz, eds., *En filosofibok tillägnad Anders Wedberg*, pp. 24–34, Stockholm: Bonniers.
- Broome, John (1991) *Weighing Goods: Equality, Uncertainty and Time*, Oxford: Blackwell.
- Fehige, Christoph (1998) 'A Pareto Principle for Possible People', in Christoph Fehige and Ulla Wessels, eds., *Preferences*, pp. 508–543, Berlin: de Gruyter.
- Gandjour, Afschin and Karl Wilhelm Lauterbach (2003) 'Utilitarian Theories Reconsidered: Common Misconceptions, More Recent Developments, and Health Policy Implications', *Health Care Analysis* 11 (3): 229–244.
- Gustafsson, Johan E. (2018) 'The Difference Principle Would Not Be Chosen behind the Veil of Ignorance', *The Journal of Philosophy* 115 (11): 588–604.
- (2019) 'Population Axiology and the Possibility of a Fourth Category of Absolute Value', *Economics and Philosophy*, pp. 1–30.
- (2022a) 'Ex-Ante Prioritarianism Violates Sequential Ex-Ante Pareto', *Utilitas* 34 (2): 167–177.
- (2022b) *Money-Pump Arguments*, Cambridge: Cambridge University Press.
- Hart, H. L. A. (1980) 'Death and Utility', *The New York Review of Books* 27 (8): 25–32.
- Hedenius, Ingemar (1955) *Fyra Dygder*, Stockholm: Bonniers.
- Hiż, Henry (1992) 'Praxiology, Society and Ethics', in J. Lee Auspitz, Wojciech W. Gasparski, Marek K. Mlicki, and Klemens Szaniawski, eds., *Praxiologies and the Philosophy of Economics*, pp. 421–429, London: Routledge.
- Hurka, T. M. (1982) 'Average Utilitarianism', *Analysis* 42 (2): 65–69.
- Klocksiam, Justin (2016) 'How to Accept the Transitivity of *Better Than*', *Philosophical Studies* 173 (5): 1309–1334.
- Knutsson, Simon (2021) 'The World Destruction Argument', *Inquiry* 64

- (10): 1004–1023.
- Locke, Don (1987) ‘The Parfit Population Problem’, *Philosophy* 62 (240): 131–157.
- MacAskill, William (2019) ‘Practical Ethics Given Moral Uncertainty’, *Utilitas* 31 (3): 231–245.
- Mathiason, Eric (2018) *Asymmetries and Ill-Being*, Ph.D. thesis, University of Toronto.
- McClennen, Edward F. (1990) *Rationality and Dynamic Choice: Foundational Explorations*, Cambridge: Cambridge University Press.
- McMahan, Jeff (2013) ‘Causing People to Exist and Saving People’s Lives’, *Journal of Ethics* 17 (1/2): 5–35.
- Metzinger, Thomas (2003) *Being No One: The Self-Model Theory of Subjectivity*, Cambridge, MA: MIT Press.
- Ord, Toby (2013) ‘Why I’m Not a Negative Utilitarian’, <http://www.amirrorclear.net/academic/ideas/negative-utilitarianism>.
- Parfit, Derek (1971) ‘Personal Identity’, *The Philosophical Review* 80 (1): 3–27.
- (1986) ‘Overpopulation and the Quality of Life’, in Peter Singer, ed., *Applied Ethics*, pp. 145–164, Oxford: Oxford University Press.
- (1995) *Equality or Priority?*, Lawrence: University of Kansas.
- Pearce, David (2005) ‘The Pinprick Argument’, <https://www.utilitarianism.com/pinprick-argument.html>.
- Popper, K. R. (1945) *The Open Society and Its Enemies Volume I: The Spell of Plato*, London: Routledge.
- (1962) *The Open Society and Its Enemies Volume II: The High Tide of Prophecy: Hegel, Marx, and the Aftermath*, London: Routledge, fourth edn.
- Santayana, George (2011) *The Works of George Santayana Volume VII, Book One: The Life of Reason or the Phases of Human Progress: Introduction and Reason in Common Sense*, Cambridge, MA: MIT Press.
- Schopenhauer, Arthur (2018) *The World as Will and Representation Volume 2*, eds. Judith Norman, Alistair Welchman, and Christopher Janaway, Cambridge: Cambridge University Press.
- Sidgwick, Henry (1907) *The Methods of Ethics*, London: Macmillan, seventh edn.
- Singer, Peter (1979) *Practical Ethics*, Cambridge: Cambridge University Press.
- Smart, J. J. C. (1989) ‘Negative Utilitarianism’, in Fred D’Agostino and I. C. Jarvie, eds., *Freedom and Rationality: Essays in Honor of John Watkins*,

pp. 35–46, Dordrecht: Kluwer.

Smart, R. N. (1958) ‘Negative Utilitarianism’, *Mind* 67 (268): 542–543.

Steele, Katie (2018) ‘Dynamic Decision Theory’, in Sven Ove Hansson and Vincent F. Hendricks, eds., *Introduction to Formal Philosophy*, pp. 657–667, Berlin: Springer.

Walker, A. D. M. (1974) ‘Negative Utilitarianism’, *Mind* 83 (331): 424–428.

Wolf, Clark (1996) ‘Social Choice and Normative Population Theory: A Person Affecting Solution to Parfit’s Mere Addition Paradox’, *Philosophical Studies* 81 (2/3): 263–282.