

# What Is the Difference Between Weak Negative and Non-Negative Ethical Views?

SIMON KNUTSSON  
Foundational Research Institute  
simon.knutsson@foundational-research.org

June 2016

## Abstract

Weak negative views in ethics are concerned with both reducing suffering and promoting happiness, but are commonly said to give *more* weight to suffering than to happiness. Such views include weak negative utilitarianism (also called negative-leaning utilitarianism), and other views. In contrast, non-negative views, including traditional utilitarianism, are typically said to give *equal* weight to happiness and suffering. However, it is not obvious how weak negative and non-negative views differ, and what it means to give happiness and suffering equal weight, or to give suffering more weight.

## Contents

1	Introduction	2
2	Terminology	3
3	The views under consideration	3
4	Weak negative utilitarianism and consequentialism	3
5	The Weak Negative Doctrine	4
6	Weak negative axiology	4
7	The difference between weak negative and non-negative views assuming happiness and suffering are objectively highly measurable	5
8	If statements about magnitudes of happiness versus suffering involve value judgements	7
9	Weak negative and non-negative views are equally affected	10
10	When is it accurate to describe non-negative and weak negative views as symmetric vs. asymmetric?	10

<b>Acknowledgments</b>	<b>12</b>
<b>References</b>	<b>12</b>
<b>Appendix: Value for an individual versus value for the world—where does the negativity set in?</b>	<b>12</b>

**List of Figures**

1	Common illustrations of non-negative and negative views, showing the value and disvalue of an individual’s happiness and suffering at a point in time. The illustrations assume that happiness and suffering are factual attributes measurable on the same ratio scale . . . . .	6
2	If happiness and suffering are measurable on <i>separate</i> ratio scales . . . . .	8
3	Illustration of views from figures 1a, 1b and 1c, but assuming that statements about magnitudes of happiness versus suffering are colored by value judgements.	9

**1 Introduction**

If happiness and suffering were factual attributes measurable on one [ratio scale](#), one way to distinguish the views would be straightforward: weak negative views could give greater moral or evaluative weight to one factual unit of suffering than one factual unit of happiness, while non-negative views could give them equal weight. It would then be accurate to call such weak negative views ‘asymmetric’ and such non-negative views ‘symmetric.’ However, such measurability is [highly controversial](#), and if it fails to hold, there is question of how one should understand the difference between weak negative and non-negative views. It has been suggested that this would be a problem for weak negative views in particular, but if it is a problem, it is as much of a problem for non-negative views.

There are several reasons why such measurability could fail to hold, and how one should understand the difference between the views depends on the details. One plausible way in

which it could fail to hold is if a person’s assessments of magnitudes of happiness and suffering are colored by her values. One could then not necessarily distinguish the views in terms of asymmetric versus symmetric and in terms of the weight they give to suffering. The reason is that there would no longer be a clear division between factual units of happiness and suffering and the value (weight) that one assigns to them. One could then understand some disagreements among people identifying as weakly negative versus non-negative in, for example, the following way: the person with the “weak negative” view can say, “When we both observe the same suffering, I judge it more severe than you do. Similarly, when we consider which individual is as happy as another is miserable, I would point to a happier individual than you would.”

One should generally not describe the difference between, for instance, a person who identifies as a weak negative or negative-leaning utilitarian, and a person who identifies as a traditional utilitarian, by saying that one gives

more weight to suffering than happiness, while the other gives them equal weight. That description can be accurate but need not be.

## 2 Terminology

By ‘happiness,’ I mean roughly to feel good, and by ‘suffering,’ I mean roughly to feel bad. So when I say ‘happiness’ and ‘suffering,’ one could roughly replace those words with ‘pleasure’ and ‘displeasure.’ Happiness and suffering are some of the things that have been conceived of as *finally* good and bad for an individual (i.e. good and bad for their own sake, as ends rather than as means). In other words, these things, and many more, have been claimed to be the facts that constitute an individual’s ‘well-being.’ How high or low someone’s level of well-being is refers to how good or bad life is for the individual who leads it. Although I focus on happiness and suffering, what I say in this essay generally applies to whatever is taken to constitute well-being, whether it is happiness and suffering, or other facts such as preference satisfaction and preference frustration, achievement or friendship.<sup>1</sup>

## 3 The views under consideration

I focus on three kinds of weak negative views: (1) weak negative utilitarianism (sometimes called negative-leaning utilitarianism) and weak negative consequentialism, (2) what

Griffin (1979) calls the Weak Negative Doctrine, and (3) weak negative axiological views. These views are called weak as opposed to strong because they give weight to *both* happiness and suffering. Strong negative views are typically conceived of as giving weight to *only* suffering.

When describing the weak negative views in this section, I talk, as is commonly done, about giving ‘more weight’ to suffering than to happiness, that promoting happiness is a ‘slighter obligation’ than eliminating suffering, and that suffering ‘matters more’ value-wise. This is to explain which views I refer to, but such talk is tentative in that I argue that one should often not think of such views as giving ‘more weight’ or the like to suffering than to happiness.

## 4 Weak negative utilitarianism and consequentialism

By ‘weak negative utilitarianism,’ I mean versions of utilitarianism that have been said to give more weight to suffering than to happiness, while still giving weight to both. I take it that versions of weak negative utilitarianism include a wide range of views that can span the whole range from caring almost only about reducing suffering to caring almost as much about happiness as about suffering.<sup>2</sup> Weak negative utilitarianism is, roughly speaking, also known as negative-leaning utilitarianism.<sup>3</sup> In contrast, traditional utilitarianism is com-

---

<sup>1</sup>More on this kind of terminology can be found [here](#).

<sup>2</sup>I broadly follow the understanding of weak and strong negative utilitarianism in Arrhenius and Bykvist (1995).

<sup>3</sup>A difference is that my use of weak negative utilitarianism is more inclusive than Brian Tomasik’s use of the term negative-leaning utilitarianism. According to Tomasik (2013), negative-leaning utilitarianism says that “suffering deserves *vastly greater* weight than happiness” (my emphasis). Similarly, Ord (2013) describes negative utilitarianism as saying that “positive experiences such as pleasure or happiness are either given no weight, *or at least a lot less weight*” (my emphasis). I use the term ‘weak negative utilitarianism’ somewhat differently from Ord and from how Tomasik uses ‘negative-leaning utilitarianism’ in that I would say that weak negative utilitarianism need not assign *vastly greater* or *a lot* more weight to suffering than to happiness; simply assigning *more* weight to suffering is enough. However, this difference in use is not crucial for this essay.

<sup>4</sup>For example, see Arrhenius and Bykvist (1995). I avoid the phrase ‘Classical Utilitarianism’ because it appears too narrow. One can be a non-negative utilitarian, and be as concerned with reducing suffering as with

monly said to give equal weight to happiness and suffering.<sup>4</sup> In addition to weak negative utilitarianism, there are weak negative versions of consequentialism that are not utilitarian. For example, I would say that one is a weak negative consequentialist, but not a weak negative utilitarian, if the part of one's morality that concerns individuals' well-being has the same form as weak negative utilitarianism, while one also believes that there are other values besides well-being, such as beauty and ugliness, or justice and injustice.

## 5 The Weak Negative Doctrine

Griffin (1979) sets utilitarianism aside and instead talks of the Positive Doctrine and the Negative Doctrine, in order to "preserve proper distance from parochial issues of utilitarianism" and because the topic does not matter "only to utilitarians," but "to anyone who in any way gives weight to people's welfare."<sup>5</sup> The Positive Doctrine says that "we have equally strong obligations to promote happiness and to eliminate unhappiness," and analogously to weak negative utilitarianism, the Weak Negative Doctrine says that we have obligations both to promote happiness and to eliminate suffering, but that "promoting happiness is a *slighter* obligation than eliminating suffering."<sup>6</sup> How is the Weak Negative Doc-

trine different from weak negative utilitarianism? One difference is that utilitarianism, including weak negative utilitarianism, says that the *only* thing that ultimately matters morally is to bring about valuable outcomes (where value is understood only in terms of individuals' well-being). However, the Weak Negative Doctrine is compatible with the idea that other things also matter morally, such as equality, rights, intentions, virtue and deontological rules that prohibit certain actions.

## 6 Weak negative axiology

A weak negative axiological view would say that both happiness and suffering affect the value of lives or worlds, but that suffering has more disvalue than happiness has value. Weak negative utilitarianism, weak negative consequentialism and the Weak Negative Doctrine are all normative views that make claims about what is morally right and wrong, moral obligations and so on. Although they are closely connected to axiology ([value theory](#)), which deals with issues of good, bad, better, and so on, I include weak negative axiology as a separate category to make the point that one can separate the normative from the evaluative, and focus on what happiness versus suffering implies about the value of lives and outcomes without talking about utilitarianism,

---

promoting happiness, without endorsing all the ingredients of Classical Utilitarianism. According to Sinnott-Armstrong (2015), "classic utilitarianism is actually a complex combination of many distinct claims, including the following [eleven] claims about the moral rightness of acts." In this essay, I conceive of utilitarianism broadly. I am concerned with versions of utilitarianism (including traditional utilitarianism and weak negative utilitarianism) that are utilitarian in the sense of holding that the only thing that matters morally is the value of outcomes, that such value should be assessed impersonally, that the value of outcomes only depends on individuals' well-being, and that distribution of well-being (e.g. equality) is morally irrelevant. I do not require that a utilitarian view considers the value of outcomes to depend on the sum of well-being (or the sum of happiness minus the sum of suffering). As I see utilitarianism, one can be a utilitarian and believe that well-being is not additive. For example, see the discussion of ordinal utilitarianism [here](#).

<sup>5</sup>Page 47.

<sup>6</sup>Ibid. Similar to the way I used 'weak negative utilitarianism,' I would prefer to use the 'Weak Negative Doctrine' to refer to the full range of positions between the Strong Negative and the Positive Doctrine, but that is not crucial to this essay.

<sup>7</sup>For example, Hurka (2010) focuses on axiology and discusses, among other things, what he calls the Moore-Mayerfeld view. Hurka quotes G. E. Moore saying that "while pleasure has 'at most some slight intrinsic value,' pain is 'a great evil'." Hurka (2010, 199). Hurka quotes Jamie Mayerfeld saying that "suffering is more bad than

consequentialism or moral obligations.<sup>7</sup>

## 7 The difference between weak negative and non-negative views assuming happiness and suffering are objectively highly measurable

Non-negative versus weak negative views are commonly illustrated using graphs as shown in Fig. 1.<sup>8</sup>

The vertical axis in Fig. 1 says value and disvalue 'for the world,' which means that the figure shows how happiness and suffering affect the overall value of an outcome or a possible world.<sup>9</sup> Fig. 1a illustrates a non-negative view. A given magnitude  $x$  of happiness adds as much positive value to the world as the same magnitude  $-x$  of suffering adds negative value to the world. In other words, happiness is as valuable as suffering is disvaluable; they have equal evaluative weight. The other graphs in Fig. 1 illustrate different weak negative views, according to which suffering has more weight than happiness. Fig. 1b says that the value for the world of a given magnitude  $x$  of happiness is strictly less than the disvalue for the world of magnitude  $-x$  of suffering. In addition, value for the world rises linearly with the magnitude of happiness, and disvalue for the world increases linearly with the magnitude of suffering. Fig. 1c is similar but says that suffering has increasing marginal disvalue for the world and happiness has decreasing marginal value

for the world. That is, an increase in one unit of an individual's happiness adds less value to the world, the higher the individual's level of happiness already is. Vice versa for suffering: an increase in one unit of suffering adds more disvalue to the world the higher the level of suffering already is. Fig 1d is very similar to 1c but adds the idea that the marginal value of additional happiness approaches zero. That is, there is a limit (the dashed horizontal line) that the positive value for the world of happiness never reaches above, regardless of how high the level of happiness gets. Fig. 1e is similar to 1b but adds a kink to the graph in the lower-left quadrant, after which the slope becomes vertical. The kink illustrates the idea that there is some level of suffering where no level of happiness can outweigh or counterbalance it.<sup>10</sup>

What does happiness and suffering on the horizontal axis refer to more specifically? Fig. 1 is too unspecific as it stands. A common view is that the horizontal axis shows facts and that the vertical axis shows value. That is, that terms such as happiness, suffering, pleasure, displeasure, and preference satisfaction, are all factual or descriptive terms.<sup>11</sup> The following is an example of one way to specify the horizontal axis from a hedonistic perspective. Hedonism says that what is good and bad for individuals or the world is pleasure and displeasure, so strictly speaking, the labels 'happiness' and 'suffering' on the horizontal axis

---

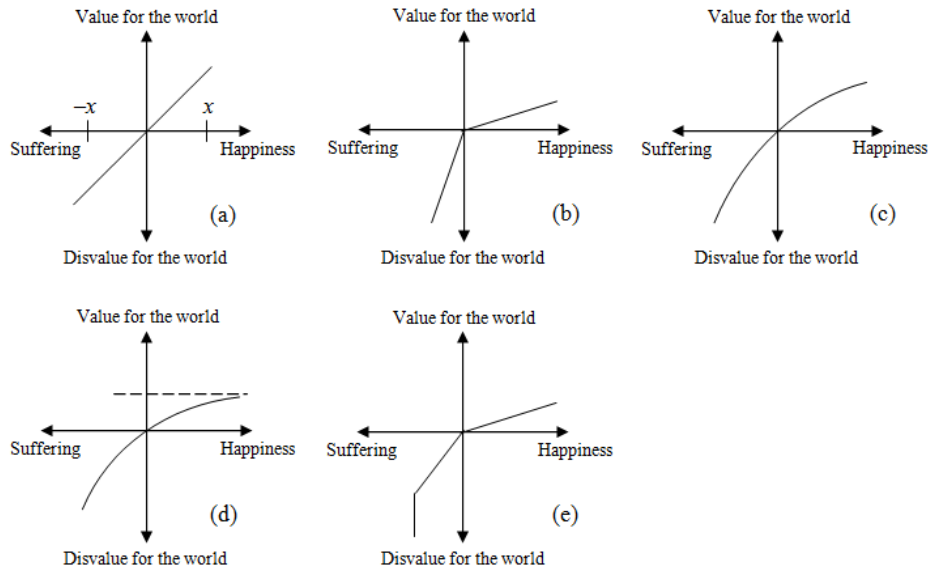
happiness is good" Hurka (2010, 200).

<sup>8</sup>Similar graphs can be found in Arrhenius and Bykvist (1995, 51), Brülde (2009), Hurka (2010), Ord (2013), and Tännsjö (2015). Those sources include more variants.

<sup>9</sup>Other terms than 'value for the world' seem to be used for the same thing in the philosophical literature, including 'value-period,' 'straightforward value,' 'value *simpliciter*,' 'absolute value,' 'ethical value,' 'contributive value' (as in the contribution that something makes to the value of a world), etc. See Brülde (1998, 9), Brülde (2009), and Fritzson (2014, 163-165). Others who have presented these kinds of graphs tend to put one of these terms, or a something term, on the vertical axis: Brülde (2009) puts 'contributive value' on the vertical axis, which is the same as value for the world. Hurka (2010) puts 'value,' by which he most likely means value for the world. For example, he speaks about the goodness and badness of worlds on page 200. Ord (2013) puts 'moral value,' which I take to be essentially the same as my value for the world. Tännsjö (2015) puts 'moral importance,' which, in his context, seems identical with or very similar to ethical value.

<sup>10</sup>This is one way to illustrate a weak threshold negative view.

<sup>11</sup>This is how Brülde (2009), Hurka (2010), and Tännsjö (2015) see the horizontal axis.



**Figure 1:** Common illustrations of non-negative and negative views, showing the value and disvalue of an individual's happiness and suffering at a point in time. The illustrations assume that happiness and suffering are factual attributes measurable on the same ratio scale

should be replaced with 'pleasure' and 'displeasure.' Then the horizontal axis would show units and magnitudes of pleasure and displeasure, which again, in this context are normally considered to be facts. So what is a unit of pleasure and displeasure? How could we say that an instance of pleasure is as pleasurable as an instance of displeasure is displeasurable? Several answers are possible; one is that a unit should be conceived of as a just noticeable difference.<sup>12</sup> If we specify the horizontal axis like this, the graphs show how much each unit of pleasure and displeasure, conceived of as perceivable increments, contribute to the value and disvalue of the world. The horizontal axis would need to be specified in that or some other way to be understandable.

Several controversial conditions are required to hold for Fig. 1 to work as an illustration of the difference between weak negative and non-negative views. The conditions include the following:

1. Happiness and suffering need in principle to be measurable on the same **ratio scale**. This means (a) that there needs to be one scale for both happiness and suffering, (b) that the scale needs to have a zero point, and (c) that magnitudes of happiness and suffering need to be able to be represented by numbers, for example 10 and  $-5$  such that an instance of happiness can be said to be, for instance, twice the magnitude of an instance of suffering.
2. Happiness and suffering need to be factual terms, and measurable, in principle, in an **objective**, non-evaluative, empirical, and factual way. This measurement would not involve value judgements on the part of the person assessing the magnitudes of happiness versus suffering (once the definition or conception of happiness and suffering is decided upon).
3. Happiness and suffering need to be *interpersonally* measurable, which means that

<sup>12</sup>The idea of smallest noticeable increments dates back at least to Edgeworth (1881). Torbjörn Tännsjö developed Edgeworth's idea. For example, see Tännsjö (2000). See [this section for more information](#).



they need to be comparable across persons. This condition is, strictly speaking, not necessary for Fig. 1 to illustrate what weak negative and non-negative views would say about the value of the happiness and suffering of *one individual at a point in time*. However, the views we are concerned with are typically applied to scenarios and situations involving several individuals.<sup>13</sup> So for Fig. 1 to illustrate what weak negative and non-negative views would say about the value of the happiness and suffering among *different individuals*, happiness and suffering would need to be interpersonally measurable (on the same ratio scale).

Fig. 2 below illustrates the first requirement, that happiness and suffering need to be measurable, in principle, on the same ratio scale. For instance, if they are *not* measurable on the *same* scale, but still measurable on *separate* ratio scales, we can only draw two separate graphs, like the following:

The slopes of the lines in Fig. 2a and Fig. 2b cannot be meaningfully compared because I have assumed for the purposes of the illustration that the units of suffering are on a different scale than the units of happiness, which is why the two figures are separate. This is not a problem in particular about formulating weak negative views—if happiness and suffering are not measurable on the same scale, we could

not draw any of the graphs in Fig 1, including the non-negative Fig. 1a.

The remainder of this essay will focus on the third requirement, that happiness and suffering are factual attributes that are (interpersonally) *objectively* measurable on the same ratio scale. I will consider what happens with the difference between the weak negative and the non-negative views if, as I find plausible, happiness and suffering are not objectively measurable (to the required degree); that is, if claims about magnitudes of happiness versus suffering are colored by value judgements.<sup>14</sup>

## 8 If statements about magnitudes of happiness versus suffering involve value judgements

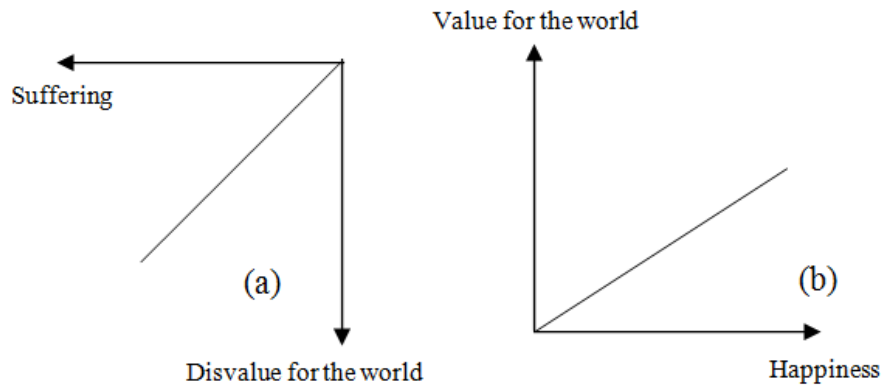
Griffin (1979) says that there are situations in which the judgement about amounts of happiness and suffering is "not independent of, not distinguishable from" one's preferences.<sup>15</sup> Relatedly, Hurka (2010) says,

Moore and Mayerfeld assume, as Bentham and Sidgwick also do, that we can compare the intensities of pleasures and pains independently of assessing their values. But an objector may challenge this assumption, saying the claim that a pain is more intense than some pleasure merely says the pain is more evil than the pleasure is good, without pointing to some independent psychological fact that makes it so.

<sup>13</sup>For example, Arrhenius and Bykvist (1995), Mayerfeld (1996), Mayerfeld (2000), and Tännsjö (2015).

<sup>14</sup>There are some interesting questions about how weak negative views can be distinguished from non-negative if happiness and suffering are objectively measurable in principle but not on the same ratio scale as Fig. 1 assumes. For example if happiness and suffering are measurable on *different* ratio scales, or only on the same interval or ordinal scale. For example, some have worked on ordinal versions of utilitarianism. See Mendola (1990) and Hardin (2009). I set these questions aside to focus on what happens with the distinction between weak negative and non-negative views if claims about magnitudes of happiness versus suffering involve value judgements.

<sup>15</sup>"Perhaps in many situations the judgement whether amounts of happiness and unhappiness are equal is not independent of, not distinguishable from, one's being indifferent to a situation in which they are combined. Similarly, perhaps in these cases the judgement whether a certain amount of happiness is just greater than a certain amount of unhappiness is not independent of the judgement that the point has just been reached where one would prefer a situation in which they are combined" (page 53). Further down on that page, he omits 'perhaps' and implies that he is suggesting that there are such situations where judgements about relative happiness and suffering are not independent from one's preferences.



**Figure 2:** *If happiness and suffering are measurable on separate ratio scales*

There is no such fact, she may argue, and thus no possibility of nonevaluative comparisons of pleasures and pains. If so, the difference between symmetry and asymmetry views about them [pleasures and pains] disappears.<sup>16</sup> This objection raises a more general issue. To formulate a pairwise asymmetry about any pair of values we must be able to compare these values non-evaluatively, or identify instances of them as equal in a way that is neutral about their comparative worth. Only then can their relative values be a further issue.<sup>17</sup>

I essentially endorse this ‘sceptical objection’ that Hurka states. What is the implication for the difference between weak negative and non-negative views? First, it could be pointless to draw graphs like those in Fig. 1 to distinguish weak negative from non-negative views. For example, all of figures 1a, 1b and 1c could be drawn as 1a. The reason is that the weak negative and non-negative views would set the scale on horizontal axis differently, based on different value judgements. In Fig.

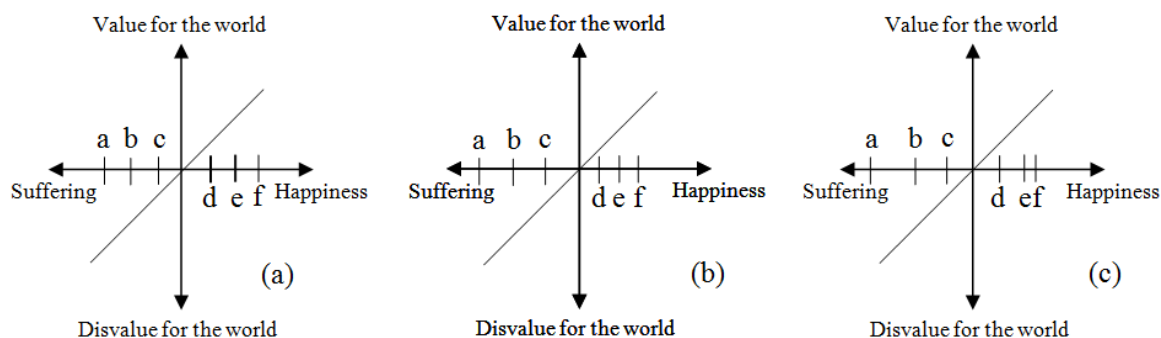
1a, there are marks  $x$  and  $-x$  on the horizontal axis that point to the happiness  $x$  that is of “the same magnitude” as the suffering  $-x$ . If assessments of magnitudes of happiness and suffering are colored by values, a person identifying with a weak negative view can say to someone identifying with a non-negative view, “Sure, Fig. 1a illustrates my view as well as yours, but when faced with the same factual situation, you and I make different value judgements about what magnitudes of happiness and suffering we see.”

Fig. 3 below illustrates my point. Say that a person identifying with a weak negative view and a person identifying with a non-negative view are to decide to what degree the following six individuals are happy and suffer: Ann, Ben, Chloe, Dan, Erica, and Fred. All six have different mental states and circumstances, and the two persons with the weak-negative and non-negative views are presented with the concrete facts about the six individuals’ mental states and circumstances. Assume that Ann, Ben, and Chloe are suffering, and that Ann is suffering most severely, Ben less severely and Chloe even less severely. Similarly, assume

<sup>16</sup>Here Hurka refers to Griffin (1979).

<sup>17</sup>Hurka (2010, 201–202). Tännsjö (2015) makes a similar point when discussing prioritarianism versus utilitarianism: “Now, if we want to keep utilitarianism and prioritarianism apart, regardless of how else we conceive of well-being, we had better see to it that well-being is an empirical (non-normative) notion. Otherwise, it is likely that, if there are prioritarian weights in the first place, then we have already factored them in, when we assess how well off an individual is at a certain moment.” (Page 241).





**Figure 3:** Illustration of views from figures 1a, 1b and 1c, but assuming that statements about magnitudes of happiness versus suffering are colored by value judgements.<sup>18</sup>

that Dan, Erica, and Fred are all happy, and Dan is the least happy, Erica is more and Fred is the most happy among them. <sup>18</sup> One can then illustrate the views from figures 1a, 1b and 1c as shown in Fig. 3.

In Fig. 3, all lines have the same slope so we cannot distinguish weak negative from non-negative views based on the slopes as we did in Fig. 1. Instead, the difference between the weak negative and the non-negative views is that they make different value judgments about the degree to which Ann, Ben, Chloe, Dan, Erica, and Fred are suffering and happy compared to one another. The view illustrated in Fig. 3a can be said to be less negative (or more optimistic) *compared* to the other two because it judges that Dan, Erica, and Fred are happier than what the views in figures 3b and 3c do. Similarly, the views illustrated in figures 3b and 3c can be said to be negative (or pessimistic) *compared* to the one in 3a, because they judge the same individuals to be suffering more and to be less happy than 3a does.

The difference between the views in Fig. 3 is not that one is symmetric and two asymmetric, nor that they give different weight or equal weight to a unit of suffering compared to

a unit of happiness. Rather, the difference is their value judgements about the degree of suffering in concrete instances of suffering compared to their judgements about the degree of happiness in concrete instances of happiness. When non-negative and weak negative views differ in this way, it is inaccurate to distinguish them in terms of symmetry versus asymmetry or in terms of giving more weight to suffering versus equal weight to suffering and happiness.

A person with a non-negative view might say that  $-x$  corresponds to how a depressed human feels who is about to commit suicide, and that  $x$  corresponds to the happy mental state of a researcher making an intellectual discovery that she worked a lifetime to reach. A person with a weak negative view might reply, "No, that happiness is not at all of 'equal magnitude' compared to that suffering. A much higher magnitude of happiness (feeling good) would be required to reach  $x$ , perhaps a heroin rush or a degree of happiness that no human has ever experienced." This disagreement would, at least partly, be a disagreement in values. In other words, because the person with the weak negative view and the person with the non-negative view would involve their different value judgements when

<sup>18</sup>Fig. 3 assumes for illustration purposes that the happiness and suffering shown on the horizontal axis are measurable on an interpersonally comparable ratio scale, but the key point in Fig. 3 is to illustrate the idea that assessments of magnitudes of happiness and suffering are colored by value judgements.

deciding which concrete mental states belong where on the horizontal axis, one could not distinguish their views with reference to, for example, the difference between figures 1a, 1b and 1c (i.e. not with reference to the idea that the views assign different values to given factual magnitudes of happiness and suffering).

So how should one understand the difference between weak negative and non-negative views if happiness and suffering are not objectively, non-evaluatively, measurable to the required degree? This is the topic of the rest of the essay, but I first want to emphasize that this is not a problem for weak negative views in particular; if it is a problem, it is just as much of a problem for non-negative views.

## 9 Weak negative and non-negative views are equally affected

In the quote from Hurka (2010) in the previous subsection, he says that comparing intensities (what I call magnitudes) of pleasures and pains non-evaluatively, in a way that does not involve value judgements, is required to formulate an evaluative *asymmetry* between pleasures and pains (and between happiness and suffering). However, this is also required to formulate a *symmetry* between them. This point may seem obvious, but I bring it up because at least Griffin (1979) seems to have considered this a problem for weak negative views in particular, as opposed to for non-negative views.<sup>19</sup> However, stating that a unit of hap-

piness has equal evaluative weight compared to a unit of suffering requires non-evaluative comparisons of units of happiness and suffering just as much as stating that suffering has more weight does. This is no more a problem for the negative views than for non-negative views, and perhaps it is not a problem at all, although it raises the question of how the difference between weak negative views and non-negative views should be understood.

In the next two sections, I describe how one can understand the difference between weak negative and non-negative views, assuming that statements about the relative magnitudes of happiness and suffering involve value judgements.

## 10 When is it accurate to describe non-negative and weak negative views as symmetric vs. asymmetric?

Even if statements about the relative magnitudes of happiness and suffering would involve value judgements, some believe that they do not. In that case, one can, at least at the principle level, distinguish their views similarly to how one would *if* happiness and suffering were objectively measurable to the required degree. For example, Mayerfeld explicitly states that he believes that happiness and suffering are objectively measurable on the same ratio scale and that "suffering is more bad than happiness is good."<sup>20</sup> Similarly Tännsjö says that

<sup>19</sup>Griffin (1979) sets out to answer the question "Is some form of the Negative Doctrine correct?" He concludes, "I know of no plausible form of the Negative Doctrine." However, he does not establish that the Weak Negative Doctrine is implausible, only that it is still unclear how it is to be distinguished from the positive doctrine. I here refer to his discussion of the CALCULUS version of the Weak Negative Doctrine and perhaps he has something similar in mind in his discussion of the BASIC SENTIMENT version. Ord (2013) perhaps sees a similar problem with weak negative utilitarianism in the section "Weak NU — The incoherence argument," in which he, among other things, says that "A major problem with Weak NU is that it appears to be incoherent. What could it mean for suffering to matter more than happiness? It suggests an image such as this: [Figure similar to Fig. 1b above] But what is the horizontal scale supposed to represent? There is no obvious natural unit of suffering or happiness to use. It might be possible to have a consistent scale in the happiness direction and a separate consistent scale in the suffering direction, but it is very unclear how they are both supposed to be on the same scale. This is what would be needed for Weak NU to be a coherent theory and for the diagram to make any sense."

<sup>20</sup>Mayerfeld (1999, 136), quoted in Hurka (2010, 200).

he believes in that degree of objective measurability and that each unit of pleasure is to count as *much* as one unit of displeasure.<sup>21</sup> In such cases, one can understand the difference in their views, at the principle level, as in Fig. 1, because that figure illustrates their 'views,' in the sense of their beliefs and their ways of thinking about these matters. One could accurately describe Mayerfeld's weak negative view as asymmetric and Tännsjö's non-negative view as symmetric.

In many cases, one should not understand the differences among weak negative versus non-negative views in terms of asymmetric versus symmetric or in terms of giving different weights to units of suffering than to units of happiness. The following are three such kinds of cases:

1. The person does not believe that happiness and suffering are factual attributes measurable to the required degree. He may, for instance, view his own statements about magnitudes of suffering and happiness as value judgements.
2. The person does not have a conception of what a magnitude or unit of happiness and suffering is, and so would not have a view on what it is that one is to assign equal or more weight to.<sup>22</sup>
3. Even if the person believes in objective measurability, in principle, and has a conception of units of happiness and suffering, she may not, or may not be able to, measure and quantify happiness and suffering in practice. For example, say that a person makes statements about the balance of happiness versus suffering in far-future scenarios or in a population of wild or domesticated animals; one can ques-

tion whether that person is really quantifying units of happiness and suffering and assigning them equal or different weight.

In such cases, how should one understand what the person is doing who is making statements about magnitudes of happiness versus suffering? One way is that they are making overall judgements of the goodness or badness of an outcome, without even approximately calculating and weighing units of happiness versus suffering.<sup>23</sup> Another way is that the person might assign numbers to magnitudes of happiness and suffering as *if* happiness and suffering were measurable to the required degree.<sup>24</sup>

How should one in such cases understand the disagreement between persons identifying with weak negative and non-negative views? Perhaps in terms of persons having different temperaments or different feelings about the situations that they contemplate, but not that they give different or equal weight to happiness versus suffering; nor that their views are necessarily asymmetrical or symmetrical. However, we can still speak of 'weak negative' and 'non-negative' as useful convenient labels to describe different persons' tendencies to make different judgements about how severe some instance of suffering is compared to some happiness (e.g. judgements like those illustrated in Fig. 3 above where six persons' relative happiness and suffering were assessed). This is how I usually think of those terms in conversations. This means that in these kinds of cases, the terms weak negative and non-negative are *comparative* in the sense that a person's view is negative compared to another person's view if the first person tends to see magnitudes of suffering and happiness less op-

---

<sup>21</sup>E.g. Tännsjö (2000).

<sup>22</sup>As mentioned above, one such conception of units of happiness (or pleasure) and suffering is that a unit is a smallest noticeable increment. One could have other conceptions, but one needs to have some conception of what a unit is to say that one is giving more or equal weight to a unit of suffering compared to a unit of happiness.

<sup>23</sup>For more discussion of such overall judgements, see [this section](#).

<sup>24</sup>As Brian Tomasik does. See [this section](#) for more.

timistically than the other person does when considering similar facts.

### Acknowledgements

I am grateful to Brian Tomasik and especially to Bengt Brülde for useful discussions, advice, and feedback on my work. Robin Raven did a nice job improving the language of this text.

### References

- Arrhenius, G. and Bykvist, K. (1995). Future generations and interpersonal compensations: Moral aspects of energy use.
- Brülde, B. (1998). The human good [phd diss]. *Gothenburg: Gothenburg university.*
- Brülde, B. (2009). *Lyckans och lidandets etik.* Thales,.
- Fritzson, F.-A. (2014). *Value Grounded on Attitudes. Subjectivism in Value Theory.* Lund University (Media-Tryck).
- Griffin, J. (1979). Is unhappiness morally more important than happiness? *The Philosophical Quarterly (1950-)*, 29(114):47–55.
- Hardin, R. (2009). Utilitarian aggregation. *Social Philosophy and Policy*, 26(01):30–47.
- Hurka, T. (2010). Asymmetries in value. *Noûs*, 44(2):199–223.
- Mayerfeld, J. (1996). The moral asymmetry of happiness and suffering. *The Southern journal of philosophy*, 34(3):317–338.
- Mayerfeld, J. (1999). *Suffering and Moral Responsibility.* Oxford University Press.
- Mendola, J. (1990). An ordinal modification of classical utilitarianism. *Erkenntnis*, 33(1):73–88.
- Ord, T. (2013). Why i’m not a negative utilitarian. *University of Oxford, published*

online at URL <http://www.amirrorclear.net/academic/ideas/negativeutilitarianism/index.html>. Date of retrieval, 20(2):2014.

- Sinnott-Armstrong, W. (2015). Consequentialism.
- Tannsjo, T. (2000). The least sub-noticeable difference. In *Imperceptible Harms and Benefits*, pages 75–93. Springer.
- Tännsjö, T. (2015). Utilitarianism or prioritarianism? *Utilitas*, 27(02):240–250.
- Tomasik, B. (2013). Three Types of Negative Utilitarianism.

### Appendix: Value for an individual versus value for the world—where does the negativity set in?

Let us assume the framework of Fig. 1, that happiness and suffering are factual attributes measurable on a ratio scale. The vertical axis in Fig. 1 says ‘value for the world’ and ‘disvalue for the world.’ Other terms for such value include ‘contributive value’ and ‘value-period.’ One can make a distinction between such value and value for individuals, which is sometimes called ‘prudential value’ or ‘value-for.’ If happiness has positive value for individuals, it contributes positively to their well-being (happiness makes their lives better). In contrast, if happiness has value for the world, it makes the world (or the overall outcome) better. This value for the world is what consequentialism says we should bring about.

One can discuss these two ways in which something can have value separately: (i) How does happiness and suffering affect an individual’s well-being? (ii) How does well-being affect the value of the world? One can get to, for example, Fig. 1b above in two steps: the first is from happiness and suffering to well-being, and the second is from well-being to value and

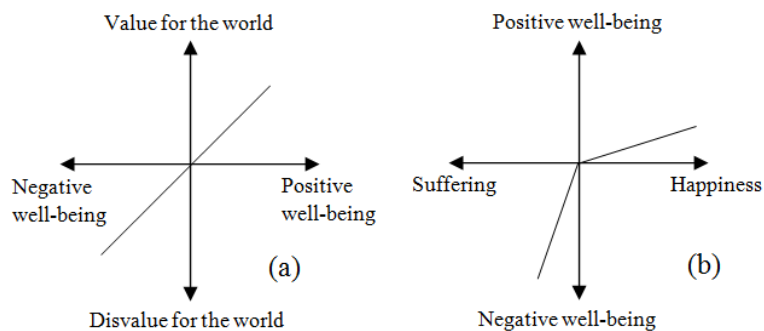


Figure 4

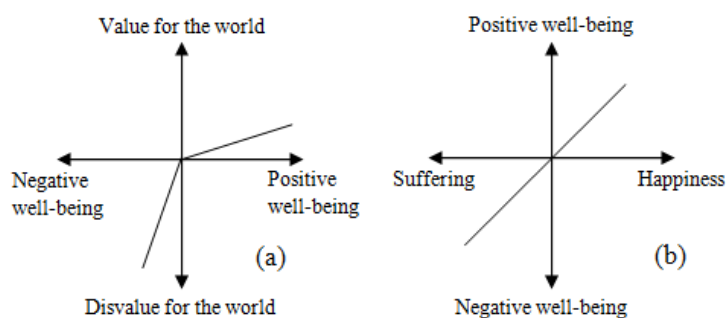


Figure 5

disvalue for the world. There is an interesting question of at which "step" the negativity "sets in." Fig. 4 illustrates the view that the negativity sets in during the step from happiness and suffering to well-being. Fig. 4a says that it is as valuable for the world if the individual's life is good for him (if his well-being is positive) at a moment as it is disvaluable for the world if his life is bad for him at that moment (if his well-being is negative). Fig. 4b says that happiness is less good for an individual than suffering is bad for an individual. Combining the two steps in figures 4a and 4b results in Fig. 1b. Alternatively, one can arrive at Fig. 1b in two steps in a different way, namely, as Fig. 5 illustrates.

Fig. 5b says that happiness is as good for an

individual as suffering is bad for an individual. In contrast, Fig. 5a says that it matters more for the value of the world if an individual's life is bad for him at a moment than if his life is good for him at that moment. In other words, happiness and suffering are equally good and bad for you, but it is more important for the value of the world that you are not suffering. Again, by combining Fig. 5a and Fig. 5b we end up with Fig 1b, but the content of the view is different compared to if we get to Fig. 1b by combining figures 4a and 4b. Similarly, one can arrive at, for example, Fig. 1c in two steps in these two alternative ways, which represent two different approaches to where the negativity sets in.