

Welfare Subjects

Philosophical debates about welfare or well-being have traditionally focused on the relative merits of different theories of well-being, such as hedonism, the desire theory, and the objective list theory. In recent years, however, philosophers have increasingly discussed which entities are welfare subjects and which of them each theory applies to: there are now debates, for example, about whether the same theory is true of all welfare subjects¹ and whether the class of welfare subjects extends beyond that of sentient beings.² These recent debates have occurred in the absence of a consensus about the definition of the expression in terms of which they are framed. While all parties agree that ‘welfare subject’ expresses the concept of an entity that is in some sense capable of well-being, there are two *prima facie* plausible ways to make this meaning precise, each with some adherents in the literature. I will argue that, because one of these definitions cannot plausibly accommodate certain comparative judgments about welfare, and because it leaves us ill-equipped to explain which entities are welfare subjects, we should prefer the other definition. In doing so, I will also explain why the claim that even inanimate objects can have a level of welfare is, on reflection, plausible.

1. Two Definitions

To my knowledge, the expression ‘welfare subject’ was introduced by Wayne Sumner, who claims, in *Welfare, Happiness, and Ethics*, that although it is clear that children and cats are welfare subjects, it is unclear whether paramecia or plants are and clear that inanimate objects aren’t.³ His explanation of why the term applies to cats sheds light on what he means by it: “It is perfectly natural for me to say that my cat is doing well, that having an ear infection is bad for her, that she has benefited from a change of diet, and so on. In making these judgements it certainly seems to me that I am applying exactly the same concept of welfare to my cat that I habitually apply to my friends.”⁴ Although he never defines the expression, he does eventually gloss it as meaning “an individual with a distinct welfare.”⁵ It appears, then, that he would define a welfare subject either as an entity with a welfare, one to which the concept of welfare applies, one that can be benefited or harmed, one that can be

¹ Lin (2018).

² Bradford (2023); Lee (forthcoming).

³ Sumner (1996), pp. 15-16, 22.

⁴ Sumner (1996), p. 14.

⁵ Sumner (1996), p. 43.

doing well or badly, or one for which things can be good or bad in the sense that is conceptually tied to well-being. Theorists who have adopted the term from Sumner have followed him in using it with this meaning or constellation of meanings in mind.

There are at least two *prima facie* plausible ways of making this meaning more precise. On the first, the term distinguishes the entities that are capable of faring well or badly from those incapable of doing either. If we adopt the convenient and standard convention of postulating amounts or levels of well-being that can be represented by real numbers, we can state this definition as follows:

Narrow Definition S is a welfare subject \equiv_{df} S is capable of having a positive or negative amount of well-being.

Eden Lin has explicitly endorsed this definition.⁶ Gwen Bradford seems sympathetic to it, too, when she glosses welfare subjects as “bearers of welfare goods and bads.”⁷ After all, barring the baroque and implausible possibility of an entity who cannot have any good without simultaneously having a bad that is at least as bad for it as the good is good for it, to be able to have a welfare good is to be able to be positive in well-being. And barring the opposite possibility, to be able to have a welfare bad is to be able to be negative in well-being. If Bradford dismisses these possibilities, as I think she surely does, then she should be receptive to the narrow definition.⁸

On the second way of precisifying the meaning of ‘welfare subject’, the expression distinguishes the entities that are capable of having an amount of well-being from those incapable of having any such amount—not even a neutral or zero amount:

Broad Definition S is a welfare subject \equiv_{df} S is capable of having an amount of well-being (even if only a neutral or zero amount).

Uriah Kriegel accepts this definition on the grounds that zombies, which he believes are not welfare subjects, differ importantly from what he calls mombies: people with “full and stormy experiential [lives] the net experiential value of [each of] which happens to be exactly zero.” Mombies qualify as

⁶ Lin (2018), p. 322.

⁷ Bradford (2023), p. 906.

⁸ I will revisit these possibilities in section 4.7.

welfare subjects on both definitions because they are positive in welfare at some times and negative in welfare at others even though each of them has a lifetime welfare of zero. Nevertheless, Kriegel thinks that contrasting mombies with zombies motivates the broad definition by showing that we must distinguish having a zero welfare level from having no welfare level at all. The life of an entity that isn't a welfare subject (e.g., a zombie) is not worth living in a deeper sense than a mombie's is, he claims. “The value of a mombie's life to the mombie is zero; but the zombie's life has *no* value to the zombie, not even the value 0—the zombie's wellbeing value is rather *undefined*.”⁹ Willem van der Deijl, who also contrasts having a zero welfare level with having no welfare level, endorses the broad definition when he equates the task of identifying the welfare subjects with that of “say[ing] to which individuals we can ascribe levels of wellbeing.”¹⁰ Andrew Lee also endorses this definition: he claims that since “having welfare level zero is distinct from lacking a welfare level altogether,” there can be welfare subjects “that cannot accrue any welfare goods/bads” and therefore cannot be positive or negative in well-being.¹¹

Since the broad definition assumes that having a zero amount of well-being is distinct from having none at all, we cannot assess it without understanding the difference between these two conditions. There are many contexts in which to say that something has no F is just to say that it has zero F . If I say that I felt no pain during the past minute and that I received no letters yesterday, what I mean is that I felt zero pain and received zero letters: after all, I am saying something that implies that I felt less pain, and received fewer letters, than I would have if I had felt any positive amount of pain or received any positive number of letters. In other contexts, however, it is important to distinguish having no F at all from having a zero amount of F . To say that the number 7 has no elevation is not to say that it has an elevation of zero, for that would imply that it has an elevation roughly one mile lower than that of Denver. Instead, it is to say that it is nowhere on the scale of elevations—not even at the zero point—so that it cannot be compared elevation-wise with anything. Likewise, to say that the color red has no pitch is not to locate it at a neutral or zero point on the scale of pitches that

⁹ Kriegel (forthcoming), pp. 9-10, 16.

¹⁰ Deijl (2021), pp. 189, 194.

¹¹ Lee (2022), p. 208n5. Lee (forthcoming) assumes a variant on the broad definition that appeals to the possession of welfare levels rather than the capacity to possess them: ‘welfare subject’, on this view, means ‘entity that has an amount of well-being (even if only a neutral or zero amount)’. Because it is unclear that an entity could be capable of having a welfare level without actually having one, it is unclear that the two versions of the broad definition would classify any cases differently. In any case, the difference between the two versions makes no difference to anything that Lee says or to anything that I will say. Since my formulation better parallels the narrow definition in structure, I will focus on it, and I will proceed as though Lee accepts it.

can be compared to the one occupied, say, by high C : it is to deny that it occupies any point on that scale, and thus to deny that it can be compared pitch-wise with anything. As these examples show, one crucial difference between occupying the zero point on a scale and occupying no point on it at all is that, in the former case but not in the latter, one can enter into comparisons with respect to that scale. An entity can have more, less, or the same amount of F as another entity does only if it has *some* amount (even if only zero) of F . One point on a scale can be compared only with another point on it: it cannot be compared with nothing at all. Thus, if an entity not only isn't positive or negative in welfare but has no amount of welfare at all, not even zero, then it cannot be compared with respect to welfare with anything: it cannot be better or worse off than, or equally well off as, anything. For these reasons, the narrow and broad definitions disagree about whether entities that are not welfare subjects can enter into welfare comparisons. Whereas the broad definition says that they cannot because they have no amounts of well-being, the narrow definition says that whether they can depends on whether they can have a zero amount of well-being.

Lifetime amounts of well-being, which represent how well entire lives have gone for their subjects, can be distinguished from amounts of well-being during shorter intervals of time, which represent how well lives are going for their subjects during such intervals. As I implied earlier when I said that zombies are welfare subjects even according to the narrow definition, I take both definitions to be concerned with any amounts of well-being, lifetime or otherwise. The narrow definition defines a welfare subject as an entity capable of being positive or negative in welfare, either on a lifetime basis or during an interval shorter than its entire life. The broad definition defines a welfare subject as an entity capable of having some amount of well-being (even if only zero), either on a lifetime basis or during an interval shorter than its entire life.

I understand both definitions to be employing a notion of capacity that is more robust than mere metaphysical or even nomological possibility—one grounded in the relevant entities' psychological or physical constitutions. You are incapable, in this sense, of flying without mechanical assistance even if there is a possible world at which you have grown and learned how to use wings. Thus, for example, if hedonism is the correct theory of welfare, then the narrow definition accords the status of welfare subjects only to entities whose constitutions enable them to feel pleasure or pain: if an entity is not so constituted but there is a possible world where it is, then it is not a welfare subject even though there is a possible world where it is one. Neither of the definitions takes a stance on

whether the capacity that it invokes can be had contingently. On each definition, whether welfare subjecthood can be had contingently turns on whether the relevant capacity can be had contingently.

Since ‘welfare subject’ is a technical term rather than an expression from ordinary language, we can’t adjudicate between the narrow and broad definitions on the basis of how well they fit pretheoretical intuitions about its application. Instead, we must ask whether either definition affords us theoretical advantages that the other does not. I will present two independent reasons of this sort to prefer the narrow definition.

2. Welfare Comparisons with Inanimate Objects

The first reason we should prefer the narrow definition to the broad one is that, unlike the former, the latter cannot accommodate certain comparative judgments about well-being without so greatly enlarging the extension of ‘welfare subject’ that the term ceases to be of much use.

Imagine a paradigmatically excellent life, overflowing with all of the basic goods postulated by your preferred theory of welfare and nearly devoid of any of the basic bads postulated by that theory, and compare that life to the existence of a rock. Surely, that life goes better for the person living it than the rock’s existence does for the rock. Now, compare the rock’s existence with a paradigmatically terrible life, one of uninterrupted torture and agony. Surely, this life goes worse for the person living it than the rock’s existence does for the rock. If you doubt these comparative judgments, perhaps because they might involve a category mistake, contrast them with the judgments that Denver has a higher elevation than the number 7 and that high C has a higher pitch than the color red. The latter judgments plainly do involve a category mistake, and we feel no inclination to accept them. But the former judgments are much more plausible and do not feel absurd in the way that the latter ones do. Moreover, the following judgments are also intuitively appealing:

- (1) There is a larger difference in welfare between the excellent life and the terrible life than there is between either of these lives and the existence of the rock.
- (2) There is a smaller difference in welfare between the excellent life and an otherwise exactly similar life that contains slightly less pleasure than there is between either of these lives and the existence of the rock.

- (3) There is a smaller difference in welfare between the terrible life and an otherwise exactly similar life that contains slightly less pain than there is between either of these lives and the existence of the rock.

But clearly, these judgments cannot be true unless a person can be better or worse off than a rock. If there couldn't be any differences in welfare between a person and a rock, then we couldn't compare the size of a difference in welfare between a person and a rock with that of a difference in welfare between two people. Because it is intuitively plausible that we can compare such sizes, it is plausible that a person can be better or worse off than a rock. Once again, it is instructive to contrast this case with one in which the relevant comparisons clearly are impossible. It is not plausible at all that there is a smaller difference in elevation between Denver and Salt Lake City than there is between either of these cities and the number 7. The absurdity of this claim, and of any other comparative claim concerning the size of the difference in elevation between anything and the number 7, corroborates the fact that the number 7 cannot be compared with anything with respect to elevation.

Some might suspect that this argument begs the question, since each of (1) – (3) obviously entails its conclusion and is thus a premise that no one who is committed to rejecting that conclusion would be willing to accept. But the fact that an argument employs such a premise doesn't make it question-begging, at least not in any objectionable sense. Every obviously valid argument has premises that together obviously entail its conclusion and whose conjunction would thus be rejected by anyone committed to rejecting its conclusion. This doesn't mean that a valid argument can avoid begging the question in an objectionable way only by failing to be obviously valid. For example, the claim that a good life spent in the real world is better than an experientially indistinguishable life spent in an experience machine obviously entails the falsity of standard quantitative hedonism. But it isn't objectionably question-begging to reject that theory because one accepts that claim.¹² An argument can provide evidence for its conclusion—it can provide good reasons for accepting it—even if it is predictably unable to convince people who are committed to rejecting its conclusion. Someone who is unsure whether people can be better or worse off than rocks, and even someone initially inclined to think that they cannot be, might realize on reflection that (1) – (3) are plausible and conclude on this basis that people can indeed be better or worse off than rocks. There would be nothing amiss about reasoning in this way.

¹² Lin (2016), pp. 324-27.

The upshot of this argument is that a person’s life can go better or worse for her than an inanimate object’s existence does for that object. Admittedly, it would be odd to go around saying that you are better off than rocks, chairs, and other inanimate objects. But something can be true even though it would be odd to say. Usually, people are so much better off than inanimate objects are that this isn’t worth saying. There is “no comparison” between your existence and that of a rock only in the sense that there is “no comparison” between a meal at a Michelin-starred restaurant and a stale cracker: the former is so much better than the latter that the act of comparing them misleads by suggesting that they are close enough in quality to be worth comparing. If we imagine a life that is just barely good on balance for the person living it, it seems less odd to say, because it is more worth saying, that it is better than the existence of a rock. There is some residual oddness, but it isn’t probative: it would also sound odd to say that this life is better than that of a cat, even though this can surely be true. We are just not used to comparing, and thus find it odd to compare, our welfare with that of non-human entities. Moreover, a large part of the point of making welfare comparisons is to provide guidance for efforts to improve the condition of those who are worse off. Because we have no way of making inanimate objects better off than they are, it is unclear why we would bother noting how they compare in welfare to people. For reasons like these, the oddness, in most contexts, of welfare comparisons with inanimate objects doesn’t show that such comparisons cannot be true. Indeed, this oddness dissipates considerably once we appreciate the aforementioned reasons for thinking that such comparisons can indeed be true.

We can accommodate the judgment that people can be better or worse off than inanimate objects by ascribing a neutral or zero amount of well-being to those objects. The person with the excellent life is better off than the rock because, being positive in welfare, he has a larger amount of well-being; the person with the terrible life is worse off than the rock because, being negative in welfare, he has a smaller amount of well-being. The claim that inanimate objects can have a level of well-being—namely, zero—isn’t implausible, since the ability to have a neutral or zero welfare level doesn’t entail the ability to be positive or negative in welfare or the ability to be made better or worse off.

On the narrow definition, according to which welfare subjects are entities capable of being positive or negative in well-being, we can make these claims while denying that inanimate objects are welfare subjects, since these objects not only do not have but are *incapable* of having positive or negative

amounts of welfare. But on the broad definition, on which the capacity to have even a neutral or zero amount of welfare is sufficient for being a welfare subject, we cannot ascribe a neutral or zero amount of well-being to such objects without considering them welfare subjects. Indeed, the broad definition implies that people cannot be better or worse off than such objects unless those objects are welfare subjects. After all, as I noted earlier, the point of distinguishing a zero amount of welfare from no amount of it, as the broad definition does, is to deny that entities that have no amount of it can enter into welfare comparisons. If we posit amounts of welfare, as both definitions do, then we must hold that one entity is better off than another if and only if the first entity’s amount of welfare is larger than the second’s. Clearly, an amount of welfare cannot be larger or smaller than *no* amount (as contrasted with a zero amount) of it. Thus, the broad definition implies that if the rock is not a welfare subject, then its existence is not worse than the excellent life and not better than the terrible one. More generally, it allows us to recognize that people can be better or worse off than inanimate objects only if we consider such objects welfare subjects. But this would contradict the consensus among theorists that inanimate objects aren’t welfare subjects and enlarge the extension of ‘welfare subject’ so much that the term wouldn’t mark the evaluatively significant boundary that it should. If the term lumps rocks and people together and ascribes to the latter nothing more than a capacity had by the former, then the distinction between welfare subjects and other entities matters less than a distinction within the class of welfare subjects that separates rocks from people. The term is more useful if it marks the latter distinction, as it does on the narrow definition. Therefore, if people can be better or worse off than inanimate objects, as I have argued, then we should prefer the narrow definition to the broad one because only the former allows us to accommodate this fact without deeming such objects welfare subjects.

Some readers might not be entirely convinced by my argument that people can be better or worse off than inanimate objects. Notice, though, that even the *epistemic possibility* of that claim renders the narrow definition preferable. The narrow definition doesn’t force us to maintain that people can be better or worse off than inanimate objects: it is consistent with the view that such objects can’t have any welfare levels (not even zero) and thus that such comparative judgments cannot be true. Thus, even if you doubt that such judgments can be true, this is no reason to prefer the broad definition to the narrow one, since you can say that such objects can’t have even zero welfare regardless of which definition you accept. By contrast, even if you are not entirely convinced that such judgments can be true, you have reason to prefer the narrow definition on the grounds that it is at least epistemically

possible that they can be true—even if, as I have assumed, inanimate objects are incapable of being positive or negative in welfare (because they are incapable of possessing anything that might be a basic good or a basic bad). If you can’t rule out the possibility that people can be better or worse off than inanimate objects, then you can’t rule out the possibility that the broad definition, unlike the narrow one, implies that such objects are welfare subjects and thus excessively enlarges the class of welfare subjects. Even if I haven’t established that people can be better or worse off than inanimate objects, I hope to have shown that, for all we know, they can be—even if, as we ordinarily assume, such objects are incapable of being positive or negative in welfare. Since, other things being equal, we should prefer definitions that allow us to plausibly accommodate hypotheses that could turn out to be true, this already gives us reason to prefer the narrow definition.

3. Explaining Which Entities Are Welfare Subjects

The second reason we should prefer the narrow definition is that, unlike the broad definition, it is compatible with a natural and attractive picture of how the facts about which entities are welfare subjects are explained by those about the true theory or theories of well-being.

It is standardly assumed that theories of well-being purport to explain how well off people are by identifying the basic goods and bads. For example, hedonism claims that pleasure and pain are the only basic good and bad, respectively, whereas the objective list theory accords the status of basic goods to a plurality of kinds, such as friendship, achievement, and knowledge. On the picture that I have in mind, the most fundamental facts about welfare, at least at the normative ethical level, are ones about which kinds are basic goods and bads and how basically good or bad instances of these kinds are. When paired with facts about which entities are capable of possessing instances of these kinds, these facts straightforwardly explain which entities are welfare subjects: an entity is a welfare subject if and only if (and if so, in virtue of the fact that) it is capable of having an instance of at least one basic good or bad. If hedonism is true, for example, then it explains why all and only creatures capable of experiencing pleasure or pain are welfare subjects. And inanimate objects aren’t welfare subjects on any plausible theory of well-being because they are incapable of possessing instances of any kinds that plausibly are basic goods or bads.¹³

¹³ Bradford (2023, pp. 919-20) notes the plausibility of this picture.

The picture is simplest if we assume, as I just did, that there is only one true theory of welfare. But it remains natural and attractive, though it is more complicated, if there are multiple true theories, each of them true of a different class of entities. Suppose that there are two true theories. The first, which is true of entities capable of valuing things, holds that the fulfillment of one’s values is the only basic good and the frustration of one’s values the only basic bad. The second, which is true of entities not capable of valuing things, holds that pleasure is the only basic good and pain the only basic bad.¹⁴ In that case, the picture that I am describing holds that the fundamental facts about welfare are (i) that value-fulfillment and value-frustration are the only basic good and bad, respectively, for valuers and (ii) that pleasure and pain are the only basic good and bad, respectively, for non-valuers. It also holds that an entity is a welfare subject if and only if (and if so, because) it is capable of having an instance of at least one kind that is a basic good or bad for it. Thus, if these two theories are the true theories, then their truth explains why the class of welfare subjects is the union of the class of entities capable of having fulfilled or frustrated values and the class of entities capable of feeling pleasure or pain. As before, the central tenets of the true theory or theories explain which entities are welfare subjects.

The narrow definition is compatible with this picture. After all, barring the baroque and implausible possibilities that I mentioned earlier, an entity is capable of being positive or negative in welfare if and only if it is either capable of having an instance of a kind that is a basic good for it or capable of having an instance of a kind that is a basic bad for it. Thus, the entities that this picture identifies on the basis of the facts about the basic goods and bads and what it takes to possess instances of them are precisely the welfare subjects, narrowly defined.

By contrast, we cannot accept this picture if we accept the broad definition. This is because, on this definition, an entity that cannot have any instance of any kind that is a basic good or bad could be a welfare subject in virtue of being able to have a zero amount of well-being: the capacity to possess an instance of a basic good or bad is sufficient but not necessary for being a welfare subject, since a distinct capacity that doesn’t entail it (*viz.*, the capacity to have a zero welfare level) is also sufficient. If the broad definition is true, then the fact that rocks aren’t welfare subjects isn’t fully explained by the fact that they cannot possess instances of pleasure, pain, or any other kinds that could plausibly be basic goods or bads. Since this incapacity of theirs is compatible with their having a zero welfare

¹⁴ Philosophers who defend theories on which value-fulfillment occupies the central role, such as Raibley (2010), Dorsey (2017), and Tiberius (2018), typically intend their theories to apply only to valuers and admit that some other theory or theories are true of non-valuers.

level, the fact that they aren’t welfare subjects is also partly explained by the fact that they cannot have even a zero welfare level. More generally, the reason why the class of welfare subjects is no larger than it is isn’t just that nothing outside it can possess any instances of any basic goods or bads: it’s also that nothing outside it can have even a zero welfare level. This view about why the entities that are welfare subjects, and no others, have that status is less explanatory than the picture available to the proponent of the narrow definition. For the fact that certain entities cannot have even a zero amount of welfare itself stands in need of explanation, and it is not obvious what could explain it. Since it is better to define a term in a way that makes it easy to explain why it has the extension that it does, we have reason to prefer the narrow definition.

A proponent of the broad definition might reply that, on that definition, we can give a different but equally good explanation of which entities are welfare subjects. But although I can’t rule this out, I struggle to see what such an explanation might be. It can’t be that the things that are welfare subjects have that status in virtue of being capable of having a welfare level. After all, on the broad definition, all it means to say that something is a welfare subject is that it is capable of having a welfare level, so this explanation would amount to the claim that the things that have that capacity have it in virtue of having it. And although it wouldn’t be circular to claim, as Lee does¹⁶, that the things that are welfare subjects have that status in virtue of being capable of phenomenal consciousness, this claim would itself stand in need of explanation. It’s easy enough to see what could explain why entities incapable of phenomenal consciousness are incapable of being positive or negative in welfare: perhaps some theory of welfare on which all of the basic goods and bads require phenomenal consciousness (e.g., hedonism), is true. But what could explain why entities incapable of phenomenal consciousness are incapable of having even zero welfare?¹⁷ Until proponents of the broad definition show otherwise, it seems to me that we can better explain which entities are welfare subjects if we accept the narrow definition. This is a reason to prefer the narrow definition.

4. Objections and Replies

4.1 *Rejecting Amounts of Well-Being*

¹⁶ Lee (forthcoming).

¹⁷ Lee (forthcoming) gives no explanation, and his view may be that none is possible: he is inclined to treat as basic the view that something is a welfare subject (broadly defined) just if and because it is capable of phenomenal consciousness (p. 7). I will say more about his paper in section 5.

As I formulated them, both the narrow and the broad definitions posit numerically representable amounts or levels of well-being. If you suspect that we shouldn't postulate such amounts (perhaps because this licenses judgments about well-being that are too precise), you might suspect that both definitions should be rejected and thus that it doesn't matter if one of them is better than the other.

The definitions can be restated in a way that doesn't presuppose such levels of well-being, however. As I said earlier, what's crucial to the narrow definition is that it draws a boundary around the beings capable of faring well or badly. Thus, it could be recast as follows:

Narrow Definition* S is a welfare subject \equiv_{df} S is capable of being well off or badly off.¹⁸

By contrast, the broad definition evidently wants to draw a boundary around the beings about which any positive judgments concerning well-being (e.g., judgments that affirm that they are better off or worse off than certain entities, as opposed to judgments that merely deny such things) could be true. After all, the effect of insisting that entities that aren't welfare subjects have *no* amounts of welfare as opposed to neutral or zero amounts is precisely to rule out such positive judgments about their well-being. So, the broad definition could be restated like this:

Broad Definition* S is a welfare subject \equiv_{df} some positive judgments concerning S's well-being can be true.

Suitably modified versions of my arguments would show the superiority of the narrow definition to the broad one even when they are so construed. The first argument would observe that, while the narrow definition can hold that you are better off than a rock even though it isn't a welfare subject (because the rock, though incapable of being well off or badly off, is both worse off than any entity that is well off and better off than any entity that is badly off), the broad definition can't. After all, if the rock isn't a welfare subject, then the broad definition can't accept the positive judgment that it is worse off than you are and thus can't claim that you are better off than it is. The second argument would observe that, according to the broad definition, the fact that rocks aren't welfare subjects is

¹⁸ We could say that for S to be well (badly) off is for S to be better (worse) off than S would be if S were not to exist, or we could take these concepts to be primitive.

not fully explained by their inability to possess instances of any basic goods or bads, as it is on the narrow definition. Instead, it is at least partly explained by their even more profound inability to be subjects of any true positive judgments about well-being—an inability that cries out for explanation but is not obviously explicable. Thus, although I have proceeded as though there are numerically representable amounts or levels of well-being, neither the interest of my subject matter nor the force of my arguments depends on that assumption.

4.2 The Capacity for Zero Welfare

Both of my arguments assumed that, from the fact that something cannot have any basic goods or bads and thus cannot be positive or negative in welfare, it doesn't follow that it cannot have a zero or neutral welfare level. To the extent that they have addressed this issue, defenders of the broad definition join me in accepting this assumption.¹⁹ However, you might question this assumption on the general grounds that an entity cannot have a zero amount of something (as contrasted with no amount of it) unless it can have a positive or negative amount of it.

That principle is false, however. An entity incapable of feeling pleasure is one incapable of having a positive amount of it—or a negative amount of it, whatever that might mean. But such an entity can and does have a zero amount of pleasure. For if it had no amount of pleasure, then the large amount of pleasure had by a person who is in ecstasy wouldn't be greater than its amount of pleasure, and it wouldn't be true, as it surely is, that such a person feels *more* pleasure than it does. There is nothing strange about the position that some things can possess zero amounts, and only zero amounts, of pleasure, since this merely implies that some things cannot have any pleasure but can have less of it than is had by anything that has some pleasure.²⁰ More generally, there is nothing strange about the view that an entity can have a zero amount of something (e.g., well-being) even though it cannot have a positive or negative amount of it.

4.3 Comparing Good Lives with Non-Existence

¹⁹ Lee (2022), p. 208n5; Lee (forthcoming), pp. 16, 21.

²⁰ The claim that rocks have zero amounts of pleasure might, for some, have the feeling of a category mistake. Given what I have just argued, I take this to show only that one can easily be misled into thinking that a true claim involves a category mistake—a fact that bolsters my first argument.

It is plausible that a person with a good life is better off than he would be if he didn’t exist. Some might think, however, that such a person would have no welfare level at all if he didn’t exist. This might suggest, contrary to my first argument, that even the broad definition can consider people better or worse off than inanimate objects without deeming such objects welfare subjects. For if a person can be better off than he would be if he had no welfare level, then he can presumably be better or worse off than an inanimate object that has no welfare level.

If people with good lives are indeed better off than they would be if they didn’t exist, however, this can only be because they would have a lower welfare level (namely, zero) if they didn’t exist. For as I explained earlier, one point on a scale can be compared only with another point on it, even if only the neutral or zero point: it cannot be compared with nothing at all. This is why we cannot compare the elevation of Denver with that of the number 7 or the pitch of high C with that of the color red. Thus, if there are welfare levels, then someone is better off in one scenario than in another just in case his welfare level is higher in the one than in the other, and someone is better off at one time than at another just in case his welfare level is higher at the one than at the other. I am not aware of any good reasons for rejecting this view, which is endorsed both by (i) theorists who hold that we *can’t* be better off existing than not existing because non-existent people have no welfare levels and by (ii) theorists who hold that we *can* be better off existing than not existing and, accordingly, that non-existent people have a welfare level (namely, zero). And although I can’t rehearse them here, I find convincing Ben Bradley’s arguments for the claim that actually existing people would have zero welfare if they didn’t exist.²¹ Thus, I don’t think my first argument is undermined by the intuition that people with good lives are better off than they would be if they didn’t exist. Those who are convinced that no person would have any welfare level if he didn’t exist should reject that intuition.

What about Fred Feldman’s deprivationist account of the badness of death, on which death is bad for the person who dies because it deprives him of the well-being that he would have accumulated if he had lived longer?²² This account can explain why death is bad for the person who dies without assuming that a person has a welfare level after he goes out of existence, so it might appear to show that a person can be better off existing than not existing even if he would have no welfare level at all if he didn’t exist. This appearance would be mistaken, though. On this account, the sense in which a

²¹ Bradley (2009), pp. 98-111; Bradley (2013), pp. 45-48.

²² Feldman (1991).

person’s death, at a given time, can be bad for him is the following: the person’s lifetime amount of welfare can be lower, given that he dies at that time, than his lifetime amount of welfare would have been if he hadn’t died at that time.²³ Since this involves a comparison of lifetime welfare levels across scenarios, it obviously doesn’t suggest that a person can be better off in one scenario than in another even if he has no welfare level in one of the scenarios. It is true that this account doesn’t assume that a person has a welfare level after he goes out of existence. But nor does it claim that a person can be worse off *at a time after he goes out existence* than he would have been at that very time if he had existed then (or than he was at an earlier time at which he existed).²⁴ On my view, a person can be worse off at a time after he ceases to exist than he would have been at that time if he had still existed then (or than he was at an earlier time at which he existed) only if he has a welfare level—even if only zero—at the later time. Since Feldman’s account makes no intertemporal comparisons of this sort, it poses no threat to this view.²⁵

4.4 *Zombies and Mombies*

Recall that Kriegel favors the broad definition because he wants to capture the difference between entities that aren’t welfare subjects (e.g., zombies, in his view) and welfare subjects who are neither well off nor badly off because the goodness and badness in their lives balance each other out exactly (e.g., mombies). You might worry that the narrow definition ignores this difference because it allows entities of both types to have the same welfare level: zero. Notice, though, that if no basic goods or bads can be had by zombies (as Kriegel holds²⁶), then zombies and mombies differ modally: zombies aren’t capable of being positive or negative in welfare, either on a lifetime basis or during intervals shorter than their lives, but mombies are. Moreover, while a zombie is never positive or negative in welfare, a mombie is positive in welfare at some times and negative in welfare at others: it’s just that, “at the end of the day, its [life’s] net worth is zero.”²⁷ Plausibly, this is why the zombie’s life is not worth living in a deeper sense than the mombie’s life is. We can capture the difference between the two even if we assign them both a zero amount of well-being.

²³ Feldman (1991), pp. 209, 215-16.

²⁴ Feldman (1991), pp. 219-20.

²⁵ Bradley (2009) defends a deprivationist account of the badness of death which, unlike Feldman’s, can say that a person can be worse off at a time when he no longer exists than he would have been then if he still existed or than he was at an earlier time when he still existed. But his account can say this precisely because it claims that people have a welfare level (viz., zero) at times after they cease to exist. See Bradley (2009), ch. 3.

²⁶ Kriegel (forthcoming), pp. 13-15, 17, 20.

²⁷ Kriegel (forthcoming), p. 9.

4.5 Moral Implications of Welfare Subjecthood

Some might wonder what the moral implications of ascribing welfare subjecthood to an entity are. Some might also feel that the case for the narrow definition would be more compelling if it showed that only this definition can accommodate those implications. I have not given an argument of this sort because I lack strong opinions, independent of my acceptance of the narrow definition, about whether welfare subjecthood has moral implications and, if so, what they are. If you are convinced that welfare subjecthood does have moral implications, however, then I can offer you the following expansion on my first argument.

If inanimate objects have zero welfare, as I argued, then welfare subjecthood has moral implications on the narrow definition but not on the broad one. If the narrow definition is true, then ascribing welfare subjecthood to an entity creates a strong presumption that we can make it better or worse off. For although we can conceive of an entity capable of being positive or negative in welfare but incapable of becoming better or worse off than it is, there seem to be no such entities: in every case that we are aware of, entities capable of being positive or negative in welfare can become better or worse off by undergoing changes in the quantities of basic goods or bads that they possess. Thus, because it is plausible that we have moral reasons to make entities better off and not to make them worse off, welfare subjecthood has moral implications on the narrow definition: if something is a welfare subject, this creates a strong presumption that we have moral reasons to take the potential effects of our actions on its welfare into account when deliberating about what to do. By contrast, if inanimate objects have zero welfare, then they are welfare subjects on the broad definition. Because there is nothing we can do to make such objects better or worse off than they are, and because there appear to be at least as many such objects as there are welfare subjects of the sort that we can make better or worse off (e.g., people), ascribing welfare subjecthood to an entity does not create any such presumption if the broad definition is true. Instead, what creates such a presumption is membership in a proper subset of the set of welfare subjects—namely, the one whose members are capable of being positive or negative in welfare.

Recall that my first argument showed that proponents of the broad definition can accommodate welfare comparisons with inanimate objects only by deeming such objects welfare subjects, which

would not only contradict the consensus among theorists that such objects aren’t welfare subjects but so greatly enlarge the extension of ‘welfare subject’ that the term would no longer be of much use. I can now elaborate on the claim that the term would no longer be of much use: it would no longer have any moral implications. If you believe that ‘welfare subject’ should be defined in such a way that it has moral implications, then you should prefer the narrow definition to the broad one. Furthermore, as before, even if you are not entirely convinced that inanimate objects do have zero welfare, the fact that this is epistemically possible gives you reason to prefer the narrow definition. For if it is epistemically possible that inanimate objects have zero welfare, it is epistemically possible that ‘welfare subject’ has no moral implications if the broad definition is true. If you are convinced that welfare subjecthood has moral implications, you should not want to leave this possibility open.

4.6 Comparisons to Similar Terms

As a general matter, you might find it unusual that a term of the form ‘F subject’ would fail to mean ‘entity capable of having an amount of F (even if only zero)’ or ‘entity capable of having an F-value (even if only zero)’, and you might therefore like me to give other examples of such terms that don’t merely mean such things. It seems to me, however, that such terms are sufficiently rare that nothing qualifies as the standard pattern of what they mean—in which case it is no objection to the narrow definition that it implies that ‘welfare subject’ doesn’t conform to the pattern described above. But if you still want an example, consider ‘moral subject’: to say that something is a moral subject is not to say that it is capable of having an amount (even if only zero) of morality, whatever that might mean. My point, obviously, is not that ‘moral subject’ is defined as I’ve argued we should define ‘welfare subject’: a moral subject isn’t an entity capable of having a positive or negative amount of morality, either. My point is rather that, as ‘moral subject’ illustrates, there’s no reason to expect every term of the form ‘F subject’ to be defined analogously to how the broad definition defines ‘welfare subject’. As a technical term, ‘welfare subject’ should suit our theoretical purposes. We shouldn’t let general hunches about what terms of the form ‘F subject’ should mean dictate how we define it.

Moreover, consider how it would make sense to define ‘hedonic subject’, if this term were in use. As I argued earlier²⁸, since a person who is in ecstasy feels *more* pleasure than an inanimate object does, such objects feel *zero* pleasure (rather than no amount of it). It would hardly be convincing to insist

²⁸ See section 4.2.

that, since it sounds odd to say that something that has an amount of F nonetheless fails to qualify as an F subject, we must define ‘hedonic subject’ in such a manner that inanimate objects fall in its extension—even if we have good reasons, analogous to the ones I have given, for taking it to mean ‘entity capable of feeling a positive amount of pleasure or pain’.

4.7 Subjects of Prudential Value or Disvalue

Earlier, I dismissed the possibility of an entity that cannot possess any instance of a basic good (bad) without simultaneously possessing an instance of a basic bad (good) that precisely counterbalances it. But however baroque and implausible the possibility of such an entity might be, its conceivability suggests an amendment to the narrow definition that we should consider. Such an entity wouldn’t satisfy the narrow definition even if it could possess things that are good or bad for it, since it could possess them only in such amounts that its welfare level would be zero. Some might think, however, that such an entity would nonetheless qualify as a welfare subject if it could indeed possess things that are good or bad for it, in the sense of ‘good for’ and ‘bad for’ that is conceptually tied to well-being. If this is correct, then welfare subjects should be defined as follows:

Narrow Definition’ S is a welfare subject =_{df} S is capable of possessing something that has prudential value or disvalue for it.

This definition is better than the broad one in the same two ways that the original narrow definition is. Since nothing has prudential value or disvalue for inanimate objects, it excludes such objects from the class of welfare subjects while allowing us to claim that, because they have zero welfare, they can be better or worse off than people are. And, even more straightforwardly than the original narrow definition does, it allows us to claim that which entities are welfare subjects is explained by the true theory or theories of well-being. Thus, I take my case for the original narrow definition to double as a case for this one. And since the two definitions, though very close, are distinct, I have shown, at best, that we should accept one or the other of them.

Whereas the case against the broad definition and in favor of one or the other of the narrow ones is quite compelling, it doesn’t seem to me that either of the narrow definitions is clearly superior to the other. The boundary between entities capable of possessing things that are good or bad for them—

even if they are nonetheless locked into a zero welfare level—and entities incapable of doing this is clearly evaluatively significant, and it would be reasonable to use ‘welfare subject’ to mark it. On the other hand, the concept of welfare is distinct, though closely related to, those of prudential value and disvalue: an individual’s welfare is an overall measure of the degrees to which things have prudential value and disvalue for it. It would therefore also be reasonable to distinguish welfare subjects from what we might call *subjects of prudential value or disvalue*, and to hold that whereas the original narrow definition tells us what the former are, the amended one defines the latter. I incline toward this latter position, but I doubt that the choice between them matters much.

5. Lee on the Phenomenal Theory of Welfare Subjects

Partly because it will help illustrate the importance of my arguments—and, more generally, that of explicitly considering how ‘welfare subject’ should be defined—I want to end by discussing a recent paper in which Andrew Lee defends the *phenomenal theory* of welfare subjects: that something is a welfare subject if and only if and because it has the capacity for phenomenal consciousness.³⁰ Lee sometimes characterizes a welfare subject as “the kind of thing that can be better or worse off,” but his official definition is the broad one.³¹ With ‘welfare subject’ so defined, the phenomenal theory claims that something is capable of having an amount of well-being (even if only a neutral or zero amount) if and only if and because it has the capacity for phenomenal consciousness.

Lee considers alternatives to the broad definition only briefly and implicitly, after assuming it for the bulk of his paper.³² This is unfortunate, since some of his arguments better fit the narrow definition than the broad one. For example, he claims that the phenomenal theory is supported by the fact that it explains why consciousness is ethically significant (*viz.*, because “consciousness is what makes an entity a welfare subject”) and why people cease to be welfare subjects when they die (*viz.*, because death involves the loss of the capacity for phenomenal consciousness).³³ If having the capacity to have an amount of well-being (even if only a neutral or zero amount) is all it means to be a welfare

³⁰ Lee (forthcoming).

³¹ Recall that I am eliding a minor difference between Lee’s formulation of the broad definition and mine. See note 11.

³² Lee (forthcoming), p. 24. His only argument against the narrow definition is that an entity whose welfare is necessarily zero because its life necessarily contains perfectly balanced amounts of goodness and badness for it is a welfare subject. Although this argument contradicts the narrow definition as I originally formulated it, it is consistent with the variant on that definition that I formulated in section 4.7. Because I am really arguing only that we should accept one of these two versions of the narrow definition, nothing that Lee says here undermines my argument.

³³ Lee (forthcoming), pp. 8-10.

subject, however, then it’s unclear that the phenomenal theory really does explain why consciousness is ethically significant because it’s unclear that welfare subjecthood is ethically significant. After all, even an entity whose welfare level is necessarily zero is a welfare subject so defined³⁴, but since such an entity cannot be made better or worse off by anything that anyone does, it’s unclear that such an entity is ethically significant. By contrast, since there are no known cases of entities that cannot be made better or worse off even though they are capable of being positive or negative in welfare, it is clearer that the phenomenal theory, understood as a theory of welfare subjects narrowly defined, explains why consciousness is ethically significant.³⁵ Moreover, on Lee’s definition, the claim that we cease to be welfare subjects when we die amounts to the claim that we become incapable of having any amount of welfare, even zero, when we die. Since, as I explained earlier, this claim implies that a person cannot be better off while he is alive than he will be after he dies³⁶, it isn’t clear that it is true. Thus, it’s unclear that the fact that the phenomenal theory explains this claim counts in its favor. By contrast, it is clear that we cease to be welfare subjects narrowly defined when we die (i.e., that we cease to be capable of being positive or negative in welfare). The phenomenal theory, understood as a theory of welfare subjects narrowly defined, is supported by the fact that it explains this. Thus, as he likely would have noticed if he had given more consideration to how ‘welfare subject’ should be defined, Lee’s case for the phenomenal theory is stronger if we define that term narrowly than if we define it broadly.

Moreover, my first argument generates a serious objection to the phenomenal theory as Lee intends it. If inanimate objects have zero welfare, then such objects qualify as welfare subjects in the broad sense: they are entities capable of having an amount of well-being. However, such objects are surely incapable of phenomenal consciousness. Thus, at least unless something like panpsychism turns out to be true, it’s false that something is a welfare subject in the broad sense if and only if and because it has the capacity for phenomenal consciousness. When paired with the narrow definition, however, the phenomenal theory avoids this objection. This is another reason to think that it is more plausible as a theory of welfare subjects in the narrow sense.

Lee’s paper is rich and interesting, and although it concerns the first-order question of which entities are welfare subjects (and why) rather than the higher-order question of how ‘welfare subject’ should

³⁴ Lee (forthcoming) is explicit about this (e.g., on pp. 16, 21).

³⁵ See section 4.5.

³⁶ See section 4.3.

be defined, it connects with mine in more ways that I have the space to describe. I hope my brief discussion of it illustrates the fruitfulness of explicitly considering how ‘welfare subject’ should be defined, as well as some of the implications of my arguments.

6. Conclusion

Because ‘welfare subject’ is a technical term, the task of defining it is not constrained by pretheoretic intuitions about its meaning. But it can be guided by the fact that theorists who use the term intend it to delimit an evaluatively significant class from which they all agree inanimate objects are excluded, as well as by the fact that we shouldn’t rule out comparisons of welfare that are intuitively possible (or, at least, not certainly impossible). It can also be guided by the fact that, other things being equal, it is better to define a term in a way that makes available a natural and attractive picture of how its extension is explained than to define it in a way that makes such a picture unavailable. I have argued on this basis that we should accept the narrow definition of ‘welfare subject’, on which this term expresses the concept of an entity capable of having a positive or negative amount of well-being—that is, one capable of being well off or badly off. In the course of arguing for that conclusion, I have also explained why it is plausible that inanimate objects can have a level of welfare—namely, zero. Such objects are not welfare subjects because, since nothing has prudential value or disvalue for them, they cannot be well or badly off. But the existence of such an object can nonetheless go worse for it than a good life does and better for it than a bad life does.

Works Cited

- Bradford, Gwen (2023), “Consciousness and Welfare Subjectivity,” *Noûs* 57(4): 905-21.
- Bradley, Ben (2009), *Well-Being and Death* (Oxford: Oxford University Press).
- _____(2013), “Asymmetries in Benefiting, Harming and Creating,” *Journal of Ethics* 17(1/2): 37-49.
- Deijl, Willem van der (2021), “The Sentience Argument for Experientialism about Welfare,” *Philosophical Studies* 178(1): 187-208.
- Dorsey, Dale (2017), “Why Should Welfare ‘Fit?’” *Philosophical Quarterly* 67(269): 685-708.
- Feldman, Fred (1991), “Some Puzzles About the Evil of Death,” *Philosophical Review* 100(2): 205-27.
- Kriegel, Uriah (forthcoming), “The Value of Consciousness to the One Who Has It,” in Geoffrey Lee and Adam Pautz (eds.), *The Importance of Being Conscious* (Oxford: Oxford University Press).
- Lee, Andrew Y. (2022), “Speciesism and Sentientism,” *Journal of Consciousness Studies* 29(3-4): 205-28.
- _____(forthcoming), “Consciousness Makes Things Matter,” *Philosophers’ Imprint*.
- Lin, Eden (2016), “How to Use the Experience Machine,” *Utilitas* 28(3): 314-32.
- _____(2018), “Welfare Invariabilism,” *Ethics* 128(2): 320-45.
- Raibley, Jason (2010), “Well-Being and the Priority of Values,” *Social Theory and Practice* 36(4): 593-620.
- Sumner, L. W. (1996), *Welfare, Happiness, and Ethics* (Oxford: Clarendon Press).
- Tiberius, Valerie (2018), *Well-Being as Value Fulfillment* (Oxford: Oxford University Press).