The Value of Consciousness to the One Who Has It

Uriah Kriegel

In G. Lee and A. Pautz, *The Importance of Being Conscious* (OUP, forthcoming)

Abstract. There is a strong intuition that a zombie's life is never good or bad *for the zombie*. What explains this? In this paper, I consider five possible explanations of the intuition that a zombie's life is never worth living, plus the option of rejecting the intuition. I point out the considerable costs of each option, though making clear which option strikes me as least problematic.

Introduction

Philosophers of mind like to talk about zombies: creatures that are behaviorally and functionally indistinguishable from us but lack any conscious experience. David Chalmers (1996) notoriously made zombies the cornerstone of his case against physicalism, though a quarter-century earlier Keith Campbell (1970: 108) already spoke of "the imitation man, who duplicates all of a typical man's acquisition, processing, and retrieval of information, and all his activity, but for whom there are no phenomenal properties." Now, when I think of the life led by a zombie imitation of me – going to its office, preparing its classes, writing papers and submitting them to journals, but in the manner of a robot or automaton, without anything going on "on the inside" – it seems to me like a life that could be useful or even inspiring to family members, colleagues, students, and so on, but a life of no significance *to the zombie itself*. Moreover, nothing that happens in this zombie's life seems to make the zombie's life better or worse *for the zombie*. In other words, I have a strong intuition that *a zombie's life is never worth living* – not in the sense that it's bound to be a *bad* life, but in the sense that it's not the kind of thing that can be good or bad.

The force of this intuition can be brought out by certain thought-experiments about possible zombification (see Siewert 2021). We may put it in the form of the following vignette.

> (*Zombified*!)
>
> Suppose God appears in the burning bush and tells you there is good news and bad news. The bad news is that you will be turned into a functionally indistinguishable zombie, irreversibly, in exactly 24 hours. The good news is that in your zombie state

> you will get to live for 900 years and gain untold riches of many sorts (you may, e.g., put together the world's biggest collection of your favorite painter's artworks), fulfill many of your life goals (e.g., writing a book that will transform philosophy), and more.

If your sole concern is your own wellbeing, you are likely to be very disappointed by this piece of news, and indeed feel slightly cheated by the announcement that there will also be good news. What good are those long years of riches and accomplishment to you if you can't experience any of it?

Suppose further that the devil, seeing you so downcast, jumps on the opportunity and threatens you as follows: if you don't carry out one of the devil's wishes in the next 24 hours, then after a week or two in your zombie state, when God isn't looking, the devil will see to it that you will be killed and replaced with a physical replica so perfect that even God won't notice the difference. (God, in this scenario, has no insight into haecceities, and sometimes loses concentration.) Here I predict that the devil's threat will underwhelm you: as far as you're concerned, once you're a zombie, the prospect of being replaced by an indistinguishable but distinct zombie will really make no difference to you. There is certainly no point in wasting your precious conscious time in the next 24 hours trying to avoid this outcome.

I am not offering these thought-experiments as *arguments* for the proposition that a zombie's life is never worth living (for the zombie). The proposition strikes me as *intuitively compelling*. I adduce the thought-experiments by way of *fleshing out* the way we are intuitively compelled to think about the role conscious experience plays in making our life good or bad for us.

It is sometimes said that a flower fares better when it has sun, water, etc., and that to that extent a flower has a life that can go better or worse for it. If a flower, why not a zombie? I think it's pretty obvious, however, that the sense of "faring better" and "life that can go well" is completely different here than in the case of conscious beings. A zombie who gets nutritious food and exercises regularly is likely to "fare better" than a chain-smoking zombie on an exclusive Taco Bell diet who never gets off the couch – fare better, that is, in the sense in which flowers can fare better. But this is, we may say, a purely "external" sense of faring better. In that sense of faring better, one corkscrew fares worse than another when it rusts for years behind the fridge. It is quite a different sense from the sense in which we want our lives to go well for us, notably when we want to be happy, lead a meaningful existence, and so on. In *this* sense – the sense in which we most saliently *want* our lives to go well for us – it is hard to make sense of the idea of a zombie's life going better or worse for it.

Intuitively, then, there is this link between consciousness and the good life: without the former, you are ineligible to have the latter. This fact raises a number of interesting

questions linking central discussions in philosophy of mind and normative ethics, more specifically between the independently vibrant literatures on phenomenal consciousness and wellbeing/welfare. The question I will focus on here may be put as follows: What is the good life that a zombie can't have it? That is, what is wellbeing that consciousness is a *sine qua non* for it? For much of this paper, I will take it as a *datum* that a zombie's life is never worth living, and will consider five possible "philosophical hypotheses" that might explain that datum. I will argue that each faces serious difficulties, though one of them seems to me tangibly less problematic than the others. I will then consider the possibility of "denying the datum" – i.e., allowing zombies to have better and worse lives – but will reject this option as even less plausible.

The link between conscious experience and wellbeing has not gone entirely unnoticed, of course. Long ago, James Griffin (1986: 13-4) introduced what he called the *Experience Requirement* on wellbeing – essentially, the claim that wellbeing supervenes on experience. Griffin's Requirement has been energetically debated of late (Bramble 2016, van der Deijl 2021, Lin 2021). As we will see, however, only two of the five "philosophical hypotheses" we will consider are consistent with the Experience Requirement. This shows that our zombie datum is far more neutral a starting point than the Experience Requirement in theorizing about wellbeing and the role of consciousness in it. This is not all that surprising, on reflection, given that the Experience Requirement is a theoretical principle, after all, whereas our zombie datum is but a deliverance of intuition.

It is important to appreciate that the kind of consciousness that matters to the zombie-life intuition is *phenomenal* consciousness – the subjectively experienced quality of our conscious life. There are other notions of consciousness, of course, notably the notion of *access* consciousness (Block 1995), but they seem irrelevant to the zombie-wellbeing intuition. Access-consciousness is, roughly, a mental state's property of being highly cognitively integrated and impacting information processing in many downstream modules. There are live debates in philosophy of mind about whether access consciousness is nomologically, metaphysically, or even conceptually separable from phenomenal consciousness; but these debates do not concern us here. *If* access consciousness is at least conceptually separable from phenomenal consciousness, then you can conceive of God allowing you to retain your access consciousness while losing all phenomenal consciousness, and I think this too will be no comfort to you.[1] Separated from phenomenality, access is after all a purely functional specification that can be instantiated or realized by any system with the right structural sophistication (e.g., a sufficiently well-orchestrated system of beer cans). In offering you to retain access consciousness, all God would be offering, then, is structural sophistication: you won't be any old automaton, but quite a complicated one! That is not what you want.

1. **First Hypothesis:** Phenomenological Hedonism

The most straightforward explanation of a zombie's life never being worth living is that the worth of a life for the one who lives it is fixed entirely by their experiences. On this view, only experiences can contribute to the goodness of a person's life. Or more precisely, only experiences can contribute *constitutively* (rather than causally), or *intrinsically* (rather than instrumentally), to the goodness of a life.[2] What doesn't "touch" a person's experience remains irrelevant to how well her life is going for her. Call this *experientialism about wellbeing*. Since *nothing* touches a zombie's experiences – there being none – experientialism entails that nothing contributes to the quality of a zombie's life, which consequently remains worthless. Datum explained.

One of the traditionally most prominent theories of wellbeing, hedonism, seems well positioned to avail itself of this explanation. Hedonism is sometimes framed as the view that wellbeing is fixed by the distribution of pleasure and pain. At other times the relevance of a wider range of affectively valenced mental states is also stressed – hope, awe, satisfaction, nostalgia, joy, gratitude, affection, contentment, and so on. Some of these it may be awkward to describe as 'pleasures,' though all seem aptly described as 'pleasant' or 'pleasurable.' Why the noun is awkward but the adjectives felicitous is a question we need not resolve here. I will construe hedonism in the more liberal way, as the thesis that the value of a life for the one who lives it is fully fixed by her affectively valenced mental states. Now, it is an open question in the philosophy of mind whether such states can occur unconsciously, that is, without being *experienced* by their subject. But the hedonist could either (a) take the substantive position that affectively valenced states are always experienced, or (b) remain silent on this question and simply restrict the scope of valenced states relevant to wellbeing to those which are experienced. On the resulting view, sometimes called "phenomenological hedonism," the value of a person's life (*to that person*) is fully grounded in her *affectively valenced conscious experiences*.

Hedonism was out of favor at the end of the twentieth century, mostly due to Nozick's (1974: 42-3) "experience machine" objection. More recently, however, it has seen something of a revival (Crisp 2006, Bramble 2016). Might we take the zombie datum as just one more reason to accept (phenomenological) hedonism? I think consideration of certain variants of the divine-intervention thought-experiments already aired may discourage this.

Nozick's experience-machine thought-experiment can itself be cast as a kind of divine-intervention scenario.

(*Better Berry*)

> Suppose God, having looked into your future, offers you to enter an experience machine that would reproduce your experiential life perfectly, except for one little

> improvement: a strawberry you'd be eating in real life when you're 70 and find somewhat underwhelming would taste *ah-mazing* in the experience machine. Moreover, God promises to install a zombie duplicate of you in your house, so that nobody else (family, students, etc.) would even notice your absence, much less be negatively affected by it.

Should you accept this offer? Most people report a preference for staying out of the experience machine and braving that mediocre strawberry. And this suggests that what we value in our life is not *just* the pleasurable experiences we have. (To be clear, when I speak, here and in what follows, of "what we would prefer," or "what you would choose" in some scenario, I mean what one prefers or chooses from an entirely prudential standpoint, and bracketing any other considerations.)

In recent years, hedonistic responses have proliferated, often attempting to debunk the experience-machine intuition (e.g., Silverstein 2000, Crisp 2006, De Brigard 2010). Still, it's hard to avoid the thought that whether our experience connects up with "real value" is something that makes a difference to the goodness of our life. Imagine two persons who lead experientially equi-valuable lives: the same amount of joy, the same amount of frustration and irritability, the same experience of meaning and fulfillment, and so on. But one of them is Shakespeare and one is Sisyphus (compare Wolf 1997). Shakespeare writes one mind-blowing play after another, Sisyphus pushes a rock up an endless hill grinning. Both experience an incredible sense of fulfillment and meaning as a result of their respective activities. The only difference is that Shakespeare's lifework *really is* meaningful, whereas Sisyphus' is entirely pointless. Do we really think they've led equally good lives? If God told you there's an afterlife and offered you to live either type of life (without remembering this conversation!), would you really be indifferent and tell God to flip a coin?

We can also tweak the thought-experiment so that Sisyphus gets one extra second in his life, and in that second has a very pleasant experience – another wonderful strawberry, or perhaps even a cheesecake bite. Would you then cease to be indifferent and ask to live the Sisyphus-type life over a Shakespeare-type life? I certainly wouldn't. This seems to suggest that we intuitively take the goodness of our life to depend in part on things that go beyond our experience and pertain to "real value."

Regardless of the ongoing experience-machine controversy, phenomenological hedonism faces significant difficulties with other divine-intervention scenarios.

> (*Vulcanized*!)
>
> Suppose God, seeing the disappointment on your face (and in your soul) after announcing to you the news that you'll be zombified, offers the following reprieve: instead of zombifying you entirely, He will "zombify" only your affectively valenced

> states. This is a procedure that retains these states' functional role but removes their valenced phenomenology, without messing with any of your non-valenced phenomenology.

In other words, what God is offering is to tune the affective dimension of your experiential life all the way down to zero, while keeping intact whatever other phenomenal dimensions your experiences have. For the remainder of your life, nothing will feel pleasant or unpleasant to you, but you will still have *conscious awareness*, and you will still undergo other, "affectively neutral" experiences. Thus, you'll have *perceptual* experiences, such as smelling freshly ground coffee; mnemonic experiences, such as recalling the first time you saw a panda at the zoo; *intellectual* experiences (with their so-called cognitive phenomenology), such as suddenly realizing the solution to some problem; *conative* experiences (with their agentive phenomenology), such as exerting effort trying to move the desk to the living room; and perhaps other experiences. None of these will occasion any pleasure or satisfaction, of course, and you may reasonably prefer keeping your shorter and affectively invested life over this affectively muted existence. Still, when God announces to you that you won't be completely zombified, but will get to keep your non-valenced experiences, and with them your conscious existence, it would probably seem to you like a *major break*. At least you'll get to *be there*, in a very basic sense. Your experiential death sentence has been commuted. But if phenomenological hedonism is true, then you should really be entirely unmoved by this development. Without affectively valenced experience, say phenomenological hedonists, life is not worth living.

Moreover, if phenomenological hedonism is true, then if the devil comes around snickering again, informing you that in a week's time you'll be destroyed and replaced by a replica, you should be just as unimpressed as before. But in fact this time I'd think you'd care a little more about being killed and replaced by someone else. Intuitively, it would be *great* to keep your non-valenced phenomenology (as compared to being completely zombified) and *terrible* to be killed if you have it.

David Chalmers (2022 Ch.18) calls creatures that are like us experientially except they lack affectively valenced phenomenology *Vulcans*. He argues that while our intuitive resistance to pushing the fat man off the bridge to stop a trolley from running over five people diminishes immensely once we're told that the man is a zombie, much of that resistance is left with us if we're told rather that he's a Vulcan. Likewise, I'm suggesting, if you yourself have to choose between *being* a zombie and *being* a Vulcan, the difference to you might be enormous. I am well aware, of course, that trolley cases probe our intuitions about *moral* value whereas the present topic is *prudential* value. But there is this relevant similarity: intuitively, as a Vulcan, your life should matter *to others* very much in trolley cases, whereas as a zombie, it should not; and likewise, as a Vulcan, your life should matter

to you in the cases we're considering, whereas as a zombie, your life – in whatever sense of "you" and "life," if any, makes it the case that *you* have a *life* – should *not* matter to you.

I'm open to the possibility that intuitions to the effect that being Vulcanized is better for one than being zombified rest on failure to imagine sufficiently vividly what it'd be like to live without any affective investment. Perhaps without affective phenomenology *nothing would matter to me* – nothing will be such that I *care* about it – and that would include my own existence. At the same time, we need to realize that this in no way falls out of the nature of caring. Caring is not itself a valenced experience: some caring experiences feel good, some feel bad, and some are neutral. So if a Vulcan is defined as lacking valenced experience but retaining all other experience, then as far as the *definition* is concerned, the Vulcan *could* care about her own life. It would therefore have to be a deeper, more surreptitious connection between valenced phenomenology and caring that would make it *impossible* to have the latter without the former.[3]

There is, however, yet another divine-intervention scenario that I find even more problematic for phenomenological hedonism.

(*Mildly Unpleasant*)

> It's another good news/bad news routine from God. This time, the bad news is that God looked into your future, and is letting you know that from here on out, your life will on the whole skew toward the unpleasant. There won't be infernal torment or anything like that – you won't be tortured either physically or psychologically. Nor will your life lack joy and fulfillment altogether. It's just that, on the whole, your life is going to be slightly more unpleasant than pleasant – there will be more dissatisfaction than satisfaction, more bad mood than good mood, and so on. The good news, now, is that, if you want, God could zombify you right away and spare you this on-the-whole-mildly-disagreeable existence that awaits you.

If you're anything like me, you'd politely decline, holding on to dear life of the mind despite its affectively negative accent. However, if phenomenological hedonism is true, this is just a mistake we'd be making here, a spurt of all-too-human irrationality. The right choice is to embrace zombiehood to ensure that our net pleasure/displeasure distribution is null rather than slightly negative. Again, this seems like the wrong verdict to return. It seems on the contrary perfectly rational, in such circumstances, to choose a continued conscious existence. What this suggests, I think, is that *experience as such* is something that brings value to our life – over and above the *character* of our experience as pleasant or unpleasant (see Nagel 1979: 2).[4,5]

2. Second Hypothesis: Non-Hedonistic Experientialism

Hedonism is only one version of experientialism. Non-hedonistic versions, allowing non-valenced experiences to influence the goodness of a life for the one who lives it, are rarer in the extant literature, but exist nonetheless. According to Richard Kraut (2018), for instance, any experience that results from the realization of our natural potential makes our life better, and our natural potential is not exhausted by the capacity for pleasure. And according to Willem van der Deijl (2019), experiences of self-understanding and novelty make life better in and of themselves; one may or may not *enjoy* having experiences of self-understanding and novelty, but even if one does, the (constitutive, non-instrumental) contribution such experiences make to the goodness of one's life is not *exhausted* by the fact that one enjoys having them. Although different in important ways, Kraut's and van der Deijl's accounts are both forms of non-hedonistic experientialism.

Non-hedonistic experientialism has a ready explanation of our zombie datum – the same explanation, in fact, as the hedonist's. Since a zombie has no experiences, and only experiences make a life better or worse for the one who lives it, a zombie's life cannot be good or bad for the zombie.

At the same time, non-hedonistic experientialism has the advantage that it does not return counterintuitive results in other cases considered above. It explains (and rationalizes) our preference for being Vulcanized over being zombified, as well as for a tolerably negatively-valenced life over a zombie life. We prefer to be Vulcanized because in our Vulcanic state we would still be able to have experiences that enhance the value of our lives to us; and we prefer a slightly negatively valenced life over a zombie life because we count on our non-valenced experiences to tip the balance of overall value. Thus non-hedonistic experientialism may seem like an improvement over hedonism, while offering just as good an explanation of the zombie datum.

But the hedonist may claim that the non-hedonistic experientialist gets wrong other cases. To wit:

(*Immortal Vulcan*)

> Suppose God's latest offer to you is this. You can live 900 years of conscious bliss, full of joy, meaning, and good vibes; *or* you can be turned into an immortal Vulcan, experiencing an eternity of perceptual, intellectual, etc. activities completely devoid of any hedonic or affectively valenced dimension.

If, as the non-hedonist experientialist claims, a Vulcan's experiences do contribute to the goodness of her life (and not just instrumentally), then at least if we assume a linear aggregation function, the goodness of an *infinite* Vulcan life is bound to exceed at some point the goodness of any finite life. It would then follow that you should choose to become

an immortal Vulcan. But if you're anything like me, you'll choose 900 years of conscious bliss in a heartbeat (compare Pummer 2018).

Might the non-hedonistic experientialist simply adopt a non-linear – specifically: asymptotic – aggregation function? This is the kind of function where there are diminishing returns on the wellbeing value that experience generates, with the marginal value tending toward zero. If the non-hedonistic experientialist adopts such a function, she can block some untoward implications of her view.

Now, it is true that it'd be very helpful for the non-hedonistic experientialist to adopt such an aggregation function. But she would have to also *motivate* doing so. Recall that the Vulcan does not experience boredom, since that's a valenced experience. So it's unclear why there should be "diminishing returns" on the value generated by her non-valenced experience. And presumably, if non-hedonistic experientialism were true, then in a world where everybody is a Vulcan, a longer Vulcanic life would be better than a shorter one, other things being equal. So at the very least, it should fall on the non-hedonistic experientialist to adduce some non-ad hoc reason to impose asymptotic aggregation.

In addition, non-hedonistic experientialism faces the same difficulties that hedonism does when it comes to experience-machine-style scenarios. Intuitively, loving and feeling loved are experiences that contribute (non-instrumentally) to the quality of our life, even when love is complicated and involves as much confusion and frustration as orgasmic euphoria. The emotional investedness as such seems valuable to us. But also intuitively, loving one's children or partner seems to make for a better life than loving computer simulations one mistakenly takes to be children and partners; and feeling loved because a real person really loves you makes for a better life than feeling loved because a sniggering scientist is activating the right subpopulation of neurons in your brain (McMahan 2020). Certainly if God approached me with the offer of implementing whichever scenario suits me better, I wouldn't be indifferent – I'd prefer to be really loved. The connection to reality thus appears to make a difference to love's prudential value.

Connection to reality also seems to make a difference, regardless of the involvement of experience machines, to the prudential value of a Vulcan's *cognitive* phenomenology.

(*Vulcan Einstein*)

> Imagine two Vulcan worlds, where two incredibly creative scientific geniuses, Vulcan Einstein and Vulcan Tweinstein, have rich and energetic intellectual lives, with extraordinarily intricate (though affectively muted) cognitive phenomenology. Both eventually come up with the holy grail of physical science – a theory of everything. We may stipulate that the history of science is identical in the two worlds and that the theory that Vulcan Einstein and Vulcan Tweinstein come up with is the same, and

> is supported by the same evidence. The only difference is this: Vulcan Einstein's is the *one true theory* of the relevant world, while Vulcan Tweinstein's is a big mistake – a brilliant construction, just as well supported by the available evidence, but unfortunately completely misdescribing the true structure of Vulcan Tweinstein's cosmos.

Intuitively, Vulcan Einstein's life is a better life than Vulcan Tweinstein's. It is certainly the life we would prefer having, and indeed the life both Vulcan Einstein and Vulcan Tweinstein are *trying* to have. But given that Vulcan Einstein's and Vulcan Tweinstein's phenomenology is strictly the same, it must be some non-experiential element in their lives – presumably, the link to the way the world really is – that accounts for the difference.

A final objection to experientialism, hedonistic and non-hedonistic alike, is what I will call the *Refined-Datum Objection*. Experientialism seemed to us to have a good explanation of why a zombie's life is never worth living. But this depends forsooth on how we take this explanandum. Compare a zombie's life to a mombie's. A mombie is a person with a full and stormy experiential life the net experiential value of which happens to be exactly zero.[6] Now, there is a clear sense in which the mombie's life is not worth living, namely, the sense that at the end of the day its net worth is zero. But the way in which the zombie's life is not worth living seems to go deeper than this. In general, there is a difference between (a) instantiating a quantitative determinable in virtue of instantiating the determinate 0 and (b) not instantiating the determinable at all (see Balashov 1999). 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 problem with experientialism is that its explanation of the worthlessness of a zombie's life to the zombie is the same as its explanation of the worthlessness of a mombie's life to the mombie. It just adds up the zombie's experiences' values and comes up with zero. But there is a way to see this as not explaining everything there is to explain here.[7]

This argument applies equally to hedonistic and non-hedonistic experientialism, of course. What it suggests is that experientialism's explanation of the zombie datum may not be as straightforward as we have suggested. In addition, both hedonistic and non-hedonistic versions of experientialism face a certain embarrassment in experience-machine-style scenarios. And each version returns odd results in some divine-offer cases: hedonistic versions return the results (i) that Vulcanization is in no way preferable to zombification and (ii) that zombification is preferable to a mildly disagreeable life; non-hedonistic versions return the result that, pending justification for non-linear value aggregation, an immortal Vulcan's life is better than any finite life imaginable. For all these reasons, there is room for hope that a better explanation of the zombie datum could be had.

3. **Third Hypothesis:** Experiential Organic Unities

G.E. Moore (1903 §18) famously claimed that the value of certain wholes is greater than the sum of the values of their parts. His chief example is of an experience as of a beautiful object, perhaps the kind of experience Derek Matravers (2003) calls "visual delight," say taken in the Chauvet Cave paintings from 30,000 years ago. Let *e* be the experience of visual delight, *c* the Chauvet Cave paintings, and V the value operator. Then according to Moore, $V(e + c) > V(e) + V(c)$. In this particular instance, Moore seems to suggest, $V(c) = 0$ and yet $V(e + c) > V(e)$. Moore calls value structures such as this "organic unities."

Consider now a view according to which life's goodness – wellbeing – is fixed by two kinds of item: (i) experiences and (ii) organic unities at least one constituent of which is an experience. Call "experiential organic unity" any organic unity at least one constituent of which is an experience. Then according to this view, wellbeing is fully fixed by the combination of (i) experiences and (ii) experiential organic unities. The view preserves something of the experientialist spirit but allows for value that comes ultimately from things that aren't experiences. Significantly, prudential value *fails to supervene on experience*. Still, although things that aren't experiences can contribute non-instrumentally to wellbeing, they can do this only if they enter into organic unities with experiences. On their own, so to speak, they cannot contribute to wellbeing.

Because of its affinity with experientialism, I will call this view "organic-experientialism," or "o-experientialism" for short.[8] O-experientialism, like experientialism, has a choice to make between hedonistic and non-hedonistic versions. A hedonistic version would ground wellbeing specifically in (i) *affectively valenced* experiences plus (ii) organic unities at least one constituent of which is an *affectively valenced* experience. A non-hedonistic version will not include this restriction on the type of experiences relevant to wellbeing.[9]

In either version, o-experientialism has a simple explanation of the zombie datum: the reason a zombie's life is never worth living is that a zombie has no experiences, and without experiences it can't have experiential organic unities either. Thus neither of the wellbeing-promoting elements that o-experientialism recognizes can occur. That is why a zombie's life could never be worth living.

But what makes o-experientialism specially interesting is that, in addition to explaining the zombie datum, it can *also* accommodate the notion, brought out by the experience machine, that the goodness of a person's life seems to depend in part on things that go beyond her experience and pertain to "real value." O-experientialism can allow, for instance, that of two persons who experience the same strength of visual delight, one is benefited more (her wellbeing is augmented more), because what she is delighted with is

really beautiful, whereas what the other is delighted with is in truth ugly; and more importantly for our purposes, that of two persons experiencing a great sense of meaning and fulfillment with their life's work, a Shakespearean character, whose work really is meaningful, has led a better life than a Sisyphean character, whose work is in truth pointless. The reason is that the experiential organic unities in the former's life add more value to their life than the experiential organic unities in the latter's life.

In this way, o-experientialism is perfectly positioned to accommodate both zombie-life-type intuitions and experience-machine-type intuitions, that is, both intuitions that pressure us to exclude the unexperienced from the sphere of wellbeing and intuitions that pressure us to recognize "real value," as opposed to mere experience-as-of-real-value, as a determinant of good living. We get to have our cake and eat it to – courtesy of experiential organic unities.

O-experientialism overlaps in important ways with so-called hybrid accounts of wellbeing (Woodward 2016), first floated by Derek Parfit (1984: 501-2) but developed and defended most perspicuously in Shelly Kagan's "enjoying the good" view (Kagan 2009). Kagan is somewhat non-committal on whether your life can be at all improved by (a) enjoyment taken in worthless things and/or (b) worthy things present in your life but not enjoyed. What matters to him most is the *big boost* to wellbeing that comes from these two things coming together, that is, from enjoyment of what is objectively worthy. This "coming together" is commonly interpreted in terms of organic unities (Woodward 2016: 167-9, Hurka 2019). To that extent, o-experientialism certainly resembles hybrid accounts. However, there are two commitments o-experientialism definitely makes that hybrid accounts as such need not. The first is that nothing can improve or worsen one's life absent an appropriate experience of it. (In hedonistic o-experientialism, an "appropriate" experience would be an affectively valenced experience; in non-hedonistic o-experientialism, it could be any of the other experiences as well.) For instance, the greatest achievements contribute not an iota to the goodness of your life if you don't enjoy them or otherwise experience them in the right way. This commitment is crucial to the o-experientialist's explanation of the zombie datum: once we remove this requirement, a zombie's life could be good for it, provided it included the right "objective goods." The second commitment of o-experientialism that it may or may not share with hybrid accounts is that the right experiences do contribute to a person's wellbeing in and of themselves, regardless of the intrinsic value of their objects. This allows the o-experientialist to return the right results in certain cases. Brad Hooker (2015: 30), for instance, asks whether we should not accept that, of two people who have enjoyed the same worthwhile things to the exact same extent in their lives, but one of whom had had one nicer dream, the life of the pleasant dreamer is not ever so slightly better. Intuitively, we should, and o-experientialism, as formulated above, returns just this result.

The main problem with o-experientialism, however, is that it's just a bit mysterious how these organic unities exactly work. Where does the alleged extra value come from? How does it come to be injected into the world?

To appreciate the mysteriousness here, consider that for *x* and *y* to form an organic unity, they must enter into some relation R, in virtue of which they form the unity, but that the "added value" of the whole does not come from R. Or better put, if it does come from R, then the whole is not an organic unity after all. For then it would be the case that V(*x* + *y* + R) = V(*x*) + V(*y*) + V(R). Anyway this is clearly not what happens in the alleged wellbeing-enhancing organic unities. Suppose Shakespeare enjoys writing *Troilus and Cressida*, such that the independent prudential value of his enjoyment = 10 prudons, the independent prudential value of *Troilus and Cressida* = 0 prudons, while the total prudential value of Shakespeare's enjoyment taken in the writing of *Troilus and Cressida* = 18 prudons. (Prudons are units of wellbeing.) Here the value of the whole exceeds the sum of the values of its parts, but crucially, the extra value in the whole does not come from the relationship between Shakespeare's enjoyment and the writing of *Troilus and Cressida*. What is that relationship? Informally, we may call it the "taken in" relationship – Shakespeare's enjoyment is *taken in* the writing of *Troilus and Cressida*.[10] What matters for our purposes is that it's clearly not the case that the taken-in relation has an independent prudential value of 8 prudons (to make up the difference between 10 and 18). For when a modern salesman takes equal enjoyment in composing his spam emails about refinancing and debt consolidation, the taken-in relation between his enjoyment and his spam-writing is the same. Yet the whole point of appealing to organic unities in this context is that it's supposed to explain why Shakespeare's life becomes better by 18 prudons whereas the salesman's only by 10. So no, the extra value in the experiential organic unity does not come from the relationship between the constituents. It clearly comes, in fact, from the non-experiential relatum, since that is the only variable we vary in the cases that motivate introducing organic unities here. Thus, in the Shakespeare/salesman case, the only thing we vary is that which the subject's enjoyment is taken in: *Troilus and Cressida* is a stunning achievement both as poetry and as a study of the psychology of pride, whereas refinancing spam is neither. And yet it's crucial to o-experientialism's explanation of the zombie datum that the non-experiential relatum's independent prudential value is strictly zero. So, it's zero when unexperienced, but the way these organic unities work, it doesn't suddenly acquire 8 prudons once experienced. Even when experienced, its own prudential value remains null. *It's the value of the whole*, not the value of any of its parts, that's augmented. That's precisely why the whole's value is more than the sum of the values of its parts. The value of the whole increases, and increases because of the nature of its non-experiential part, and yet the value of that part does not increase and remains null. It's all very mysterious.

It may be objected that, however difficult it is to account for theoretically, reflection on concrete cases suggests that this kind of organic unity is common in everyday life. Here is one such case:

> (*Jeff's dream car*)
>
> Jeff has only one goal in life: to get a really awesome car. For the car to be awesome, he needs it to have leather seats, sweet rims, and a kick-ass sound system. He toils day and night as a tuba player in carnivals and bar-mitzvahs to afford all this. Importantly, Jeff has no need for leather seats, sweet rims, etc. outside the context of a car. It's only if these elements come together to form the right whole that the whole becomes valuable to him.

It may be theoretically challenging to give a proper *account* of how such value-wholes work, but cases like Jeff's prove that these wholes do in fact exist.

My response is that, in this scenario, once the seats and rims are integrated into the car, they do have value to Jeff. Their value does *change* as a result of being integrated into the right whole. Accordingly, the whole is not an organic unity in Moore's sense. Compare: carbon monoxide is disvaluable even though neither carbon nor oxygen is disvaluable. But if we take a concrete, specific carbon monoxide molecule, and stipulate that it is disvaluable, then the concrete, specific carbon atom and oxygen atom constituting it *are* disvaluable – these *specific* atoms do some harm. The fact that other carbon and oxygen atoms don't does nothing to show that these ones are likewise harmless.[11]

4. **Fourth Hypothesis:** Experience-Conditioned Value

The offending feature of organic unities is the fact that the parts don't change their value when they come together, and it's just the whole that has some extra value. Can't we simply reject this feature? Can't we, in fact, forget about axiological wholes and simply say that certain things can change their value when they enter into the right relationship with an experience? On this suggestion, even though an experience of visual delight has 10 prudons and an unexperienced beautiful sculpture has 0 prudons, once the visual delight is taken *in* the sculpture, this *changes* the sculpture's (non-instrumental) prudential value to the perceiver to 8, with the result that the experience of the beautiful sculpture improves the subject's life by 18 prudons rather than 10.

Thomas Hurka (1998) has distinguished between two kinds of organic unity, which he calls "holistic" and "conditional." The holistic variety is the structure we considered in the previous section. The conditional one is essentially the one just proposed. The difference between the two is that in a holistic unity the parts don't change their value inside the whole, and the added value accrues *only* to the whole; whereas in a conditional

organic unity the parts do change their value inside the whole, and once that happens the whole as such becomes in fact irrelevant to the value calculus. As a result, a holistic organic unity's value is greater than the sum of values of its parts, whereas a conditional organic unity's value is equal to the sum of values of its parts once they have come together.

One may question whether "organic unity" is the right name for the conditional structure, given that the whole plays no real role in it. But that's just a verbal issue. What matters for our purposes is that Hurka's conditional organic unities, or whatever we choose to call them, offer us a distinct way to try to accommodate the zombie datum. I have in mind a view where things other than experiences can contribute to life's goodness, but only conditionally on being experienced. Call prudential value that is conditional on experience "experience-conditioned value," and call things that *bear* such value "experience-conditioned valuables." Then, on the view I have in mind, a life's degree of goodness to the one who lives it is fixed by two kinds of item: (i) experiences and (ii) experience-conditioned valuables. Call this view "conditional-experientialism," or "c-experientialism" for short.[12] If two persons take comfort and joy in their children's apparent love, but one of them *is* loved by her children while the other is despised and derided, then the c-experientialist can say that the loved person's life is better (provided she counts being loved among the experience-conditioned valuables and comfort or joy among the value-conditioning experiences).

Experientialism and o-experientialism faced a decision about *which experiences count*: there was a hedonistic version that counted only affectively valenced experiences and a non-hedonistic version that counted also non-valenced experiences. C-experientialism faces a *double* decision here: one concerning which experiences have independent prudential value, the other concerning which experiences may condition the value of things that aren't experiences. It is natural, perhaps, to take these two sets to be coextensive, or at least for the latter to be a subset of the former; otherwise it's hard to explain why something that has no independent prudential value can confer prudential value in something else is otherwise prudentially valueless. Still, there is no overt inconsistency in positing value-conditioning experiences that are not themselves valuable. These are issues for a c-experientialist research program to resolve.

C-experientialism explains the zombie datum fairly straightforwardly: the reason a zombie's life is never worth living is that a zombie has no experiences, and without experiences no experience-conditioned value can occur either. At the same time, c-experientialism can accommodate experience-machine-type intuitions. The reason Shakespeare's life is better than Sisyphus', even if they experience the same amount of satisfaction and fulfillment from their life's work, is that the products of Shakespeare's work have greater experience-conditioned value than the products of Sisyphus' work; the reason a life in the "real world" is better than an experientially indistinguishable life in the

experience machine is that the former includes more experience-conditioned value than the latter; the reason Vulcan Einstein's life is better than Tweinstein's is that it contains more experience-conditioned valuables; and so on. Here too, then, the real success is not just in explaining the zombie datum, but explaining it despite rejecting the supervenience of wellbeing on experience.

Notice that c-experientialism is extensionally indistinguishable from o-experientialism, in that it returns all the same verdicts in specific scenarios. There is thus no *normative* difference between them. The only difference is at the level of the metaphysics of value they presuppose. For some purposes, then, the difference won't matter much. From the perspective of rendering intelligible the "axiological mechanics" of the view, however, it seems to me to make a big difference.

An unlovely feature of c-experientialism, which it shares with o-experientialism, is the unprincipled pluralism of sources of prudential value it implies. As someone who believes that we experience our experiences not only in the sense that we smile our smiles and dance our dances but also in a more substantial sense grounded in the fact that all conscious experiences are self-representing (Kriegel 2009), I could generate more unity here by claiming that experiences themselves are experience-conditioned valuables (compare Pallies ms), so that all and only prudential valuables turn out to be experienced-conditioned valuables. There would still be an open question, however, whether experiences are prudentially valuables *only because* they are experienced, or also simply in virtue of being experiences. I will leave this question unresolved here because I think that even if c-experientialism is saddled with unprincipled pluralism about sources of prudential value, that is a relatively minor strike against it. If that is the price of returning the right results on all of the cases we have discussed, it is a price well worth paying.

What really matters for whether c-experientialism represents progress over o-experientialism, it seems to me, is the question of whether conditioned value is less mysterious than organic unities of value. I am not sure how to go about assessing the matter, but perhaps I could start by reporting a gut feeling that, yes, it is less mysterious. It is certainly an advantage that c-experientialism doesn't postulate wholes whose value increases thanks to one of their parts but without that part increasing in value (sometimes without that part even *having* value). Still, there are legitimate questions as to why and how experiences get to condition the value of things other than experiences. Experience somehow "unlocks" prudential value which inheres in things, so to speak, *in potentia*. This category of "axiological potential" seems somewhat mysterious, though perhaps not frighteningly so.

We may press the mysteriousness of c-experientialism as follows (see Bradford 2023). Suppose $X_1, \ldots, X_n$ are such that experiencing them in one's life makes one's life

better for one, and better not just because of the experiencing, but also because of the (experienced) presence in one's life of $X_1, \ldots, X_n$. In contrast, suppose further, anything outside $X_1, \ldots, X_n$ is such that experiencing it in one's life makes no difference to one's wellbeing. Then there is clearly some axiological dividing line between what falls *within* $X_1, \ldots, X_n$ and what falls without. There is something value-y that inheres in $X_1, \ldots, X_n$ and only in them, something which gets manifested when someone experiences some X_i. But if there is something value-y that inheres in $X_1, \ldots, X_n$, it becomes mysterious why the presence of $X_1, \ldots, X_n$ in a zombie's life would not make that life more valuable than a zombie life devoid of $X_1, \ldots, X_n$. If the things that fall within $X_1, \ldots, X_n$ are objectively different, in an axiologically significant way, from things outside $X_1, \ldots, X_n$, then you'd think that their objectively adorning some lives and not others would make an axiological difference to those lives.

As Bradford points out, this difficulty derives from a more fundamental tension in the wellbeing literature. On the one hand, a majority of wellbeing theorists posit welfare goods that go beyond conscious experience (often on the strength of experience-machine-type considerations). On the other hand, a majority of wellbeing theorists also hold that only conscious ("sentient") beings are *welfare subjects*. But if some welfare goods are objective, why wouldn't a living being that instantiated them in their life not be a welfare subject?

This point connects with our "refined datum," the idea that what needs explaining is why a zombie's life has *no* value to it as opposed to just *zero* value. Note that c-experientialism, as developed so far, doesn't really help with this. On the contrary, the c-experientialist's explanation of a zombie's life's worthlessness to the zombie is not substantially different from her explanation of a mombie's life's worthlessness to the mombie. In both cases, we have just added up the units of prudential value – all the experiences and all the experience-conditioned valuables in the zombie's and mombie's lives – and have come up with zero. To all appearances, then, something remains unexplained here.

At the same time, in §6 I am going to argue that (c-)experientialism *can* provide a more satisfying, if more complex, explanation of the refined datum. And the level of mysteriousness involved in conditioned value strikes me as considerably lower than that involved in organic wholes, mindful though I am that we have no methodological canon for evaluating such claims. The fact that c-experientialism can return the right results on *all* the cases we have considered, and yet is not embroiled in the excessive mysterianism of o-experientialism, makes me prefer it over all the views we have considered; and indeed over the views we are about to consider.

5. **Fifth Hypothesis:** Experientialism about Wellbeing Subjects

It's one thing to claim that only experiential goods are welfare goods, another to claim that only experiencing subject are welfare subjects. The latter claim can be defended independently of the former (van der Deijl 2021, Lee ms). Let us call the view that all and only experiencing subjects are welfare subjects – in other words, that all and only conscious creatures are the kinds of thing that can have wellbeing – *WS-experientialism*. Importantly, WS-experientialism is *not* a version of experientialism. It can't be, since these are theories *of different things*: one is a theory of what determines an entity's level of wellbeing, the other a theory of which entities *have* a level of wellbeing to begin with. On the face of it, WS-experientialism has a nicely minimalistic explanation of the zombie datum: a zombie's life is never worth living because a zombie is not a conscious creature and so does not *have* a life that can be good or bad to the one who has it. This is a "minimalistic" explanation in that it involves no commitments on what makes life good or bad. Its only commitment concerns what kind of thing can *have* a life that's good or bad, and that is really all the zombie datum is directly concerned with (the datum, after all, is that a certain kind of thing – the zombie – doesn't have that kind of life). Insofar as modesty is one of the theoretical virtues by which we evaluate competing hypotheses, WS-experientialism's superior modesty speaks in its favor.

The *main* virtue of WS-experientialism, however, is that it explains nicely the "refined" datum that a zombie's life does not instantiate the wellbeing determinable. Since a zombie isn't the kind of thing whose life can be good or bad (to it), it doesn't instantiate *any* wellbeing-level property, not even the property *having a wellbeing level of zero*. In this way, WS-experientialism's explanation of the worthlessness of the zombie's life is different from its explanation of the worthlessness of the mombie's life.

The main problem with WS-experientialism, however, is that the question of which kinds of thing can have a good or bad life does not seem to be independent, at the deepest level, from the question of what determines the goodness or badness of a life. On the contrary, it's very natural to *derive* one's view on who the wellbeing subjects are from one's view on what wellbeing consists in. The derivation mechanism is simple: For any view according to which wellbeing is fixed by $x_1, \ldots, x_n$, we say that an entity E is a wellbeing subject just if E is capable of instantiating any of $x_1, \ldots, x_n$ in its life. For instance, if the hedonist is right that a life is good to the extent that it contains experiences that are pleasant rather than unpleasant, then it's natural for her to hold that the kind of thing that can have a life that's good or bad is the kind of thing that can have experiences that are pleasant or unpleasant.

But if this is how we end up with an account of wellbeing-subjecthood, then the truth of WS-experientialism itself would be explained by the truth of some substantive account of wellbeing goods, e.g. by the truth of hedonism, non-hedonistic experientialism, o-experientialism, or c-experientialism. Each of these *generates* WS-experientialism through

the derivation mechanism just considered. And so the WS-experientialist explanation of the zombie datum, including the refined zombie datum, doesn't really stand on its own, but on the contrary must be traced back to one of those four theories of wellbeing.

For WS-experientialism to offer a genuinely independent explanation, it must find a way to ground *x*'s status as a welfare subject in facts that don't concern what would make *x* fare well or ill.[13] Since WS-experientialism claims that only experiencing beings are welfare subjects, it is natural for it to claim that *x* is a welfare subject *because* it is an experiencing subject. In principle, this may then be combined with any number of views on what would determine how well *x* fares. Perhaps some objective goods, such as achievement or knowledge, make *x*'s life better when *x* is an experiencing subject but don't make any difference to non-experiencing subjects, since the latter don't have what it takes (according to WS-experientialism) to have a life that can go well or badly for them.

The question arises, however, of *why* experiencing makes *x* a welfare subject. Of course, if experiencing makes *x* a welfare subject because experiences make *x* fare better or worse, then we can understand why it is that experiencing is what makes *x* a welfare subject. But then welfare subjectivity is grounded in the presence of welfare goods after all. What WS-experientialists have to do, in order for their explanation of the zombie datum to be independent of the views we have discussed on welfare goods, is to tell some story about why experiencing makes *x* a welfare subject that does not invoke the welfare goods that experience delivers; or else give us a reason to believe that the fact that experiencing makes someone a welfare subject can be the kind of fact for which there is no explanation – a brute and inexplicable bedrock truth about prudential value. It is unclear, at least to me, how either option could be defended.

This doesn't mean that our detour through WS-experientialism has been useless. On the contrary, it helps us see that the four theories of wellbeing we've considered *can* explain the refined datum, albeit indirectly – in two steps, so to speak. Suppose c-experientialism is true and wellbeing is fixed by (i) certain experiences and (ii) certain experience-conditioned valuables. Then using the above derivation mechanism, we can derive from c-experientialism the view that a wellbeing subject is any entity capable of having the relevant experiences and/or the relevant experience-conditioned valuables in its life. Since a zombie is capable of having neither, a zombie is not the kind of thing whose life can be good or bad. And so the zombie's life has *no* wellbeing level – not even the zero level. Refined datum explained.

So far, c-experientialism strikes me in fact as the best among our options. Like the other options, it can explain the zombie datum. But unlike experientialism, it makes room for things that go beyond experience to make a difference to how well a life has gone; unlike o-experientialism, it does not involve the axio-mereological mysteries of (holistic) organic

unities; and unlike WS-experientialism, it has an *explanation* for why all welfare subjects are conscious beings. C-experientialism does have its own debt, though: to explain how and why experience manages to "unlock" prudential value that inheres "*in potentia*" in things that are not experiences.

There is also a sixth course of "dialectical action" we should consider: *denying the datum* rather than trying to explain it. On this approach, the intuition that a zombie's life is not worth living is an irrational residue of a commonsensical but misguided understanding of the good life. In truth, if God offered you to become a rich and handsome zombie, you should jump on this generous offer, not for anybody else's sake but from an entirely selfish standpoint. ("Rich and handsome" stands here for whatever non-experiential good you might want in your life; you may replace it with anything else, and I hope that you do. More on this below.)

6. A Sixth Option: Denying the Datum

I should confess upfront that I find it strictly unfathomable that I should jump on the opportunity to become a rich and handsome (or whatever) zombie. Being the conscious subject I am, zombification seems to me tantamount to death, and *choosing* zombification tantamount to suicide. As long as my life seems to me worth living, then, I would prefer it over zombification regardless of what the zombification comes with.

Shelly Kagan (2019: 28) tries to jog our empathy toward galactic zombies with the following vignette about robots on a faraway planet (keep in mind, as you read it, that it's *stipulated* that these robots are zombies):

> Imagine that you are an Earth scientist, eager to learn more about the makeup of these robots. So you capture a small one – very much against its protests – and you are about to cut it open to examine its insides, when another robot, its mother, comes racing up to you, desperately pleading with you to leave it alone. She begs you not to kill it, mixing angry assertions that you have no right to treat her child as though it were a mere thing, with emotional pleas to let it go before you harm it any further.
>
> Would it be wrong to dissect the child?

The obvious answer is No. No matter how many experiential terms the vignette is surreptitiously peppered with ("desperately," "angry," "emotional"), and how many automatized projections it counts on from what similar behavior in conscious beings indicates about their likely experiential state, it is strange to think that one is in any way *harming* a collection of metal plates by intervening in its internal organization.

Roger Crisp has suggested to me that while the offer to become a rich and handsome zombie is underwhelming, this may be because wealth and good looks are underwhelming goods (to true philosophers, at any rate!); an offer to become a zombie who accomplishes much of value may tug intuitions more. Suppose you are offered to become a zombie who will transform the face of philosophy with a simple but penetrating argument, or with a trilogy of well-argued and well-written tomes constituting a comprehensive philosophical system that reveals the deep nature of the true, the good, and the beautiful; and who plays the piano like nobody ever has and the violin like nobody ever has (and more!); and who leaves behind an enormous corpus of plays each of which makes Shakespeare look superficial and graceless. Are you not more tempted?

I agree that this is more tempting, but only in the way it is tempting to become a zombie whose desires for world peace and the end of hunger are satisfied. What is tempting in both cases is to sacrifice one's own wellbeing for the sake of what one takes to be of immense intrinsic value. After all, your music and plays will not delight *you* and your trilogy will not make *you* finally grasp the deep natures of the true, the good, and the beautiful (as a zombie, you will grasp nothing, in the relevant sense of grasping – see Siewert 2013, Bourget 2017). The only advantage of these feats being accomplished by your zombified self, rather than by somebody unrelated to you, is that posthumous fame will attach to *your* name. But is fame any better than wealth and good looks?

In a similar vein, Gwen Bradford has pointed out to me that given a chance to choose between just dying and being zombified with the zombie finishing all her projects, she would vastly prefer zombification (see also Bradford 2023). Here too, the intuition is surely sound, but it seems to me that there is in fact no difference between those projects being brought to conclusion by a zombified version of me versus by a zombie entirely unrelated to me or a conscious student. What I value is my projects' products, not the life I would have pursuing them in my zombie state. In the sense of "life" relevant to wellbeing, I contend that my zombified self *has* no life.

I have been defending the zombie datum – the intuition that a zombie's life is never good or bad *for the zombie* – against pressure from competing intuition-pumps. But a completely different way to challenge the zombie datum is to offer a debunking explanation of why we *have* the intuition, thus undermining its trustworthiness. In a debunking explanation, we try to show that what *causes* the intuition that *p* is not the putative fact that *p*, but some other, unrelated set of facts; with the consequence that if the intuition is correct, it is so purely accidentally, as it was never formed sensitively to the fact, if it is a fact, that *p*. In this case, it might be argued that the zombie intuition is a form of self-congratulatory chauvinism only slightly more sophisticated than speciesism (compare Bradford 2023), or more perniciously, a form of gatekeeping designed to justify monopoly

over resources for the conscious, but in any case has its sources in self-serving rather than truth-linked factors.

In response, the main thing I would like to say is that there's a difference between the (plausible) claim that we have some tacit desires to elevate ourselves over other denizens of our world, as well as to control resources for our benefit, and more generally to serve our own self-interest, on the one hand, and the (much more speculative) claim, on the other hand, that it is these desires which causally explain our intuition that the life of non-conscious beings cannot go better or worse for them. The former claim is plausible, but the latter is highly speculative, and on reflection not particularly plausible. For one thing, there are many other things that could cause these intuitions, such as considering various divine-offer scenarios in good faith. Secondly and more probingly, we who might read and write papers such as this one belong to many much narrower groups than the group of conscious beings: we are all adult humans. Mere self-interest would therefore be maximally effective if it produced intuitions to the effect that only adult humans have lives that can go better or worse for them; to focus on zombies is to be willing to share resources with many conscious beings who are not adult humans, which fails to serve *our* self-interest. Finally, the self-interested desires at issue seem more relevant, in the first instance, to intuitions about *moral status*: to serve our self-interest, we need intuitions to the effect that others – e.g., non-conscious beings – do not have a moral statues (they "don't count" morally speaking). Now, it is true that a natural view gives all and only welfare subjects moral status. But it is precisely commitment to this additional proposition that comes under strain when self-interest enters the picture. Slave-holders did not deny that their slaves were capable of doing better or worse; they just thought *it didn't matter* whether they were doing better or worse (or didn't *much* matter, at any rate). So, if we the conscious were so blinded by our self-interest as to allow it to generate in us certain intuitions, they would be intuitions to the effect that the non-conscious don't have moral status. This would leave it open whether they don't because they're not welfare subjects (a highly theoretical inference) or because their welfare doesn't matter (the much more "instinctual" reaction). All these considerations cast considerable doubt over the plausibility of the proposed debunking explanation qua psychologically real causal explanation of our intuition that a zombie's life cannot be better or worse for the zombie. I don't think that's really believable qua psychologically real causal explanation.

In the end, I think the only legitimate way to deny the zombie datum is through a sort of "reversal argument," openly running a modus tollens on our insinuated modus ponens. The kind of argument I have in mind would take as its starting point some theory of wellbeing that appears to make room for well-faring zombies (e.g., an objective-list theory with various non-experiential items on its list); and it would conclude from this theory that zombies can have better or worse lives. Schematically, our modus ponens goes like this: 1)

zombie datum; 2) if zombie datum, then either experientialism or o-experientialism or c-experientialism; therefore, 3) either experientialism or o-experientialism or c-experientialism. The corresponding modus tollens would go like this: 1) Neither experientialism nor o-experientialism nor c-experientialism; 2) if zombie datum, then either experientialism or o-experientialism or c-experientialism; therefore, 3) not zombie datum.

However, all the availability of such a reversal argument shows is that the intuition that a zombie's life is never worth living does not *rationally compel* us to the kind of experience-friendly account of wellbeing considered in §§1-4, in the sense that it does not render other theories of wellbeing *incoherent*. This may be readily conceded, though, as long as it's also accepted that the zombie datum creates independent and substantial dialectical pressure in favor of approaches to wellbeing of roughly the shape we have been examining here.

Conclusion

I hesitate to issue a recommendation at the end of our discussion. A hedonist could embrace the implications that a fulfilled Shakespeare's life is no better than an equally fulfilled Sisyphus' life, that becoming a Vulcan is no better than becoming a zombie, and that becoming a zombie is better than leading a mildly unpleasant life. A non-hedonistic experientialist could embrace the implications that a Truth-revealing Einstein's life is no better than a big-mistake-producing Tweinstein's life, and that it's better to become an immortal Vulcan than to lead any finite life or unbridled joy and fulfillment. An o-experientialist could shrug at the axio-mereological mysteries of organic unities and mutter that the world is full of mysteries. A c-experientialist could welcome a world full of in-potentia value inhering in insentient things and awaiting unlocking by conscious experience. A WS-experientialist could insist that what makes an entity a wellbeing subject is independent of, perhaps even prior to, what determines the entity's wellbeing. And everybody else could scratch their head and say that, come to think of it, maybe one should accept God's offer to become a long-lived, rich, handsome, and incredibly accomplished zombie. The truth is that all of these seem to me highly uncomfortable positions to end up in. As noted, as I contemplate my own levels of discomfort, I find the embarrassments of c-experientialism substantially more tolerable; but perhaps I am unrepresentative in this.[14]

References

- Balashov, Y. 1999. 'Zero-Value Physical Quantities.' *Synthese* 119: 253-386.
- Barker, M. Ms. 'Charged Experience.' Unpublished.
- Block, N.J. 1995. 'On a Confusion About the Function of Consciousness.' *Behavioral and Brain Sciences* 18: 227-247.
- Bourget, D. 2017. 'The Role of Consciousness in Grasping and Understanding.' *Philosophy and Phenomenological Research* 95: 285-318.
- Bradford, G. 2023. 'Consciousness and Welfare Subjectivity.' *Noûs* 57: 905-921.
- Bramble, B. 2016. 'A New Defense of Hedonism about Well-Being.' *Ergo* 3: 85-112.
- Campbell, K. 1970. *Body and Mind*. New York: Doubleday.
- Chalmers, D.J. 1996. *The Conscious Mind*. Oxford and New York: Oxford University Press.
- Chalmers, D.J. 2022. *Reality+: Virtual Worlds and the Problems of Philosophy*. New York: Penguin Books.
- Crisp, R. 2006. 'Hedonism Reconsidered.' *Philosophy and Phenomenological Research* 73: 619-645.
- De Brigard, F. 2010. 'If You Like It, Does It Matter If It's Real?' *Philosophical Psychology* 23: 43-57.
- Griffin, J. 1986. *Well-being*. Oxford: Clarendon Press.
- Hurka, T. 1998. 'Two Kinds of Organic Unity.' *Journal of Ethics* 2: 299-320.
- Kagan, S. 2009. 'Wellbeing as Enjoying the Good.' *Philosophical Perspectives* 23: 253-272.
- Kagan, S. 2019. *How to Count Animals, More or Less*. Oxford and New York: Oxford University Press.
- Kraut, R. 2018. *The Quality of Life: Aristotle Revised*. Oxford and New York: Oxford University Press.
- Lee, A.Y. Ms. 'Consciousness Makes Things Matter.' Unpublished.
- Lin, E. 2021. 'The Experience Requirement on Well-Being.' *Philosophical Studies* 178: 867-886.
- McMahan, J. 2020. 'Review of R. Kraut, *The Quality of Life: Aristotle Revised*.' *Notre Dame Philosophical Reviews*.
- Matravers, D. 2003. 'The Aesthetic Experience.' *British Journal of Aesthetics* 43: 158-174.
- Moore, G.E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.
- Nagel, T. 1974. 'What is it Like to Be a Bat?' *Philosophical Review* 83: 435-450.
- Nagel, T. 1979. 'Death.' In his *Mortal Questions*. Cambridge: Cambridge University Press.
- Nozick, R. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Pummer, T. 2018. 'Lopsided Lives.' *Oxford Studies in Normative Ethics* 7: 275-296.
- Siewert, C. 2013. 'Speaking up for Consciousness.' In U. Kriegel (ed.), *Current Controversies in Philosophy of Mind*. London and New York: Routledge.
- Siewert, C. 2021. 'Consciousness: Value, Concern, Respect.' *Oxford Studies in Philosophy of Mind* 1.
- Silverstein, M. 2000. 'In Defense of Happiness: A Response to the Experience Machine.' *Social Theory and Practice* 26: 279-300.

- van der Deijl, W. 2019. 'Is Pleasure all that is Good in Experience?' *Philosophical Studies* 176: 1769-1787.
- van der Deijl, W. 2021. 'The Sentience Argument for Experientialism about Welfare.' *Philosophical Studies* 178: 187-208.
- Wolf, S. 1997. 'Happiness and Meaning: Two Aspects of the Good Life.' *Social Philosophy and Policy* 14: 207-225.
- Woodward, C. 2016. 'Hybrid Theories.' In G. Fletcher (ed.), *The Routledge Handbook of Philosophy of Well-Being*. London and New York: Routledge.

[1] Obviously, if access consciousness is inseparable from phenomenal consciousness, then when God deprives you of one s/he deprives you of the other; and so in contemplating their offer you'd be contemplating lacking phenomenal consciousness.

[2] Any number of things can contribute causally and instrumentally to the goodness of a life on this view: a good pianist can contribute to the goodness of my life, but only by causing certain auditory experiences in me.

[3] Thanks to Dan Pallies and Galen Strawson for pressing me on this point.

[4] Here I am grateful to Géraldine Carranante and Anna Giustina for making me appreciate this point.

[5] Might the hedonist say that just having a conscious life is something that we value, quite apart from the un/pleasant experiences in it? She could, of course, though it's unclear that she would still count as a hedonist. Labels aside, it's also unclear what advantage there is to stressing the having of a conscious life as opposed to the having of experiences occurring *in* that life? Either way one is distancing oneself from the idea, definitive of hedonism, that wellbeing is fixed only by affectively valenced experience. In addition, it's unclear that having a conscious life is really something distinct from having experiences. Conscious life is not some big container in which individual experiences crawl about. It just is the stream of experiences. Well, so it seems to me at least – the question is vexed.

[6] If one is a hedonist experientialist, this means that the amount of pleasantness and the amount of unpleasantness are exactly equal; if one is a non-hedonist experientialist, it means the amount of unpleasantness is greater than the amount of pleasantness by the exact amount of value generated by the occurrence of non-valenced experiences plus experience as such.

[7] Thanks to Roy Sorensen for suggesting to me this way of putting the difference between the way the zombie's and mombie's lives are not worth living.

[8] There is a close neighbor of o-experientialism, considered in Bradford 2023, according to which the independent prudential value of some things that are not experiences is not zero but just relatively small. On this view, there are three sources of prudential value: (i) certain experiences, (ii) certain non-experiences, and (iii) certain experiential organic unities – with the third category generating the majority of value. The problem with this view is that it won't recover the zombie datum. For the zombie life God offers you could be an eternal life, such that adding an infinity of negligibly valuable items in it would produce a non-negligibly valuable life. Depending on the alternatives, it could then be rational for you to choose zombification. Intuitively, however, that could never be a rational choice.

[9] Strictly speaking, it is also possible, prima facie, to hold that wellbeing-enhancing experiences must be affectively valued but wellbeing-enhancing experiential organic unities may have for their experiential

component non-valenced experiences; or that non-valenced experiences can enhance wellbeing on their own but only valenced ones can enter into the kinds of organic unity that can enhance wellbeing. These views seem prima facie coherent, though poorly motivated.

[10] Arguably, the taken-in relation is just a special case of intentionality, aboutness, or directedness – the case that characterizes the intentionality of enjoyment. But this will not matter for our purposes.

[11] There is another point to make here – a point pertaining to pure axiology, so to speak. I've been careful to stress that when I speak of elements contributing to the good life, I have in mind constitutive, non-instrumental contributions. Arguably, however, Jeff's rims have only instrumental value to him: they help him get what he really values, namely, the awesome car. Although the more standard way *x* acquires instrumental value is by being the *cause* of some *y* that has intrinsic/final value, another way is for *x* to be a *part* or *component* of a *y* with intrinsic/final value. Compare: if knowledge has final epistemic value, and belief is a necessary component of knowledge, then belief also has epistemic value, not intrinsically however, but only insofar as having a belief is instrumental to having knowledge. Likewise, having the right rims is instrumental to Jeff's having the right car, and is worthless to him if it stops being so instrumental. Moorean organic unities are different from this: they are structures of *intrinsic/final* value.

[12] There is also the view that the goodness of a life is fixed entirely by experience-conditioned value, that is, by things that are not (or not necessarily) experiences but contribute to wellbeing when they are the objects of experiences. In this version, experiences themselves do not in general contribute to wellbeing. We can either think of this is a limit case of c-experientialism, where the set of experiences that have independent prudential value is the null set, or think of it is a neighbor of c-experientialism. Much of what I will have to say about the more standard c-experientialism will apply to this view as well.

[13] The other option is to claim that *nothing* grounds *x*'s status as a welfare subject – it is a primitive, bedrock fact. This seems odd for a view that has a strong commitment to who is and who isn't a welfare subject, namely, that it being a conscious creature to be a welfare subject. If it is so systematic a fact that only conscious creatures can be welfare subject, it would seem that their being conscious is relevant to *why* they are welfare subjects.

[14] This paper was shaped by sustained philosophical exchange with three people: Gwen Bradford, Andrew Lee, Geoffrey Lee, Eden Lin, and Charles Siewert. It also benefited from comments on a previous draft by Gwen Bradford, David Chalmers, Roger Crisp, Anna Giustina, Lorenza D'Angelo, Andrew Lee, Geoff Lee, Eden Lin, Dan Pallies, George Sher, Charles Siewert, Roy Sorensen, Galen Strawson, and Willem van der Deijl, as well as from presentations at the Jean Nicod Institute in Paris, the London Mind Group, Rice University in Houston, University of Barcelona, UNAM in Mexico City, at the University of Texas at Austin; I am grateful to the audiences there, in particular Aarón Álvarez Gonzalez, Géraldine Carranante, Filippo Contesi, Brigitte Gill, Manuel Garcia-Carpintero, Anna Giustina, Alex Gzrankowski, Márten Gönöri, Steven Gubka, Andrew Lee, Tricia Magalotti, Michelle Montague, Seyed Razavi, David Sosa, Hamid Taieb, Josh Weisberg, and Nick Zangwill.