CrossMark

Utilitas (2021), **33**, 399–416
doi:10.1017/S0953820821000133

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

A New Argument Against Critical-Level Utilitarianism

Patrick Williamson*

Australian National University, Canberra, Australia
*Corresponding author. Email: patrick.williamson@anu.edu.au

Abstract

One prominent welfarist axiology, critical-level utilitarianism, says that individual lives must surpass a specified 'critical level' in order to make a positive contribution to the comparative status of a given population. In this article I develop a new dilemma for critical-level utilitarians. When comparatively evaluating populations composed of different species, critical-level utilitarians must decide whether the critical level is a universal threshold or whether the critical level is a species-relative threshold. I argue that both thresholds lead to a range of axiological puzzles and objections as yet undiscussed within the literature, and therefore conclude that critical-level utilitarianism should not be taken as a morally plausible welfarist axiology. I show that certain competitive formulations of critical range utilitarianism are subject to the argument too, and that further attempts to relativise critical levels to a particular group or category of welfare bearer (in particular, individual-relative critical levels) are unsustainable.

1. Welfarist population axiology

In population axiology we want to compare a set of outcomes, that is, possible ways the world could be. In each outcome there is a population, that is, the set of morally significant individuals who exist. And in each population there is some distribution of welfare over the morally significant individuals, that is, whichever thing it is that makes an individual life go better rather than worse for the one who experiences it. Consider, for instance, the following two outcomes:

Table 1. Small Utopia and Mass Mediocrity

	Population size	Well-being per individual
Small Utopia	1	100
Mass Mediocrity	1000	1

In this simplified scenario we want to know whether *Small Utopia* is better than *Mass Mediocrity*, or whether *Mass Mediocrity* is better than *Small Utopia.* A population axiology hopes to tell us as much, since a population axiology is a comparative betterness ordering that holds across outcomes even when population size varies.

© The Author(s), 2021. Published by Cambridge University Press

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

This article considers exclusively *welfarist* population axiologies. Welfarists maintain that the comparative betterness ordering depends only on the way in which welfare is distributed across respective outcomes (Arrhenius 2000a: 257). So, welfarists maintain that the comparative betterness ordering does *not* depend on certain non-welfare considerations sometimes thought to be intrinsically or instrumentally valuable, non-welfare considerations such as desert or equality (Greaves 2017: 2). For the purposes of this article, we can remain deliberately ecumenical on the significance of non-welfare features like desert. Perhaps the welfarist only views these non-welfare features as instrumentally valuable inasmuch as they promote positive welfare within a population, or perhaps the welfarist holds these non-welfare features fixed with a *ceteris paribus* clause (Greaves 2017: 2). In other words, perhaps the welfarist simply assumes that all outcomes under consideration are equally good as regards distribution of desert.

Even in the simplified example of *Small Utopia* versus *Mass Mediocrity*, we have made certain assumptions about the nature of welfare consistent with the broader axiological literature. First, we have associated each individual member of any given population with a *lifetime well-being sum*, and assumed that these lifetime well-being sums are interpersonally comparable (Blackorby, Bossert and Donaldson 2005: 3). That is, we have assumed that many statements of the form *Dantès' life in this outcome is better than Mondego's life in that outcome* are meaningful. Second, we have assumed that these lifetime well-being sums are cardinally comparable. That is, we have assumed that it is meaningful to talk about the degree of difference between lifetime well-being sums, and to insist that *Dantès' life is going x-much better in this outcome than Mondego's life in that outcome.* And finally, we have assumed that lifetime well-being sums are measured against a zero level. Lives that exist above the zero level are worth living for a given individual, while lives that exist below the zero level are not worth living for a given individual, in addition to being worse than those lives at the zero level. Despite some degree of nuance in defining the zero level, for now we can simply think of the zero level as a 'neutral level for continuing to live', a level of welfare and a period of time by which we could extend individual lives, where this level of welfare and period of time are equally as good as ceasing to live for that given individual (Broome 2004: 234).

2. Totalism and the Repugnant Conclusion

To begin with, consider totalism, the canonical welfarist axiology:

> *Totalism*: *A* is better than or equal to *B iff A* contains at least as great a total well-being as *B*.

Totalism says that the better (or equivalent) population contains greater (or equivalent) quantities of total welfare. Totalists prefer *Mass Mediocrity* to *Small Utopia*, for instance, since *Mass Mediocrity* contains greater total welfare. This is true even though individual quality of life in *Mass Mediocrity* is not particularly enviable. (Perhaps members of *Mass Mediocrity* spend their short lives living inside cardboard boxes, watching reruns of a television show that is mildly entertaining but nothing to write home about.)

Totalism captures many of our common-sense intuitions about the comparative status of populations. For instance, totalism tells us that it is always good for more well-being to exist rather than less. Totalism also entails that, holding the welfare of an existing population fixed, it is never wrong to add moderately good lives to that population (Parfit 1984: 420). But totalism also entails Parfit's (1984: 388) Repugnant Conclusion.

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

Given the totalist ordering principle, a large population of very-liveable lives is always outranked by a sufficiently large population of barely-liveable lives, since a sufficiently large population of barely-liveable lives contains greater total well-being. Consider, for instance, *World A* and *World Z*:

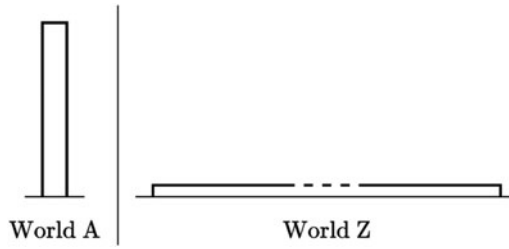


Figure 1. The Repugnant Conclusion.

Suppose for now that *World A* and *World Z* are populations each ten-billion times the size of *Small Utopia* and *Mass Mediocrity*. *World A* contains billions of individuals, and every individual member of *World A* experiences a very high lifetime well-being sum. *World Z* contains billions more individuals, and every individual member of *World Z* experiences a lifetime well-being sum above but arbitrarily close to zero (Blackorby, Bossert and Donaldson 1997: 208). Given these numbers, *World Z* contains far greater total well-being than *World A*. Totalists must affirm that the better outcome contains greater total well-being. Hence, totalists must affirm that *World Z* is better than *World A*.[1]

3. Critical-level utilitarianism

This article aims to test whether or not critical-level utilitarianism is a morally plausible welfarist axiology. The previous detour through totalism and the Repugnant Conclusion was necessary since critical-level utilitarianism represents a rejection of classical totalism, and an attempt to avoid the Repugnant Conclusion. On this view, the Repugnant Conclusion is an unacceptable axiological consequence, and we must therefore develop some axiological modification or invention that implies that populations of barely-liveable lives never outrank sufficiently large populations of highly-liveable lives.

Here is how this strategy promises to work. Blackorby and Donaldson's (1984: 22) original critical-level utilitarianism (CLU) can be stated as follows:

$$V_{CLU} = \sum_{i=1}^{n} (x_i - \alpha)$$

V_{CLU} refers to the value of the population, n refers to the population size, x_i refers to the welfare contained within each individual life i within the population, and α refers to the specified critical level. So this formula reads: the value of a population is equivalent

[1]We might also have arrived at the Repugnant Conclusion via a process of Mere Addition, that is, the iterative addition of lower (but still positive) lifetime well-being sums to a given population (Parfit 1984: 419). I put the issue of Mere Addition to one side, since the Repugnant Conclusion follows straightforwardly from the aggregative principles of total utilitarianism (Arrhenius 2000b: 49).

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

to the sum of all individual contributions to the population. Each individual's contribution to the population is equivalent to that individual's total lifetime well-being sum, minus the specified critical level α. In Blackorby and Donaldson (1984), the assumption is that the specified critical level α exists at some point above the zero level.

Here is a concrete example. Suppose we positioned the critical level at 10 welfare units, with this marker representing the point at which individual lives become robustly worth living for the individuals who live them. It would follow that *Mass Mediocrity* is valued at −9000 units; each of its 1000 inhabitants exist 9 welfare units below the critical level, α. It would follow that *Small Utopia* is valued at 90 welfare units; its single, ecstatic inhabitant exists 90 welfare units above the critical level. Hence, *Small Utopia* would outrank *Mass Mediocrity*.

We see that if the critical level α is set higher than bare-liveability, then barely-liveable lives can never add value to a population. They instead contribute negative value to the world in which they feature. With this in mind, suppose we return to the standard Repugnant Conclusion:

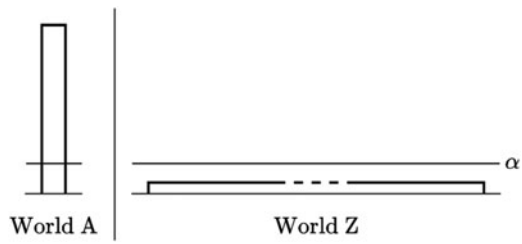


Figure 2. Critical-level utilitarianism and the Repugnant Conclusion.

We see that every life within *World A* exists above the critical value, and hence, contributes positive value to that outcome. We see every life within *World Z* exists below the critical level, and hence, contributes negative value to that outcome. It follows that *World Z* is *not* better than *World A*, and critical-level utilitarians are not subject to the standard Repugnant Conclusion (Arrhenius 2000b: 72). This fact does not change, even if we increase the number of barely-liveable lives featuring in *World Z*. When we do so, we decrease rather than increase the impersonal value contained within *World Z*.

It is already well known that critical-level utilitarianism entails the so-called Sadistic Conclusion (Arrhenius 2000a: 256), in which the addition of a small number of torturous lives to a given population might still be better than the addition of many more positive welfare lives, so long as these positive welfare lives exist below the critical level. In light of this point, we might think that the development of a new argument against critical-level utilitarianism is somewhat redundant – the theory is already unsustainable. It is worth making two brief comments in response to this point. First, every axiology in welfarist population ethics is objectionable in virtue of violating at least one *prima facie* plausible moral adequacy condition (Arrhenius 2000a; Arrhenius 2000b). Hence, the development of multiple arguments against a given axiology may be exactly what is required of us when determining whether the theory is a success or a failure. But second, as we will see, the argument I develop in the following sections applies even to competitive formulations of critical-level theory which escape the standard Sadistic

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

Conclusion. For instance, the argument applies to certain versions of critical range utilitarianism which only entail the Weak Sadistic Conclusion and the Weak Mirrored Sadistic Conclusion (Gustafsson 2020: 8).

All this being said, here are some brief caveats on the scope of this article. First, I restrict my attention to utilitarian formulations of critical-level theories, and of other prominent axiologies more generally, since even these utilitarian formulations are enough to generate some of the distinctive problems that emerge when weighing populations of variable size. This is not meant to imply that these utilitarian formulations are more plausible, say, than their prioritarian counterparts. Second, I don't pretend here to argue against all the possible variations of critical-level theory one might adopt. This goes in particular for critical-level theories which assign no intrinsic moral value to lives which exist below the critical level. Although these sorts of *Neutral Critical Views* (Parfit 2016) might well escape the argument of the following sections, I take them already to entail implausible and unsustainable axiological consequences elsewhere. For instance, neutral critical views violate Arrhenius' (2000a: 257) *Dominance Principle*: on these views, two equally sized populations *A* and *B* might be just as good as each other, even when every individual in *A* is better off than every individual in *B*. In light of these independently problematic consequences I put neutral critical views to one side.

4. Two variations on critical-level utilitarianism

Now for the argument against critical-level utilitarianism. Take it to be the case that, for a wide range of species, welfare is cardinally comparable across species.[2] Consider, then, the following two populations:

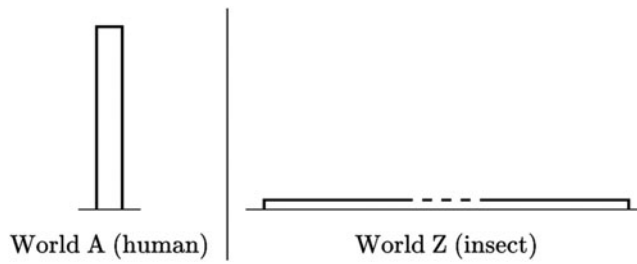


Figure 3. Human versus insect populations.

[2]Here are three pragmatic reasons why the following sections assume (at least some) cardinal interspecies comparisons of welfare. First, those working within welfarist population axiology have regularly expressed openness when it comes to ranking across populations composed of different species (Arrhenius 2000a: 250; Tännsjö 2002: 339; Parfit 2016: 118; Sebo n.d.: 27). For this audience, at least, cardinal interspecies comparisons are likely to come across as unobjectionable. Second, guided by our common-sense folk intuitions, we regularly *do* make at least some interspecies welfare comparisons. We say things like 'my pet ferret is better off than the mouse it is eating' or 'that man is better off than the bug he just squished under his boot'. If we put any weight on these folk comparisons, then this gives us provisional grounds for continuing the following discussion. And third, for the purposes of the following discussion it does not matter if one particular species (some variety of insect, say) turns out to have welfare incommensurable to human welfare. It need only be the case that *some* species have morally significant welfare that is commensurable to human welfare, since this is all it will take to generate the objection to the critical-level view.

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

These outcomes are not dissimilar to those populations featured in the original Repugnant Conclusion. *World A*, for instance, is a human utopia, in which every individual member of the population lives a long and fruitful life, filled with meaningful relationships as well as with intellectual and creative pursuits. On the other hand, *World Z* is a much larger swarm of insects, where each individual insect experiences some minuscule lifetime well-being sum before dying and falling to the ground.[3]

How do we construct a welfarist betterness ranking spanning *World A* and *World Z*? Assuming that the non-human animals in *World Z* really are capable of experiencing morally significant and interspecies comparable welfare, then we might think that it is business as usual; we consult our pre-established axiology (totalism, say), and then find that, given the totalist ordering principle, the better population contains the greater quantity of total welfare. If *World A* contains greater total welfare *World A* is the better outcome. If *World Z* contains greater total welfare *World Z* is the better outcome.[4]

But when it comes to critical-level theories, things are not so simple. Recall that the critical level α is the point at which individual lives begin to contribute impersonal value to the goodness of an outcome, the point at which individual lives become robustly worth living for those who live them. When evaluating populations composed of different species, critical-level utilitarians must make a further specification about the nature of the critical level. They must decide if the critical level remains constant regardless of species membership, or if the placement of the critical-level changes with species membership. We can call this the distinction between *universal critical-level utilitarianism* and *species-relative critical-level utilitarianism*.[5]

We might worry whether *species* is even a natural kind upon which to ground the existence of different critical levels. As we will see later in Section 7, finding relevant groupings (whether in terms of species, breed, individual welfare capacity, and so on) upon which to ground the existence of different critical levels is a task fraught with difficulty. In fact, as I argue in Section 7, the attempt to find sufficiently justified groupings for different critical levels can serve as an argument in and of itself against critical-level theories.

5. Universal critical-level utilitarianism

To begin with, consider the universal view. Suppose that the critical level were an unchanging, cross-species threshold. In other words, assume that the same critical level arbitrated the welfare of all possible individuals, human and non-human in all possible worlds. When comparatively evaluating *World A (human)* and *World Z (insect)*, for instance, the universal critical level would look as follows:

[3]For a similar interspecies comparison, see Sebo (n.d.: 27).

[4]This answer suffices for now, though it does neglect the possibility that one might hold pre-established views from the domain of environmental ethics on what *kinds* of entities have morally considerable welfare, and on the way in which these different kinds of welfare must be evaluated. Speciesists, for instance, privilege the welfare of fellow species members, and hence might be tempted to assign some kind of special and overriding priority to the human welfare featuring in *World A (human)*.

[5]Although this terminology is new, in articulating their theory, Blackorby, Bossert and Donaldson (1997: 217) once suggested it ought to be interpreted along what I have called species-relative lines.

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

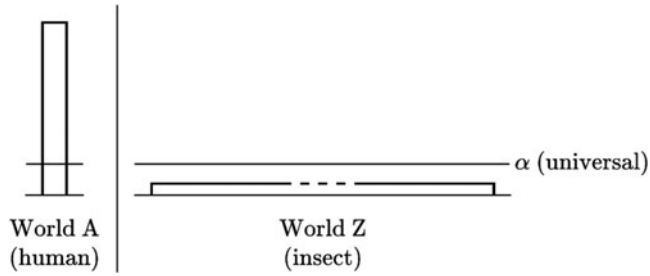


Figure 4. Universal critical-level utilitarianism.

World A (human) is above the universal critical level, while *World Z (insect)* is below the very same universal level. Hence, every life in *World A (human)* contributes positive value to the population in which it features, while every life in *World Z (insect)* contributes negative value to the population in which it features. Given these facts, *World A (human)* outranks *World Z (insect)*.

The universal critical level fits quite naturally with a certain interpretation of critical-level utilitarianism. We might think that only a certain type of life is capable of adding value to a population – lives that contain, say, certain higher order goods like rational discourse or self-awareness through time, the appreciation of beauty, or perhaps the ability to write novels or paint watercolours (Griffin 1986: 340). The universal critical-level reflects this intuition; it represents the point at which certain higher order goods typically obtain, and the point at which individual lives begin to flourish rather than merely subsist. It turns out that for some species this critical level constitutes a trivially low bar, and for other species this critical level constitutes an impossibly high bar, but the same bar remains in place for all creatures irrespective of secondary facts like species membership.

There is also a certain simplicity in the universal view. Species membership does not directly factor into the comparative goodness or badness of different populations. If creature x exists above the universal critical level they make an outcome better, and if creature x exists below the universal critical level they make an outcome worse. Just as non-welfare facts (shoe size, say) did not influence the betterness ordering before, non-welfare facts like species membership need not influence the betterness ordering now.

In spite of these intuitive reflections, it seems there are strong reasons for rejecting the universal critical level. The first reason is the general consideration that, for many of us, the universal critical level will not quite reflect the weight that ought to be assigned to non-human animal populations. For instance: if the critical-level level represents the point at which human lives become robustly (rather than barely) worth living for the ones who live them, then there will be many non-human animals with shorter lifespan and comparatively limited welfare capacities, which are probably *incapable* of reaching the universal critical level. It would follow that these non-human animal lives are incapable of making the world a better place, even when living the best life it is possible for them to live.[6] In fact, the mere existence of such non-human animals would be

[6]On this view perhaps the same point would follow for other welfare bearers, for instance, humans with reduced welfare capacities.

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

axiologically worse than their non-existence, since any life that exists below the critical level detracts value from the population in which it features.

We might worry, in other words, that the universal critical level commits the sin of axiological elitism. Although the universal critical level might certainly tell us that it is better for a given animal life to go better rather than worse for the animal who lives it, it also tells us that the mere existence of many non-human animals, even when they live the best lives it is possible for them to live, could never be as axiologically valuable as their non-existence. Of course, one might reply that this is simply consistent with the theoretical rationale underpinning the universal critical level given only a few paragraphs ago: the universal critical level represents the point at which certain higher order goods typically obtain, the point at which individual lives begin to flourish rather than merely subsist, and many non-human animals are simply incapable of reaching this point, given the particular biological and neurological facts governing their existence. But something does not seem quite right in this reply. It is one thing to say that the universal critical level represents the point at which individual lives begin to flourish, and that some non-human animals are incapable of reaching this standard; it is another thing altogether to say that animals which exist below this standard make the world an actively *worse* place, and that such lives should not exist rather than exist. As the theory stands, the universal critical level utilitarian is committed to both claims.

This leads us, in fact, to the second reason for rejecting the universal critical level. We can call this the problem of *disproportionate disvalue*. Suppose we return to the example of insect and human welfare. Suppose the average human life contains up to *10,000,000 welfare units*, the average insect life contains up to *100 welfare units*, and the universal critical level is placed at *100,000 welfare units*: combine these two facts and we discover that the critical value of an insect life is *−99,900 welfare units*. That is, the creation of a single insect life offsets the creation of an entire human life containing *199,900* welfare units. Back in our ignorant days we might have thought it was a fairly benign thing whenever a swarm of mosquitos hatched. Not so, says the universal critical level. The universal critical level tells us that this is a cataclysmically bad outcome, enough to outweigh the creation of entire human cities containing extremely high levels of positive welfare.

The immediate response is that these were only dummy numbers, and that there is nothing stopping us from lowering the critical level as much as we want, until we have mitigated the problem of disproportionate disvalue. Only there *is* something stopping us from lowering the critical-point as much as we want: if we lower the critical level too far, then we once again allow that barely-liveable human lives might add value to a population, and we once again fall subject to the standard Repugnant Conclusion. So, we can frame this as a dilemma for the universal critical level: either we maintain a high universal critical level, in which case we face the problem of disproportionate disvalue, or we lower the universal critical level, in which case we face the standard Repugnant Conclusion.

One possible response to the problem of disproportionate disvalue, already employed frequently enough by critical-level utilitarians elsewhere (Blackorby, Bossert and Donaldson 1996; 1997), involves postulating a universal 'critical range'. On this view, we suppose that there is a band of possible α locations, that is, a band of possible points at which the critical level α might exist. We say that an outcome A is better than (equal to) B if A is better than (equal to) B for every possible α point within the critical range. Otherwise, we say that A and B are incomparable under vagueness, that is, there

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

is simply no fact of the matter as to whether *A* is better than *B*, whether *B* is better than *A*, or whether *A* and *B* are equally good (Broome 2004: 185).[7]

Now, how would this critical range strategy promise to defuse the problem of disproportionate disvalue? Well, if there exists a broad critical range, containing many possible points at which the universal critical level might exist, then even if certain non-human animals were unable to reach the original critical level, it may still be the case that their lives exist within the *universal critical range*:

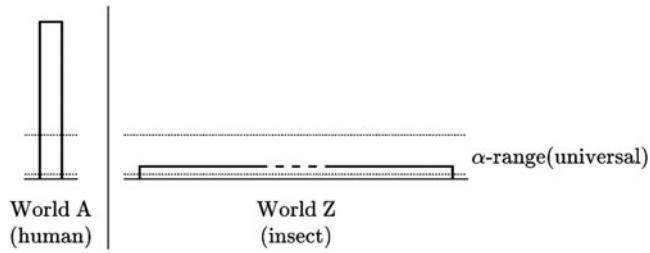


Figure 5. The universal critical range.

Defenders of the universal critical range would then be able to mitigate the problem of disproportionate disvalue as follows.[8] They could first say that *World Z (insect)* contains negative value for some possible α locations within the critical range, and that *World Z (insect)* contains positive value for other possible α locations within critical range. They could second say that *World Z (insect)* contains the lesser value for some possible α locations within the critical range, and contains the greater value for other possible α locations within the critical range. They could third say that *World A (human)* and *World Z (insect)* are incomparable under vagueness, since each population contained the greater value for different possible values of α within the critical level. Hence, the disproportionate disvalue found in a population like *World Z (insect)* is negated; it is replaced with comparative indeterminacy in value.

But this strategy seems unlikely to succeed for the following reasons. First, we would need an extremely broad critical range in order to catch, say, the disproportionately disvaluable lives featuring in *World Z (insect)*, and render them only indeterminately valuable. Perhaps we would have to stretch the critical range down to the zero level, such that no insect life contributed determinate value to the outcome in which it featured. This flies in the face of the whole theoretical justification we gave for the universal critical level in the first place – namely, that the universal critical level represents the point at which individual lives become genuinely and robustly liveable in virtue of containing

[7]Alternatively, one can employ a critical range that entails incommensurability rather than vagueness: *A* is better than (equal to) *B* if *A* is better than (equal to) *B* when every point within the critical range is taken to be the value of α. If not, *A* and *B* are incommensurable, that is, *A* is not better than *B*, *B* is not better than *A*, and *A* and *B* are not equally good (Arrhenius 2000b: 75; Blackorby and Donaldson 1984: 21). I restrict my attention to the vague critical range because appeals to incommensurability seem inappropriate in this context: all we are doing is weighing different distributions of the same good, welfare, against one another (Arrhenius 2000b: 77).

[8]They could mitigate the problem for at least *some* multispecies comparisons, comparisons like *World A (human)* and *World Z (insect)*. Perhaps there are other multispecies comparisons for which the critical range strategy would do little.

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

some combination of higher-order goods, thus contributing positive value to an outcome. Wherever this point is, it certainly does not exist at or below the zero level.

More generally, though, this strategy is *ad hoc* because the vague critical range is being used for a purpose for which it was not designed. The vague critical range is meant to reflect the intuition that there is genuine vagueness regarding the point at which the critical level exists, in other words, that there is genuine vagueness regarding the point at which individual lives begin to contribute positive impersonal value to the world in which they feature. But now, the vague critical range is being extended however far is necessary in order to block those betterness orderings that we find intuitively unacceptable. The consequence of using the critical range for this purpose will lie in having a betterness ordering infected with widespread indeterminacy at places where such indeterminacy seems inappropriate; when asked whether *World A (human)* is better than *World Z (insect)*, or whether *World Z (insect)* is better than *World A (human)*, we are left to answer that there is simply no fact of the matter, and that the two populations cannot be determinately weighed against one another, in spite of the fact that many of us will share the exact opposite ethical intuition: that *World A (human)* is clearly and determinately better than *World Z (insect)*. For those of us who share this intuition, the universal critical range does not constitute an adequate response to the problem of disproportionate disvalue.[9]

6. Species-relative critical levels

This leaves us with the species-relative view, on which the critical level exists at different levels for different species in different possible populations. It is not difficult to provide an intuitive rationalisation for species-relative critical levels. Different biological facts govern the welfare of different species – humans have long lifespans compared to insects, and giant tortoises have long lifespans compared to humans. These biological facts mean, among other things, that the lifetime well-being sums of different species will tend to exist in different ranges. It is hence appropriate that we furnish each species under consideration with its own species-relative critical level, where these species-relative critical levels are calculated on a case-by-case basis, and where these species-relative critical levels are based on the relevant biological facts governing the welfare of individual species members (Blackorby, Bossert and Donaldson 1997: 217).

(As I mentioned previously, we might worry whether *species* is even a natural kind upon which to ground the existence of different critical levels. This is a concern to which I return in the next section. I don't mean to imply, either, that the problems emerging within the following section are unique for our evaluations of non-human animal welfare. Critical-level utilitarians might face similar problems when comparing between exclusively human populations, since even human populations vary considerably in welfare capacity in virtue of, say, differences in lifespan or certain mental impairments.)

[9]Even after all this we have not touched on a further point, namely, that the universal critical level seems to entail a kind of 'Animal Sadistic Conclusion'. On the universal view, for instance, the addition of many positive welfare animal lives (which exist below the critical level) might be worse than the addition of fewer torturous human lives. Or alternatively, the addition of many positive welfare human lives (which exist below the critical level) might be worse than the addition of fewer torturous animal lives. I leave this point as only an addendum since the result is substantially the same as the standard Sadistic Conclusion, already entailed by critical-level utilitarianism. Thanks to an anonymous reviewer for emphasising this point.

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

On this species-relative view we avoid both the problems of axiological elitism and disproportionate disvalue. We avoid the problem of axiological elitism because the worth of each species member is found by consulting its own species-relative critical level, rather than by consulting some universal critical level, a universal critical level which will inevitably exist above the maximum possible welfare levels available to certain species. We avoid the problem of disproportionate disvalue for the same reason, namely, because we find the worth of each species member by consulting that species' own custom-made critical level.

Now, how might we go about setting the critical level for any given species? The function of the critical level remains the same: the critical level marks the point at which individual lives begin to make the world an impersonally better place rather than an impersonally worse place. We had a useful strategy for locating the critical level α in the case of humans: the Repugnant Conclusion told us that the critical level must exist somewhere above bare-liveability (*World Z*), but somewhere beneath robust-liveability (*World A*). But things are perhaps not so simple in the case of non-human animals. For instance, I do not claim to know what it means for an insect life to be 'robustly-liveable' versus 'barely-liveable' for the insect who lives it, or whether these distinctions even apply to creatures like insects with extremely limited lifespan and welfare capacities.

Blackorby, Bossert and Donaldson (1997: 217) briefly speculate whether one might set the critical level for non-human species at some fraction of the total welfare typically experienced by members of that species in the wild. Suppose we embellish this account with some more details. Suppose we set the critical level at *0.5 positive welfare capacity*, and suppose the typical deer has the capacity for *1000 welfare units*. It would follow that the critical level for deer is *500 welfare units*. Deer that exist below *500 welfare units* make the world a worse place, and deer that exist above *500 welfare units* make the world a better place.

I doubt whether this approach is plausible. The level of welfare typically experienced by animals in the wild is merely an empirical and descriptive fact. It may turn out that, due to environmental factors, North American deer have miserable lives filled with fear and that end prematurely and violently through predation, while European deer have ecstatic lives filled with foraging and berry-eating. Does it follow that North American deer have a low critical level, while European deer have a high critical level? This result is counterintuitive: a deer is a deer, and the value within a deer life should remain constant regardless of arbitrary environmental factors. It seems we must rely on some other method for setting non-human species-relative critical levels.[10]

I suspect the right question to ask is simply, *at what point do positive deer lives begin to make the world better, impersonally speaking?* I do not know where this point lies. It is enough for now to stipulate that such a point does exist, and that this point is the critical level. For some species (humans, say) this point exists somewhere above the zero-level; for other species with very limited welfare capacities (insects, say) it would not be surprising if this point were equivalent to the zero level.[11] All this, to emphasise,

[10]The proposal noted by Blackorby, Bossert and Donaldson (1997) closely resembles the *Species Norm Account* of individual fortune outlined in McMahan (1996: 10), on which 'how well off a being is depends on how its level of well-being compares to the levels accessible to those with cognitive and emotional capacities that are the norm for the species of which the being is a member'.

[11]For instance, it might sometimes be appropriate to set a critical level at the zero level for 'animals with rudimentary brains' which 'exist in the present only' (Blackorby, Bossert and Donaldson 1997: 217).

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

has still only been an initial investigation of the *way* in which we might go about setting species-relative critical levels, provided we find the species-relative interpretation appealing in and of itself.

So far we might think that the species-relative view has a certain intuitive pull: there are different species that have the capacity for different ranges of lifetime well-being sums, and it is only appropriate that we measure the welfare of these different species against different standards, that is, against different critical levels. It would not be appropriate to measure deer welfare against a human critical level, or human welfare against a deer critical level. But as with the universal view, I maintain there are good reasons for rejecting the species-relative interpretation of critical-level utilitarianism.

First, the species-relative view paves the way for an animal-based version of the Repugnant Conclusion – the Animal Repugnant Conclusion, we might call it – and hence undercuts the motivation for adopting critical-level utilitarianism in the first place.[12] To illustrate, suppose we grant that human lives are measured against a substantive positive critical point, while insect lives are measured against a zero-level critical point (or a critical level extremely close to the zero level). It would follow that, for each population of well-off humans, there exists some (sufficiently large) possible population of insects which is better in virtue of containing greater total value above the relevant critical level:

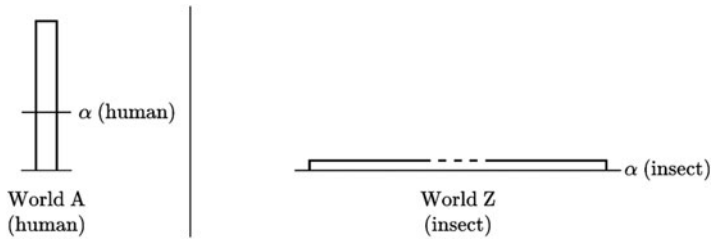


Figure 6. The Repugnant Conclusion for species-relative critical levels.

To briefly summarise the reasoning: *World A (human)* is a human utopia, and *World Z (insect)* is a large swarm of insects fluttering in and out of existence. Provided that *World Z (insect)* is sufficiently large, the species-relative critical-level utilitarian is forced to admit that *World Z (insect)* is better than *World A (human)*, since *World Z (insect)* might always contain the greater aggregated critical value. *World Z (insect)* might always contain the greater aggregated critical value since each individual life within the population exists above the insect-specific critical level, even though each individual life only contains a small quantity of positive welfare.

For anyone who found the original Repugnant Conclusion genuinely repugnant, the Animal Repugnant Conclusion is unlikely to be good news. If it was repugnant to admit that any human utopia can be bettered by some sufficiently large population of barely-liveable human lives, it should be similarly repugnant to admit that any human utopia can be bettered by some sufficiently large population of mediocre insect lives.

[12]Previous references in the literature to an animal-based Repugnant Conclusion include Parfit (2016: 118) and Sebo (n.d.: 27). As far as I am aware, neither author considers that the Animal Repugnant Conclusion might be employed as an argument in particular against critical-level utilitarianism.

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

Importantly, although we have picked insects as our examples to play the *World Z* role, the Animal Repugnant Conclusion is a generalised result: any non-human animal could play the *World Z* role, so long as we had some reason for believing that that species possessed interspecies comparable welfare. (For this reason, it would matter little if, for example, insects themselves turned out to possess non-interspecies comparable welfare.)

Further, the species-relative view forces us to accept one other unexpected consequence. We must now admit that any insect population which exists just above the insect critical level will inevitably outrank some large population of human lives, all of which exist *just* below the human critical level. A single mayfly is better than a large human civilisation, in which everyone experiences positive lives that *almost* reach the human critical level. We can call this result the Flipped Repugnant Conclusion:

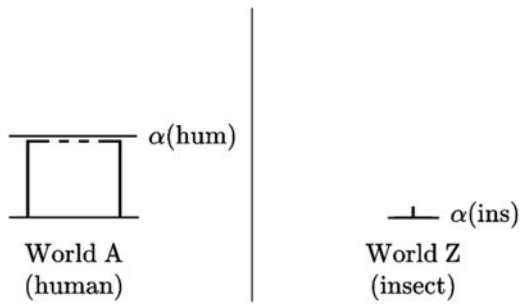


Figure 7. The Flipped Repugnant Conclusion.

The Flipped Repugnant Conclusion is an especially implausible axiological result. Simply put, it does not capture our intuitions about how different quantities of positive welfare should be weighed against one another.

We face the Flipped Repugnant Conclusion because we employ different critical levels for individuals within different populations. Accordingly, our populations do not have the comparative weight we expect them to have, and a single insect might sometimes outweigh an entire human civilisation. A small quantity of positive welfare might have positive weight when it constitutes the lifetime well-being sum of some species members; that same quantity of positive welfare might have great negative weight when it constitutes the lifetime well-being sum of a member of some other species. The quantity of welfare remains the same, but the impersonal value assigned to that welfare changes based on a further fact, namely, the sort of life in which that welfare is found.

As before, we could try to resolve these puzzles and problems by employing critical ranges under vagueness. The details of this strategy become slightly more complicated when dealing with a multiplicity of species-relative critical levels. We say that for every species x, there is a species-relative critical range, a band of possible points at which the critical level α might exist. One population A comes out better than or equal to another population B *iff* A comes out better than or equal to B for every possible α point within every possible *species-relative vague critical range*. Otherwise, the two populations are incomparable under vagueness. On this view, we rewrite the above Flipped Repugnant Conclusion scenario as follows:

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

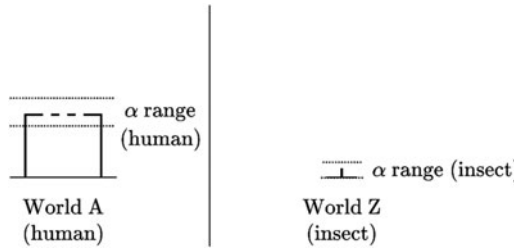


Figure 8. Species-relative critical ranges.

World A (human) comes out as better than *World Z (insect)* if *World A (human)* contains the greater total critical value for every possible combination of α points within every species-relative critical range. Vice versa, *World Z (insect)* comes out as better than *World A (human)* if *World Z (insect)* contains the greater total critical value for every possible combination of α points within every species-relative critical range. Otherwise, we say that *World A (human)* and *World Z (insect)* are incomparable under vagueness, that is, there is simply no fact of the matter as to whether *World A (human)* is better than *World Z (insect)*, whether *World Z (insect)* is better than *World A (human)*, or whether *World A (human)* and *World Z (insect)* are equally good. Since *World A (human)* and *World Z (insect) do* contain the greater values for different possible α points within the critical range, the species-relative critical ranges successfully mitigate the force of the Flipped Repugnant Conclusion, leaving us with only indeterminacy in the comparison.

On this view, we can also rewrite the Animal Repugnant Conclusion as follows:

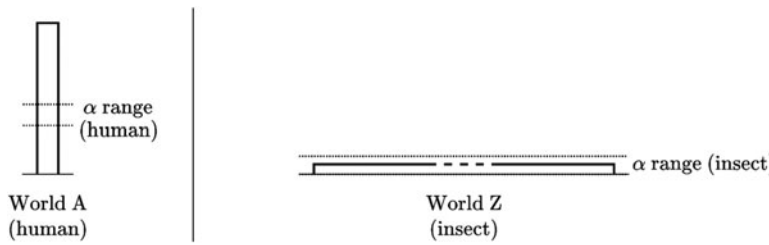


Figure 9. The Repugnant Conclusion for species-relative critical ranges.

Again, given a judicious placement of the critical ranges, the Animal Repugnant Conclusion loses its sting, since *World A (human)* will contain the greater value for some possible α points within the insect-specific critical range, while *World Z (insect)* will contain the greater value for other possible α points within the insect-specific critical range. Hence, we are only forced to admit that the two populations are indeterminately comparable under vagueness.

So, can critical ranges ultimately salvage the species-relative view? First of all, we should make the same criticisms that applied in the case of the universal critical range. (1) The vague critical range is being used to *ad hoc* block those betterness orderings we find intuitively unacceptable, rather than to merely reflect the fact that there is some degree of vagueness regarding the point at which individual lives begin to contain

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

positive impersonal value. (2) One consequence of employing a multiplicity of species-relative critical ranges in this way will be widespread indeterminacy. For an extremely large range of population comparisons, comparisons in which the declaration of indeterminacy seems inappropriate, we will now be forced to say that there is no fact of the matter as to whether *A* is better than *B*, or *B* is better than *A*. For instance, take the Flipped Repugnant Conclusion. Is it really appropriate to say that the two populations featuring in the Flipped Repugnant Conclusion are incomparable under vagueness? Many of us will have the opposite intuition, namely, that *World A (human)* is determinately better than *World Z (insect)*, and that our axiology must be capable of telling us as much. For those of us who share this intuition, the critical range strategy is unlikely to count as an adequate response to the Flipped Repugnant Conclusion or the Animal Repugnant Conclusion.

I am not claiming that these bullets of (1) and (2) cannot be bitten, but I am claiming that they count as serious marks against the species-relative critical range view. Even if we were willing to accept the sort of widespread indeterminacy entailed by a multiplicity of species-relative critical ranges, a further problem would remain: we would find ourselves in want of some principle for knowing how to set the critical ranges themselves. We previously rejected common heuristics for finding where the species-relative critical *levels* were likely to exist, for instance, Blackorby, Bossert and Donaldson's (1997: 217) comment that one might set the critical level for non-human species at some fraction of the total welfare typically experienced by members of that species in the wild.[13] In the same vein we do not have some heuristic for measuring where the species-relative critical *ranges* exist, that is, where the ranges begin and where they end. If we rely on arbitrary estimation then the betterness order, too, will rely on arbitrary estimation.

Perhaps someone might reply that this result is not too troubling. All good ethics will eventually defer to arbitrary estimation, and one philosopher's arbitrary estimation is another philosopher's moral intuition. But if this were the situation in which we found ourselves, we would have no guarantee that species-relative critical ranges constituted a genuine response to the Animal Repugnant Conclusion or the Flipped Repugnant Conclusion after all. For instance, it might be the case that the insect-relative critical range is best characterised as occupying a range below the maximum level of insect welfare, a view on which individual insect lives were still capable of adding determinate impersonal value to the goodness of an outcome, and on which any given human utopia might still be outranked by a population of happy (but nevertheless still insectile) insect lives.

Unless the critical range utilitarian knows that the species-relative critical range encompasses all possible positive welfare levels for the species-*x* featuring in *World Z* of the Animal Repugnant Conclusion, they cannot know whether their strategy genuinely blocks the Animal Repugnant Conclusion involving a *World Z* composed of species-*x*. For this reason, even if we were willing to bite the bullets entailed previously by species-relative critical ranges, these ranges seem unlikely to guarantee avoidance of axiological puzzles like the Animal Repugnant Conclusion.

[13]Further issues would emerge for critical-level utilitarians here, if it turned out that most wild animals lived lives that were not worth living. If this proved to be the case for members of a particular species, say, then neither a positive critical level nor a negative critical level would seem appropriate. (A positive critical level would represent an impossible standard; a negative critical level would constitute sadism.) This might serve as further evidence against the possibility of setting morally plausible critical levels.

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

7. Individual-relative critical levels

We have considered the proposal on which different species are governed by different species-relative critical levels (ranges). Recall the rationale for this view: both that (i) the universal critical level was unsustainable, and that (ii) different species tend to exhibit different psychological capacities for welfare, and that it is hence appropriate to measure their welfare against different species-relative critical levels.

This whole time, though, we have been ignoring a serious problem. This is the issue of whether 'species' is even a natural kind upon which to ground the existence of different critical levels in the first place. After all, some breeds of dog are likely to live much longer than others, and are therefore likely to have greater welfare capacities than others. Why didn't we therefore frame the discussion in terms of 'breed-relative' critical levels, or even in terms of the further variations in welfare capacity that might appear within a single breed? Then there is the question of whether we are indexing the critical level to a dog breed at time t (since dog breeds may well exhibit a great deal of variation in welfare capacities based on their historical period) or of whether we are indexing the critical level to a dog breed in an environmental context e (since dog breeds may well exhibit a great deal of variation in welfare capacities based on their environmental context).

Worse than this, though, is the fact that even within a species and within a breed there exists a great deal of variation in welfare capacities from individual to individual. Consider the case of two greyhounds, Alfredo and Linguine. Due to a chronic and irreversible illness, Alfredo is born with diminished welfare capacities and a comparatively limited lifespan. Applying the same reasoning that led to species-relative critical levels, namely, that critical levels must vary to respect the fact that different creatures have different capacities for well-being, it would seem appropriate to furnish both Alfredo and Linguine with individual-relative critical levels. It would be inappropriate to measure Alfredo against a (high) Linguine-relative critical level, or to measure Linguine against a (low) Alfredo-relative critical level, just as it was inappropriate to measure human welfare against a deer-relative critical level or vice versa.

But Blackorby, Bossert and Donaldson (1997: 216) rightly dismiss the idea of individual-relative critical levels. Such theories would tell us that creating Alfredo two units above his (low) individual-relative critical level is better for the world than creating Linguine one welfare unit above her (high) individual-relative critical level, even though, for most of us, this result is extremely counterintuitive. Our intuition is instead that it would be impersonally better to create a dog that was not shackled by chronic illness from birth to death rather than a dog that was, even if the chronically ill dog scored slightly comparatively better with respect to its (low) individual-relative critical level than the healthy dog scored with respect to its (high) individual-relative critical level.[14]

[14]The issues raised here closely mirror McMahan (1996). Perhaps the suggestion of individual-relative critical levels most closely parallels what McMahan (1996: 18–19) calls the *Native Potential Account of Fortune*: that whether or not 'a being is well or badly off depends on how its level of well-being compares to the range of well-being made possible by the highest cognitive and emotional capacities that it has actually achieved or that it natively had the potential to achieve'. The difference, if I have pinpointed it correctly, is that McMahan's discussion is still framed in terms of the personal value contained in a life; by contrast, critical-level utilitarians are concerned purely with defining the *impersonal* value contained within a life. This difference is why axiological puzzles (like that of Alfredo versus Linguine) quickly emerge for individual-relative critical levels, and render the view implausible.

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

If (i) there is no principled reason to stop the move from a universal critical level to species-relative critical levels, from species-relative critical levels to breed-relative critical levels, and from breed-relative critical levels to individual-relative critical levels, and (ii) it is true that individual-relative critical levels are simply untenable, then (iii) we might see these two facts as the basis of a *reductio* against any variety of critical-level theory which required group-relative critical levels. I say 'any variety of critical-level theory' perhaps hastily, but not without reason. For instance, critical-level theories which might assign merely no value (rather than negative value) to lives below the critical level seem just as subject to the problem of Alfredo and Linguine as the sorts of critical-level theory I have been discussing previously, since, if we really *were* compelled to assign an individual-relative critical level to both Alfredo and Linguine, both of them would still exist above (rather than below) their individual-relative critical levels. Hence, we could not resolve the dilemma of Alfredo and Linguine by merely changing the way we operated upon lives which existed below the critical level. We find, then, that these considerations from animal welfare threaten even more varieties of critical-level theory than were presented as the focus of this article.

8. Conclusion

This article has outlined a new dilemma for critical-level utilitarians. Either critical level-utilitarians adopt a universal critical level, or they adopt a plurality of species-relative critical levels. The universal critical level is an implausible axiological principle, entailing the problems both of axiological elitism and of disproportionate disvalue. This leaves critical-level utilitarians with the species-relative interpretation of their theory, on which the critical level exists at different points for different members of different populations. But on this view, welfare often does not contain the comparative weight it ought to. We are subject to new axiological puzzles, including the Animal Repugnant Conclusion and the Flipped Repugnant Conclusion. Adopting critical ranges rather than critical levels does little to ameliorate the situation, either on the universal or on the species-relative view, and further approaches to the threshold, such as the individual-relative interpretation, are unsustainable. We have good reason, then, for thinking that critical-level utilitarianism cannot be a morally plausible welfarist axiology.

This article has not claimed to be an exhaustive discussion of the issues that will emerge when incorporating non-human animal welfare into the project of population axiology. But hopefully, it has gone some way to showing that the task is not a simple one; at the very least, we see that critical-level utilitarianism has been problematically underspecified all along.[15]

References

Arrhenius, Gustaf. 2000a. An Impossibility Theorem for Welfarist Axiologies, *Economics and Philosophy*, **16**(2): 247–66.

Arrhenius, Gustaf. 2000b. *Future Generations: A Challenge for Moral Theory* (Uppsala: Uppsala University Printers).

[15]I am especially indebted to Katie Steele. Thanks, too, to Jeff McMahan for many comments on an earlier draft, and to an anonymous reviewer for *Utilitas*. For comments and feedback on the project that led to this article I am also thankful to Renée Bolinger, Mark Budolfson, Kida Lin, Timothy L. Williamson and Matthew Wiseman. My research is supported by an Australian Government Research Training Program Scholarship.

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press

Blackorby, C., W. Bossert and D. Donaldson. 1996. Quasi-Orderings and Population Ethics, *Social Choice and Welfare*, **13**: 129–51.

Blackorby, C., W. Bossert and D. Donaldson. 1997. Critical-Level Utilitarianism and the Population Ethics Dilemma, *Economics and Philosophy*, **13**(2): 197–230.

Blackorby, C., W. Bossert and D. Donaldson. 2005. *Population Issues in Social Choice Theory, Welfare Economics, and Ethics* (New York: Cambridge University Press).

Blackorby, C. and D. Donaldson. 1984. Social Criteria for Evaluating Population Change, *Journal of Public Economics*, **25**: 13–33.

Broome, John. 2004. *Weighing Lives* (Oxford: Oxford University Press).

Greaves, Hilary. 2017. Population Axiology, *Philosophy Compass*, **12**(11): 1–15.

Griffin, James. 1986. *Well-Being: Its Meaning, Measurement and Moral Importance* (Oxford: Oxford University Press).

Gustafsson, Johan E. 2020. Population Axiology and the Possibility of a Fourth Category of Absolute Value, *Economics and Philosophy*, **26**(1): 81–110.

McMahan, Jeff. 1996. Cognitive Disabilities, Misfortune, and Justice, *Philosophy and Public Affairs*, **25**(1): 3–35.

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

Parfit, Derek. 2016. Can We Avoid the Repugnant Conclusion?, *Theoria*, **82**(2): 110–27.

Sebo, Jeff. n.d. Animals and Climate Change, *Philosophy and Climate Change*, forthcoming.

Tännsjö, Torbjörn. 2002. Why We Ought to Accept the Repugnant Conclusion, *Utilitas*, **14**(3): 339–59.

Cite this article: Williamson P (2021). A New Argument Against Critical-Level Utilitarianism. *Utilitas* **33**, 399–416. https://doi.org/10.1017/S0953820821000133

https://doi.org/10.1017/S0953820821000133 Published online by Cambridge University Press