Discussion Paper Series

Institute-Wide Joint Research Project
Methodology of Social Sciences: Measuring Phenomena and Values

Multi-species population ethics with critical levels

Susumu Cato
University of Tokyo

Shu Ishida
Osaka University

June 10, 2022

E-22-004

Institute of Social Science, University of Tokyo
[https://web.iss.u-tokyo.ac.jp/methodology/en/dp/dp/]

Multi-species population ethics with critical levels*

Susumu Cato
Institute of Social Science
University of Tokyo
7-3-1, Hongo, Bunkyo
Tokyo 113-0033
Japan
susumu.cato@gmail.com

Shu Ishida
Research Center on Ethical, Legal, and Social Issues
Osaka University
2-8, Yamadaoka, Suita
Osaka 565-0871
Japan
isd.shu@gmail.com

Abstract. This paper explores the concept of species-relative critical levels. It is a crucial issue in multi-species population ethics. First, the formal conditions are provided under which there are species-relative critical levels (e.g., the critical level for human beings is different from that for non-human beings). In particular, we find it a salient moral question of animal ethics whether the existence of a human being is morally better than that of a non-human animal when their utility levels are the same. Subsequently, we illustrate two general classes of multi-species critical-level utilitarian orderings. One class employs species-relative critical levels; this entails either the animal repugnant conclusion or the animal sadistic conclusion. The other employs species-relative critical levels plus species-lexical ordering. Although it can avoid both the animal repugnant conclusion and the animal sadistic conclusions, it is speciesist.

Keywords. lexical view, population ethics, repugnant conclusion, sadistic conclusion, species-relative critical levels, utilitarianism

* We deeply thank Patrick Williamson for his helpful comments to an earlier version of this paper. This paper was financially supported by KAKENHI (20H01446, 22K01387, 22H05086) and the Mitsubishi Foundation (ID201920011).

1 Introduction

Since Parfit (1976, 1982, 1984) established his famous argument on the repugnant conclusion, population ethics has seen a considerable number of works concerning the idea of critical levels. Critical levels address the moral question of when the existence of an individual is morally valuable (as opposed to being valuable for the individual in question). As a typical interpretation, a critical level is a particular utility level at which the existence of an individual is morally neutral. Thus, the existence of an individual whose utility level is below the critical level is of negative moral value, even though the individual's utility level is positive and the life is worth living (Blackorby, Bossert, and Donaldson, 2005).

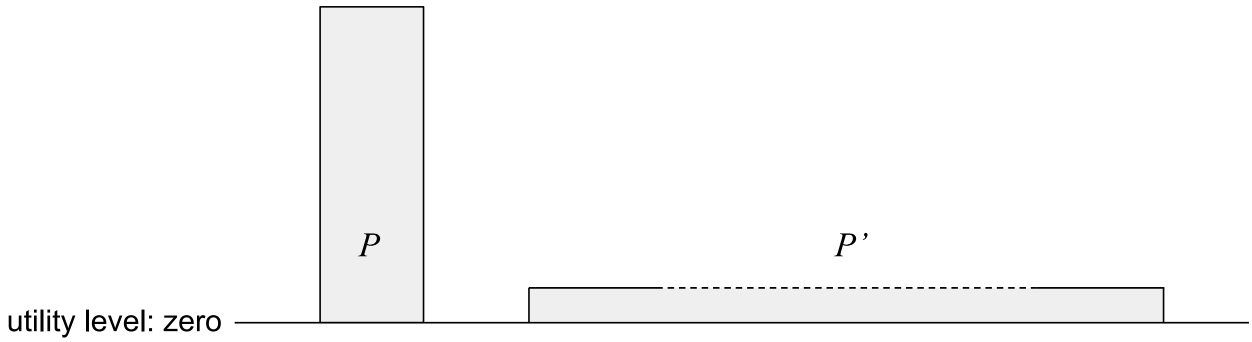
This paper develops a formal approach to assess the logical possibility and moral implications of *species-relative critical levels*. While many studies have been conducted on critical levels regarding human-only populations, our central focus in this article is on critical levels for non-human beings. That is not a mere extension. The very idea of critical levels is a matter of moral value, which means we should face a salient question of animal ethics concerning critical levels for non-human beings. Where should it be set? Is it higher or lower than the critical level for human beings? Based on this assessment, we identify and examine several classes of multi-species population principles that employ the idea of critical levels. Some such population principles can avoid the problems recently raised by Williamson (2021).

The rest of this paper is organized as follows. Section 2 reviews the basic ideas and discussions behind our study in this paper. Section 3 introduces our formal setting. Section 4 addresses several animal-ethical questions concerning species-relative critical levels. It also illustrates the implications of the existence of species-relative critical levels to multi-species population ethics. Section 5 identifies and examines a general class of population principles that involve species-relative critical levels. Section 6 considers a related class of population principles with species-lexical features. As is clear from our discussion below, one significant puzzle for this class of population principles is to avoid morally wrongful speciesism. Section 7 concludes this study. Appendix A includes technical materials, while Appendix B offers arguments on another paradoxical result.

2 Population ethics with human-beings

This section sets the stage of our study. We provide a short explanation of three vital issues in population ethics; that is, the repugnant conclusion, critical levels, and the sadistic conclusion.

Figure 1: Repugnant Conclusion

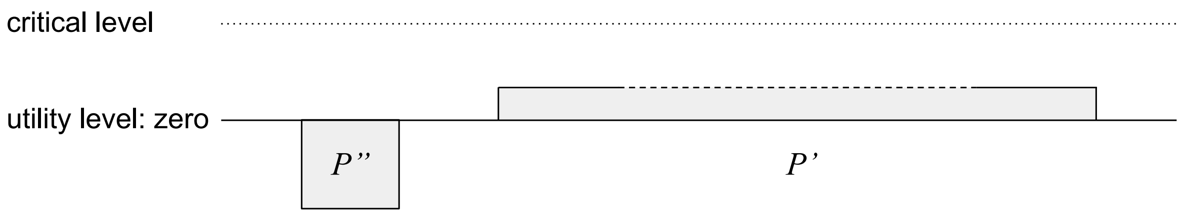


The **repugnant conclusion** is one of the most well-discussed topics in population ethics.[1] A population principle entails the repugnant conclusion if, for any alternative in which each person's utility level is positive, there is a morally better alternative with many persons whose utility level is barely above the neutral level (Parfit 1984: sec. 131). To see the point more vividly, consider total utilitarianism in an ordinary sense. To be precise, total utilitarianism makes moral judgments over populations of different sizes based on the summations of individual utility levels. Let $\mathcal{P}$ represent a large population that consists of happy persons. The total sum of utility in $\mathcal{P}$ is quite large, absolutely speaking. However, for any given $\mathcal{P}$, we can conceive of a much larger size of population $\mathcal{P}'$ consisting of miserable (i.e., barely worth-living) persons, such that the total sum of utility in $\mathcal{P}'$ is larger than that in $\mathcal{P}$. Figure 1 illustrates the two populations. According to total utilitarianism, then, $\mathcal{P}'$ is morally better than $\mathcal{P}$, which seems "repugnant." This is the gist of the repugnant conclusion.

In response to the repugnant conclusion, one suggestion is to introduce a constant, "critical" utility level. As a typical interpretation, the moral value of a life with the utility level below the **critical level** is negative, even if the utility level itself is positive (Blackorby, Bossert, and Donaldson, 2005). That is, such a life makes the world *impersonally* worse, even though that life is *personally* worth living for the one who lives it. Similarly, if a person's utility level equals the critical level, the life of such a person is morally inert. The population with that person is indifferent, impersonally speaking, to the population without that person. The most well-known population principle that involves a critical level is critical-level utilitarianism, i.e., utilitarianism

[1] Ng (1989), Blackorby, Bossert, Donaldson, and Fleurbaey (1998), and Carlson (1998) show a general difficulty of avoiding the repugnant conclusion. See also Spears and Budolfson (2021) and Stewart (forthcoming) for recent results on variations of the repugnant conclusion.

Figure 2: Sadistic Conclusion



involving a constant critical level. According to this principle, moral judgments for populations are made based on the (sum of) differences between individual utility levels and the constant critical level. Let us assume that the constant critical level is set at a substantially high utility level. In that case, we can reasonably expect that the utility level of individuals in $\mathcal{P}$ is above the critical level, while that in $\mathcal{P}'$ is below it. Hence, the moral value of $\mathcal{P}$ is positive and that of $\mathcal{P}'$ is negative, which means that $\mathcal{P}$ is morally better than $\mathcal{P}'$. Thus, critical-level utilitarianism can avoid the repugnant conclusion.

However, critical-level utilitarianism faces another problem, the **sadistic conclusion** (Arrhenius, 2000, forthcoming). We understand it in the following sense; a population principle entails the sadistic conclusion if, for any alternative in which each person's utility level is barely positive, there is a morally better alternative with fewer persons whose utility level is negative. According to the abovementioned critical-level utilitarianism, the moral value of $\mathcal{P}'$ is negative, even though the utility level of the individuals in $\mathcal{P}'$ is positive. However, for any given $\mathcal{P}'$, we can conceive of a smaller-sized population $\mathcal{P}''$ with heartbreakingly painful lives, such that $\mathcal{P}''$ is morally better—or, more naturally, morally less bad—than $\mathcal{P}'$. That is so even if the lives of those involved in $\mathcal{P}''$ are heartbreakingly painful while the lives of those involved in $\mathcal{P}'$ are worth living. Figure 2 illustrates the two populations. This upshot seems quite "sadistic," which is the gist of the sadistic conclusion.[2]

[2]This problem is usually called the *very* sadistic conclusion or the *strong* sadistic conclusion (Arrhenius, 2000). The original version of the sadistic conclusion states that given a population, adding a population $\mathcal{P}'$ where everyone obtains a positive utility level is worse than adding another $\mathcal{P}''$ where everyone has a negative utility level. See Arrhenius (2000) for a detailed explanation. The distinction sometimes matters, although their axiological implications are almost the same; see Bossert, Cato, and Kamaga (2021). This paper ignores that distinction because its focus is on population principles that satisfy "existence independence" under which the two conditions

3 Formal setting

Roughly speaking, our study compares the moral value of human and non-human populations in terms of utility. Before delving into the discussion, let us clarify our fundamental assumption that human and non-human utility levels are numerically represented and cardinally comparable. We do not deny the possibility that human and non-human utility may consist of different items—typically, the utility of human beings may depend on something more "sophisticated" than that of non-human beings. Even in that case, however, the utility *levels*, not the *components* thereof, could be comparable between human and non-human beings. Our study below need not assume that the same item contributes to both human and non-human utility.

In this study, we assume that there are only two species, α and β. All results can be extended to a general case with a finite number of species. As a typical interpretation, α denotes human beings, and β denotes a non-human species such as insects; we use this interpretation later regarding the animal repugnant conclusion and the animal sadistic conclusion. Let $u_\alpha = (u_{1\alpha}, \ldots, u_{n\alpha})$ be a utility distribution for species α with n individuals existing; $u_{i\alpha}$ denotes the utility level of individual i of species α. Similarly, we consider a utility distribution for species β with m individuals, which is given by $u_\beta = (u_{1\beta}, \ldots, u_{m\beta})$; $u_{j\beta}$ denotes the utility level of individual j of species β. In this study, a pair of distributions of the two species, (u_α, u_β), is examined. These numerical values represent prudential goodness for individuals.[3] The life of an individual is deemed to be worth living when his/her utility level is above zero; the zero level is said to be neutrality.[4] We assume that the population size of each species can vary; that is, n and m can change. To make this framework properly general, we allow one of the two species to be empty. Let $u_\emptyset$ be the null situation where there is no individual. For example, $(u_\alpha, u_\emptyset)$ represents a distribution where there are only individuals of species α.

Let Ω_α (resp. Ω_β) be the sets of possible distributions of species α (resp. species β).[5] Each of these sets includes the empty set, as mentioned above. The set of the entire populations is denoted

in question become equivalent. See Blackorby, Bossert, and Donaldson (2005) and Section 5 of this paper for the precise definition of existence independence.

[3]See Broome (1993, 2004) and Vallentyne (1993, 2009) for related arguments on prudential goodness.

[4]There are several versions of the neutrality level. Here, following Williamson (2021, p. 400), we employ the "neutral level for continuing to live," which is proposed by Broome (2004, pp. 234–245). This is basically a level of utility such that, for each individual, extending their lives with the level is indifferent to their deaths.

[5]Formally, the set of possible distributions of species α is given as follows: $\Omega_\alpha = \{u_\emptyset\} \cup \left(\bigcup_{n=1}^{\infty} \mathbb{R}^n\right)$, where $\mathbb{R}$ is the set of real numbers. This formulation of Ω_α allows species α to be empty. Similarly, we define the set of possible distributions of species β as follows: $\Omega_\beta = \{u_\emptyset\} \cup \left(\bigcup_{n=1}^{\infty} \mathbb{R}^n\right)$.

by Ω. Assuming that at least one of the two species exists, the set of possible distributions of the entire population (consisting of α and β) is defined as follows: $\Omega = \{(u_\alpha, u_\beta) \in \Omega_\alpha \times \Omega_\beta \mid u_\alpha \neq u_\emptyset \text{ or } u_\beta \neq u_\emptyset\}$. We often write (u_α, u_β) and (v_α, v_β) simply as u and v, respectively. The world only with human-beings is obtained by assuming $u_\beta = u_\emptyset$.

An inter-species population principle is an *ordering* R on Ω, which is an at-least-as-good-as relations over utility distributions. An ordering is a binary relation (or preference) that satisfies completeness and transitivity. Notably, R is said to be complete if each pair of distinct alternatives is ranked; and transitive if u is at least as good as w as long as u is at least as good as v and v is at least as good as w.[6] A typical expression is "uRv", which means "u is at least as good as v." The corresponding strict and indifference relations are denoted by P and I. Then, "uPv" means "u is strictly better than v," while "uIv" means "u is indifferent to v." Population principles considered in this paper are more general than the one which has been examined by the existing literature. Indeed, a distribution $u = (u_\alpha, u_\emptyset)$ is regarded as a single-species distribution, which is equivalent to the standard variable-population setting of Blackorby and Donaldson (1984) and Broome (2004).

Now, we define the concept of critical levels. Take a distribution $u = (u_\alpha, u_\beta)$ and assume that one individual of species α with the utility level c_α is added to this distribution. A critical level for species α under (u_α, u_β) is the utility level of this individual if the original (u_α, u_β) is indifferent to the distribution $((u_\alpha, c_\alpha), u_\beta)$ obtained by this addition. A critical level for species β is similarly defined. Notably, the existence of such a utility level is not always guaranteed. Moreover, critical levels can be dependent on distributions (u_α, u_β) in general. For instance, variable-value principles established by Hurka (1983) and Ng (1986, 1989) have critical levels dependent on the number of individuals under the standard settings only with human beings.

As the key feature of critical-level principles, Blackorby, Bossert, and Donaldson (1997, 2005) require critical levels to be always the same independently of existing populations. In that case, critical levels are *constant* under critical-level principles. To be precise, we say that there is a constant critical level for species α if and only if there exists a utility level c_α such that, for all $(u_\alpha, u_\beta) \in \Omega$, (u_α, u_β) is morally indifferent to $((u_\alpha, c_\alpha), u_\beta)$, where (u_α, c_α) is a distribution of species α including a new individual with the utility level c_α. A constant critical level for species β can be similarly defined. Species-relative critical levels exist if and only if there are constant critical levels for each species. Moreover, we can say that a universal critical level exists if and

[6] Reflexivity requires that each alternative is as good as itself. This follows from completeness.

only if species-relative critical levels exist and their values are the same between each species.

4 Species-relative critical levels and repugnant conclusions

We examine a general class of multi-species population principles in Section 5. Before that inquiry, let us see several issues about multi-species population ethics in this section.

4.1 The animal repugnant conclusion

Blackorby, Bossert, and Donaldson (2005) provide a basic formulation of multi-species critical-level utilitarianism. It is essentially the same as Williamson's (2021) species-relative critical-level utilitarianism. We use the former terminology in this paper. An ordering R on Ω is **multi-species critical-level utilitarian** if, for all population sizes n, m, p, q, and for all distributions $(u_\alpha, u_\beta) = ((u_{1\alpha}, \ldots, u_{n\alpha}), (u_{1\beta}, \ldots, u_{m\beta}))$ and $(v_\alpha, v_\beta) = ((v_{1\alpha}, \ldots, v_{p\alpha}), (v_{1\beta}, \ldots, v_{q\beta}))$, $(u_\alpha, u_\beta) R (v_\alpha, v_\beta)$ if and only if

$$\sum_{i=1}^{n}(u_{i\alpha} - c_\alpha) + \sum_{i=1}^{m}(u_{i\beta} - c_\beta) \geq \sum_{i=1}^{p}(v_{i\alpha} - c_\alpha) + \sum_{i=1}^{q}(v_{i\beta} - c_\beta). \tag{1}$$

Notably, c_α is the constant critical level for species α, while c_β is that for species β. It can be the case that one of the two species does not exist; some of the abovementioned operations may be a summation that includes no term. We assume that such an empty sum is zero by following the standard practice in mathematics. Moreover, if we assume that a universal constant critical level exists, the abovementioned population principle is essentially the same as critical-level utilitarianism in an ordinary sense, with non-human species β just added; see Broome (2004: 43). Although that may raise unique ethical questions (Williamson 2021: sec. 5), hereafter we exclude the case where c_α is equal to c_β. That is, c_α is assumed not to equal to c_β to focus on critical levels that are purely species-relative. Also, we assume that critical levels are non-negative for both species; it is easy to extend our analysis to the case with negative critical levels.

There are two possibilities; either (i) $c_\alpha > c_\beta$ or (ii) $c_\alpha < c_\beta$. In their argument on multi-species critical levels, Blackorby, Bossert, and Donaldson (1997, 2005) assume that the critical level for human beings is higher than that for non-human beings (i.e., $c_\alpha > c_\beta$). This assumption is taken by Williamson (2021) as well. They find this assumption justifiable by appealing to the difference in their capacity for enjoying one's "integrated lives"; while human beings can care for one's past and future, non-human beings (such as mice, rats, and fish) cannot. That makes it "natural," they take, that the critical level for human beings is higher than that for non-human beings.

However, multi-species critical-level utilitarianism can face a problem. The problem, dubbed as "the animal repugnant conclusion" by Williamson (2021), is formally stated as follows:

The animal repugnant conclusion. For all positive integers n, for all positive numbers ξ, and for all positive number ε smaller than ξ, there exists a sufficiently large number m such that

$$(u_\emptyset, \varepsilon \mathbf{1}_m) P (\xi \mathbf{1}_n, u_\emptyset),$$

where $(u_\emptyset, \varepsilon \mathbf{1}_m)$ is a distribution of species β and $(\xi \mathbf{1}_n, u_\emptyset)$ is a distribution of species α.

Williamson (2021) showed that the animal repugnant conclusion is entailed if the critical level of non-human beings is zero. Does this conclusion hold when it is not zero? Interestingly, it is not the case that multi-species critical-level utilitarianism necessarily entails the animal repugnant conclusion. Our proposition states the full characterization with regard to this problem; the detailed proof is in Appendix A.

Proposition 1. Assume that both c_α and c_β are non-negative.
(i) If c_α is higher than c_β, then a multi-species critical-level utilitarian ordering entails the animal repugnant conclusion if and only if $c_\beta = 0$ (Williamson, 2021);
(ii) If c_α is lower than c_β, then a multi-species critical-level utilitarian ordering avoids the animal repugnant conclusion.

We note that Williamson (2021) proved the "if" part of the first component of this proposition. Our proposition shows that the converse is also correct. That is, the animal repugnant conclusion holds only when there is a species whose critical level is zero. This proposition suggests that the result provided by Williamson (2021) concerned a very extreme case. Perhaps this is why he offered another argument against critical-level utilitarianism as a backup; what he calls the *flipped repugnant conclusion* holds for a more general setting of critical levels.[7]

Then, one can ask: Is the animal repugnant conclusion ignorable or insignificant as a paradox that is false almost everywhere in the space of critical levels? We think that the animal repugnant conclusion still has its relevance in a way. The reason is that a similar problem can occur as long as the critical level for human beings is significantly higher than that for non-human beings. For

[7]We provide a detailed analysis of the flipped repugnant conclusion in Appendix B, although it is not the main focus of this study.

non-human beings, take a utility level just above their critical level. Then, for any large human population with tremendously high utility levels, there is a much large non-human population with the utility level above the critical level. This argument can apply as long as the critical level for human beings is significantly higher than that of non-human beings. More formally, there exists a positive number ν such that for all positive integers n, for all number ξ larger than ν, and for all number ε in the middle of ν and ξ (i.e., $\nu < \varepsilon < \xi$), there exists a sufficiently large integer m such that $(u_\emptyset, \varepsilon\mathbf{1}_m)$ is strictly better to $(\xi\mathbf{1}_n, u_\emptyset)$, where $(u_\emptyset, \varepsilon\mathbf{1}_m)$ is a distribution of species β and $(\xi\mathbf{1}_n, u_\emptyset)$ is a distribution of species α. We may call this the strong animal repugnant conclusion. Indeed, this extension of the animal repugnant conclusion more generically holds, and it shows the relevance of the animal repugnant conclusion.

As mentioned above, the critical level for human beings is supposed to be significantly higher than that for non-human beings in the existing works. However, there is a *prima facie* reason to have it that $c_\alpha < c_\beta$ if a multi-species critical-level population principle should avoid the animal repugnant conclusion and its strong version. Is there anything that would mandate us to see the critical level for human beings as higher than that for non-human beings? First, one might appeal to biological facts. As we have seen before, Blackorby, Bossert, and Donaldson (2005) refer to the difference in the cognitive capacity between human and non-human beings. In a similar vein, Williamson (2021) appeals to the difference in natural longevity, which typically results in larger sums of human lifetime well-being compared to those of non-human lifetime well-being. Call them *naturalistic* views. However, naturalistic views face a problem even assuming the biological difference. In general, it is implausible that a mere biological fact alone entails any moral claim. The claim that the critical level for human beings is higher than that for non-human beings is a *moral* one, given that critical levels are defined in terms of moral value (i.e., of the moral value of the existence of an additional individual with particular utility level). Such a moral claim might have a biological fact as an empirical premise, but a normative one is also required.

The required normative premise is perhaps a kind of perfectionism, according to which it is morally valuable to exercise one's characteristic and/or essential capacities (Dorsey 2010, 59).[8] Then one could argue as follows: given that human beings are capable of higher utility levels thanks to their sophisticated cognitive capacity or long lifespan, an individual human being's utility level must be substantially high for it to be morally valuable. That is, there is a "mediocre"

[8] Some might find it a misnomer to call this view perfectionist. Precisely speaking, the view we are considering is a "capacity-relative" one, which we think could be seen (somewhat controversially) as a form of perfectionism. Setting this terminological issue aside, we use "perfectionism" as a shorthand for this view.

utility level, μ, such that (a) the existence of a non-human individual with utility level μ is morally good but (b) the existence of a human individual with utility level μ is not. Call it the *perfectionist* view.

The perfectionist view is hard to support at its face value. Let α denote a species consisting of actual human beings and let $\hat{\alpha}$ denote another species (say, the Eloi). Eloi individuals are counterfactual human beings whose cognitive capacity is substantially lower than that of actual human beings. The perfectionist view holds that c_α is substantially higher than $c_{\hat{\alpha}}$. That is, for a real number μ such that $c_{\hat{\alpha}} < \mu < c_\alpha$, the existence of an individual in $\hat{\alpha}$ with utility level μ is morally *better* than that of an individual human being with utility level μ. We find this implication quite implausible. If anything ever is valuable from the viewpoint of perfection-focused ethics at all, the existence of a lower-capacity individual seems to be morally *worse* than that of a higher-capacity individual, other things being equal.

Given these observations, we cannot simply assume that $c_\alpha > c_\beta$; not only does the alleged case for this assumption fail, but it has even an implausible implication. We must take a serious consideration of the possibility that $c_\alpha < c_\beta$.

4.2 The animal sadistic conclusion

Although the animal repugnant conclusion is absolutely avoided under the assumption that the critical level of human-beings is lower than that of non-human beings, critical-level utilitarianism with this assumption may not be a definite resolution. As alluded in Section 2, the repugnant conclusion is not the only problem in population ethics. Indeed, a population principle immune to the repugnant conclusion can face another problem, such as the sadistic conclusion. That is what also happens in the context of multi-species population ethics.

As argued in the previous subsection, the key is how the existence of human beings compares, morally speaking, to that of non-human beings (such as insects), when the utility levels of the two groups are the same. If the critical level for human beings is lower than that for non-human beings, it allows the possibility that the existence of human beings is morally better than that of non-human beings. Then one can ask: is it the case that a society of human beings who are not worth living is better than a society of worth living animals?

Let us take a society within which only one human being with a negative utility level exists. Hypothetically, one can compare this society with another society which consists of n non-human beings with a positive utility level. If there is only one animal (i.e., $m = 1$), it may be morally acceptable that the human society is better than the animal society. However, if the number of non-human beings is substantially large, it sounds "sadistic" to say that the human society (with a single non-worth-living individual) is morally better than the animal society (with many worth-living individuals). If this priority to the society with a single human being is persistent for all m, we say that the animal sadistic conclusion holds. Formally, this conclusion can be expressed as follows.

The animal sadistic conclusion. There exist a negative number ε, a positive number ξ, and a positive integer m^* such that for all integers m larger than m^*,

$$(\varepsilon \mathbf{1}_1, u_\emptyset) P (u_\emptyset, \xi \mathbf{1}_m),$$

where $(\varepsilon \mathbf{1}_1, u_\emptyset)$ is a distribution of species α and $(u_\emptyset, \xi \mathbf{1}_m)$ is a distribution of species β.

In a footnote, Williamson (2021) mentioned a conclusion related to this. He wrote: "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." (408, Footnote 9). However, this conclusion is proposed as an argument for the case of universal critical levels. Thus, Williamson (2021) states: "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" (408, Footnote 9). Our point is that this conclusion is indeed significantly important even for the case with purely species-relative critical levels, and that trying to avoid both conclusions leads us to a new perspective, as shown in Section 6.

The animal sadistic conclusion has a similar spirit to the original, human-only version of the sadistic conclusion. Both versions are objections to theories under which a population consisting of worth-living lives are morally worse than one consisting of lives that are not worth living. Furthermore, the animal sadistic conclusion holds only when sadistic judgments are persistent; that is, such a sadistic moral judgment is available *no matter how many non-human beings* there are in the society. Because of this nature, we have a *prima facie* reason to avoid the sadistic conclusion even when it is about a comparison between different species.

How do critical levels matter for the animal sadistic conclusion? Let us assume that the critical level for human beings is lower than that for insects. Assuming that the critical level for human beings is non-negative, the critical level for insects is positive. Then, one can find a positive utility level, ξ, of insects, which is lower than their critical level. Notably, if adding an insect with utility level ξ yields a new population, without affecting the original individuals' welfare, then the new population is worse than the original one. Indeed, even though ξ is a positive utility level, the more insects with the utility level ξ are added, the worse the aggregated utility level is. Then, there must be a large number m^* such that a society with one human being who is not worth living is better than a society with m insects worth living, for all m that is larger than m^*. Thus, the animal sadistic conclusion holds. However, the animal sadistic conclusion must be avoided if the critical level of insects is zero, as presumed by Blackorby, Bossert, and Donaldson (2005) and Williamson (2021). That is because for any number of insects that are worth living, the sum of the differences between its utility and the critical level of insects must be positive, while the corresponding value for the single human being with a negative utility is negative.

As a summary, the following proposition formally states the condition under which a species-relative critical-level utilitarian ordering faces the animal sadistic conclusion; the proof is relegated to Appendix A.

Proposition 2. Assume that both c_α and c_β are non-negative.
(i) If c_α is higher than c_β, then a multi-species critical-level utilitarian ordering avoids the animal sadistic conclusion if and only if $c_\beta = 0$;
(ii) if c_α is lower than c_β, then a multi-species critical-level utilitarian ordering entails the animal sadistic conclusion.

Compare a human being with a negative utility, on the one hand, with non-human beings with positive utility levels, on the other hand. If the utility level of animals is below their critical level, the world with a single miserable human can be better than the world with a lot of satisfied animals. That is the gist of the animal sadistic conclusion. As mentioned above, a population principle always entails the animal sadistic conclusion when the critical level for animals is higher than that for human beings. Can the animal sadistic conclusion be avoided if the critical level of animals is lower than that of human beings? It is not sufficient for its avoidance. Proposition 2 shows that a multi-species critical-level utilitarian ordering can avoid the animal sadistic conclusion only when the critical level for non-human beings is zero.

Finally, the conjunction of Proposition 1 and 2 implies the following result.

Proposition 3. Assume that both c_α and c_β are non-negative. A multi-species critical-level utilitarian ordering avoids the animal repugnant conclusion if and only if it entails the animal sadistic conclusion.

One might see it analogous to the general issue of population ethics—either avoiding the (non-animal) repugnant conclusion or avoiding the (non-animal) sadistic conclusion—and look for a way out from this alleged dilemma. However, in the context of multi-species population ethics, perhaps we have a case for avoiding the animal repugnant conclusion rather than its sadistic counterpart. As is mentioned, it is more plausible that $c_\alpha < c_\beta$ than the reverse, which means that the animal sadistic conclusion seems to be the correct bullet to bite.

5 Characterizing species-relative critical levels

In the previous section, we examined population axiology based on critical-level utilitarianism. To sum up, it has been shown that a utilitarian population principle with species-relative critical levels faces either the animal repugnant conclusion or the animal sadistic conclusion. This section provides a general class of multi-species population principles, which allow us to consider a resolution of the dilemma provided in Section 4. First, we examine the axiological nature of species-relative critical levels. Our approach is the *axiomatic* analysis, which is common in the field of social choice theory. To be precise, we introduce normative conditions (or axioms) on R and use them to characterize species-relative critical levels. After this, we propose the definition of what we call *multi-species critical-level generalized utilitarianism*. It is a plausible synthesis of *critical generalized utilitarianism* of Blackorby, Bossert, and Donaldson (1995) and their multi-species critical-level utilitarianism.

Axioms that we use here are natural extensions of those employed in the axiomatic analysis of population ethics (including only a single species); see Blackorby, Bossert, and Donaldson (2005). The first axiom directly requires that there is a critical level in a certain way. For example, consider the following ordering: assuming that there are only human beings, (i) if a distribution has a larger number of people than another, the former is better than the latter; (ii) if the population size is the same between the two, utilitarianism applies. None of the distributions has a critical level because the population size is absolutely prioritized. Our axiom does not allow this. That is, for each species (α or β), there must exist at least one distribution u consisting only of this species, such that the distribution is indifferent to another distribution where one individual is added to the original distribution, u. It imposes the existence of only one pair of a distribution and a critical level for each species, and thus, we call this "weak existence of critical levels." It can be formally stated as follows.

Weak existence of critical levels: (i) There exist a non-empty distribution $u_\alpha \neq u_\emptyset$ of species α and a utility level c_α such that $(u_\alpha, u_\emptyset) I ((u_\alpha, c_\alpha), u_\emptyset)$, and (ii) there exist a non-empty distribution

$u_\beta \neq u_\emptyset$ of species β and a utility level c_β such that $(u_\emptyset, u_\beta) I (u_\emptyset, (u_\beta, c_\beta))$.

The next axiom is a condition for the separability of individuals. As argued by many authors, separability is one of the key features of modern theories of distributive justice. Unconcerned individuals who obtain the same utility levels between two distributions do not matter for the ranking between the two distributions. That is, the same ranking is obtained under the distributions without these unconcerned individuals. Notably, our version of separability applies different population sizes with different species. Let us consider two distributions (u_α, u_β) and (v_α, v_β) where the former is better than the latter. The axiom of existence independence states that the existence of a third distribution (w_α, w_β) does not affect the ranking between them. That is, (u_α, u_β) is better than (v_α, v_β) if and only if $((u_\alpha, u_\beta), (w_\alpha, w_\beta))$ is better than $((v_\alpha, v_\beta), (w_\alpha, w_\beta))$. We can state this concisely as follows.

Existence independence: For all distributions u, v, w that are elements of Ω, uRv if and only if $(u, w)R(v, w)$.

Our third axiom is a condition for impartiality. Our version requires that the moral value of a distribution is invariant for any permutation over individuals in each species. For example, consider a distribution where there are two individuals for each species. We can denote this distribution by $((u_{1\alpha}, u_{2\alpha}), (u_{1\beta}, u_{2\beta}))$. According to our version of anonymity, this distribution is indifferent to $((u_{2\alpha}, u_{1\alpha}), (u_{1\beta}, u_{2\beta}))$, where the names of two individuals in species α is permuted. A similar permutation can applies to the other species. Call this condition "species-relative anonymity" because a permutation across species is not in its scope.

Species-relative anonymity: For all population sizes n, m, for all distributions $u_\alpha = (u_{1\alpha}, \ldots, u_{n\alpha})$, $v_\alpha = (v_{1\alpha}, \ldots, v_{n\alpha})$ of species α, and for all distributions $u_\beta = (u_{1\beta}, \ldots, u_{m\beta}), v_\beta = (v_{1\beta}, \ldots, v_{m\beta})$ of species β, (i) if u_α is obtained by applying a permutation to the components of v_α, then $(v_\alpha, u_\beta)I(u_\alpha, u_\beta)$, and (ii) if u_β is obtained by applying a permutation to the components of v_β, then $(u_\alpha, v_\beta)I(u_\alpha, u_\beta)$.

We now present a characterization of constant species-relative critical levels. The aforementioned three axioms are sufficient for deriving species-relative critical levels. The proof is relegated to Appendix A.

Proposition 4. If a population principle R satisfies existence independence, species-relative anonymity, and weak existence of critical levels, then there are constant species-relative critical levels.

As alluded above, species-relative anonymity should be distinguished from the stronger version of anonymity, which employs all permutations over individuals. For example, if any permutation can be used, $((u_{1\alpha}, u_{2\alpha}), (u_{1\beta}, u_{2\beta}))$ is morally indifferent to $((u_{1\alpha}, u_{1\beta}), (u_{2\alpha}, u_{2\beta}))$. That is, all living beings are impartially treated, regardless of one's species. Call this condition "inter-species anonymity." Notably, if inter-species anonymity is imposed instead of species-relative anonymity, critical levels cannot be different across species and, thus, there must be constant universal critical levels; see Blackorby, Bossert, and Donaldson (2005, p. 191, Theorem 6.9). More formally, we have the following: if a population principle R satisfies existence independence, inter-species anonymity, and the existence of critical levels, then there are constant universal critical levels. The upshot of this observation is that species-relative anonymity is crucial for species-relative critical levels.

Note that the aforementioned population principle (1) satisfies existence independence, species-relative anonymity, and the existence of critical levels. Now, consider the following population principle:

$$\sum_{i=1}^{n}(H\sqrt{u_{i\alpha}} - H\sqrt{c_\alpha}) + \sum_{i=1}^{m}(u_{i\beta} - c_\beta), \tag{2}$$

where H is a positive constant larger than one. This population principle associated with (2) also satisfies existence independence, species-relative anonymity, and the existence of critical levels. However, it is different in two respects from multi-species critical-level utilitarianism, defined in (1). First, inequality aversion is introduced for species α, while the standard utilitarian approach applies among insects. Thus, the population principle (2) has different degrees of egalitarian concern for different species. Second, the parameter H represents the priority of utilities of human beings over those of insects. If H is substantially large, the loss of one unit of human utility can be covered only by a large increase in insects' utilities. We think that this is a possibly plausible generalized form of critical-level utilitarianism The reason why we mention it here is to see how asymmetry among species is introduced. That is, H corresponds to the degree of asymmetric treatment between human beings and insects. Given this form, one can ask if a sufficiently large H is helpful to avoid the animal repugnant conclusion or animal sadistic conclusions. In particular, the animal repugnant conclusion shown in Proposition 1 may make someone wonder if priority to human beings is a possible source of resolving the dilemma of critical-level utilitarianism. (As is seen later in a general form, however, we are not suggesting it as a conclusive resolution.)

Besides (1) and (2), we can conceive of other population principles as well that satisfy the three conditions: existence independence, species-relative anonymity, and the existence of critical levels. In other words, once we weaken the "anonymity" condition, i.e., once we drop universal anonymity and take species-relative anonymity, a wide range of population principles comes into the scope. Now we introduce a general class of population principles with species-relative critical levels. An ordering R on Ω is **multi-species critical-level generalized utilitarian** if there exists a pair of concave and increasing transformations g_α and g_β applied to each individual utility values of species α and β such that, for all population sizes n, m, p, q, and for all distributions $(u_\alpha, u_\beta) = ((u_{1\alpha}, \ldots, u_{n\alpha}), (u_{1\beta}, \ldots, u_{m\beta}))$ and $(v_\alpha, v_\beta) = ((v_{1\alpha}, \ldots, v_{p\alpha}), (v_{1\beta}, \ldots, v_{q\beta}))$, $(u_\alpha, u_\beta) R (v_\alpha, v_\beta)$ if and only if

$$\sum_{i=1}^{n}[g_\alpha(u_{i\alpha}) - g_\alpha(c_\alpha)] + \sum_{i=1}^{m}[g_\beta(u_{i\beta}) - g_\beta(c_\beta)] \geq \sum_{i=1}^{p}[g_\alpha(v_{i\alpha}) - g_\alpha(c_\alpha)] + \sum_{i=1}^{q}[g_\beta(v_{i\beta}) - g_\beta(c_\beta)].$$

The two transforming functions, g_α and g_β, are utility weighting functions, which can be different between different species. The concavity of these functions guarantee the prioritarian nature of this class.[9] Notably, multi-species critical-level generalized utilitarian orderings satisfy existence independence, species-relative anonymity, and the existence of critical levels; thus, it has constant species-relative critical levels.[10] Propositions 1–3 are valid for this class of population principles; the proofs in Appendix A apply to this general class. The severe dilemma holds; the animal repugnant conclusion is avoided if and only if the animal sadistic conclusion.

6 Resolution with lexical views and the "theory X"

Is there any population principle that simultaneously avoids *both* the animal repugnant conclusion *and* the animal sadistic conclusion? Let us consider another class of population principles outside the scope of multi-species critical-level generalized utilitarian orderings. Parfit (1984, sec. 140) suggests the "lexical" view, according to which the utility of individuals in one species has lexical moral priority over that of individuals in another species, regardless of their utility levels.[11] As a

[9] Brown (2007) examines a utility weighting function for prioritarianism in a variable-population setting.

[10] Here, the principal purpose of introducing these functions is to show the validity of our propositions under a general class of population principles (rather than to defend the particular class of prioritarian principles as such).

[11] In his more recent paper, Parfit (2016, p. 118) considered a comparison between animals ("earliest sentient animals") and human beings.

typical interpretation, it holds that the utility of human individuals has moral priority over that of non-human individuals.

Parfit (1984, sec. 141) seems to have in mind a simple choice; either to employ critical levels or to take the lexical view. However, we can combine the lexical feature into critical-level generalized utilitarianism. Even in the case of a single species, the critical-level lexicographic ordering is well-defined (Blackorby, Bossert, and Donaldson, 1996). For any two populations with different sizes, say n and m ($n > m$), the critical-level lexicographic ordering applies the lexicographic comparison after adding $n - m$ people with critical levels to the smaller population. This ordering obviously has a constant critical level, and it is "lexical." However, it does not exhibit the aggregative nature, which is the core of utilitarianism. In the case of multiple species, the lexical approach can be combined with utilitarianism. The following two-step procedure is considered for any two distribution, A and B. First, if the sub-distribution of human beings in A is better than that in B according to critical-level (generalized) utilitarianism for human beings, A is morally better than B. Second, if the sub-distributions of human beings are indifferent, critical-level (generalized) utilitarianism for insects applies to the sub-distributions of insects. Notably, the "lexical" view holds for inter-species comparisons, while the critical-level utilitarianism applies to intra-species comparisons.[12]

Formally, an ordering R on Ω is **species-lexical critical-level generalized utilitarian** if if there exists a pair of concave and increasing transformations g_α and g_β applied to each individual utility values of species α and β such that, for all population sizes n, m, p, q, and for all distributions $(u_\alpha, u_\beta) = ((u_{1\alpha}, \ldots, u_{n\alpha}), (u_{1\beta}, \ldots, u_{m\beta}))$ and $(v_\alpha, v_\beta) = ((v_{1\alpha}, \ldots, v_{p\alpha}), (v_{1\beta}, \ldots, v_{q\beta}))$, $(u_\alpha, u_\beta) R (v_\alpha, v_\beta)$ if and only if

$$\sum_{i=1}^{n}[g_\alpha(u_{i\alpha}) - g_\alpha(c_\alpha)] > \sum_{i=1}^{m}[g_\alpha(v_{i\alpha}) - g_\alpha(c_\alpha)]$$

or

$$\sum_{i=1}^{n}[g_\alpha(u_{i\alpha}) - g_\alpha(c_\alpha)] = \sum_{i=1}^{m}[g_\alpha(v_{i\alpha}) - g_\alpha(c_\alpha)]$$

and

$$\sum_{i=1}^{p}[g_\beta(u_{i\beta}) - g_\beta(c_\beta)] \geq \sum_{i=1}^{q}[g_\beta(v_{i\beta}) - g_\beta(c_\beta)].$$

[12] A similar nature is found in a certain type of sufficientarianism; see Bossert, Cato, and Kamaga (forthcoming). Under this type of sufficientarianism, the entire set of individuals is divided into two categories: those who are below the sufficiency threshold, and those who are above or at the sufficiency threshold. The lexical applies to inter-category comparisons, while the critical-level utilitarianism applies to intra-category comparisons.

To avoid confusion, we hasten to stress that our species-lexical critical-level generalized utilitarian ordering is an extension of the standard critical-level generalized utilitarian orderings proposed by Blackorby, Bossert, and Donaldson (1997, 2005). In particular, if species β does not exist, the ordering will be the same as a standard critical-level generalized utilitarian ordering for species α. That is, $(u_\alpha, u_\emptyset)R(v_\alpha, u_\emptyset)$ if and only if $\sum_{i=1}^{n}[g_\alpha(u_{i\alpha}) - g_\alpha(c_\alpha)] \geq \sum_{i=1}^{m}[g_\alpha(v_{i\alpha}) - g_\alpha(c_\alpha)]$. Moreover, if species α does not exist, the ordering will be the same as a critical-level generalized utilitarian ordering for species β. That is, $(u_\emptyset, u_\beta)R(v_\emptyset, u_\beta)$ if and only if $\sum_{i=1}^{p}[g_\beta(u_{i\beta}) - g_\beta(c_\beta)] \geq \sum_{i=1}^{q}[g_\beta(v_{i\beta}) - g_\beta(c_\beta)]$.

In our formulation of species-lexical critical-level generalized utilitarianism, c_α and c_β are constant critical-levels for species α and β, respectively. Thus the condition, weak existence of critical levels, is satisfied. Next, species-relative anonymity is satisfied because individuals are symmetrically treated within each species. Finally, we can show that existence independence is also satisfied; see Appendix A. In sum, a species-lexical critical-level generalized utilitarian ordering satisfies existence independence, species-relative anonymity, and weak existence of critical levels.

What is the implication of this proposal to multi-species population ethics? The following proposition shows that no condition over species-relative critical levels to avoid animal repugnant/sadistic conclusions for species-lexical critical-level generalized utilitarian orderings; see Appendix A for the proof.

Proposition 5. Assume that c_α and c_β are non-negative. A species-lexical critical-level generalized utilitarian ordering avoids both the animal repugnant/sadistic conclusions.

That seems to be a welcome result. Is this lexical view an ultimate solution? One might challenge that such a lexical view is utterly speciesist. That is, a species-lexical critical-level generalized utilitarian ordering might be seen as morally flawed in that it gives moral priority to human lives (*qua* human lives) over non-human lives. Such a view requires a justification; otherwise, it would count as discrimination against non-human beings, as the objection goes.

In response, first, our discussion about lexical views need not apply to *all* non-human beings. If the relevant comparison is between human beings and (say) whales, a moral priority to human lives over whales' lives may sound morally problematic. That said, there could be *some* kind of non-human beings such that a moral priority to human lives over lives of the non-human beings is justifiable. Perhaps because of this observation, Williamson (2021) compared the lives of human beings with those of insects rather than other "higher" animals. In that case, we can restrict the scope of our discussion thus far; that is, we are considering the animal repugnant/sadistic conclusions regarding two species, α and β, that permit such a lexical moral priority.

Second, the charge of speciesism can be relevant in a different form in population ethics. For instance, how c_α compares to c_β could be seen as a matter of speciesism. Some might find it speciesist that the critical level for human beings is lower than that for non-human beings (i.e., $c_\alpha < c_\beta$). It is speciesist, goes the objection, in that the existence of a human individual with utility level μ would be morally better than that of a non-human individual with utility level μ when $c_\alpha < \mu < c_\beta$. Thus, a notable feature of species-lexical critical-level generalized utilitarianism is that it bypasses the complicated issue comparing c_α and c_β. Alternatively, one could see the speciesist nature of the view that $c_\alpha < c_\beta$ as *less* problematic than that of lexical moral priority. We do not mean to settle the "which is more speciesist" question. Our point is that we must choose between two population principles that are more or less speciesist; multi-species critical-level generalized utilitarianism (involving the view that $c_\alpha < c_\beta$) on the one hand, and species-lexical critical-level generalized utilitarianism (involving a lexical moral priority to human beings) on the other hand.

Third, and finally, it is not surprising at all if we cannot have the "theory X" in multi-species population ethics, given that we do not have one even in human-only population ethics. The "theory X," a (human) population principle that is plausible in every respect, must satisfy several requisites such as avoiding the repugnant conclusion and avoiding the sadistic conclusion.[13] These requisites also apply to the multi-species version of the "theory X," *mutatis mutandis*; it must avoid the animal repugnant conclusion, it must avoid the animal sadistic conclusion, along with other requisites. On top of them, however, the multi-species "theory X" must satisfy another requisite, according to the charge of speciesism; it must avoid speciesism. Perhaps, our discussion thus far has shown that finding such a multi-species "theory X" becomes a difficult task if we hope to steer clear from speciesism.

7 Concluding remarks

We examined a general class of critical-level utilitarian principles that applies to multi-species population ethics. First, the conditions were clarified under which there are species-relative critical

[13] Remember a population principle provided by Sider (1991). It avoids the mere addition paradox, and thus, it seems to be a good candidate for the "theory X." However, it exhibits equality aversion. That is, anti-progressive transfers are considered normatively desirable.

levels. Second, we illustrated how a multi-species critical-level general utilitarian ordering faces a dilemma; it entails either the animal repugnant conclusion or the animal sadistic conclusion, depending on whether the constant critical level for human beings is zero or not. This observation implies a correlation between the animal and human variants of the repugnant conclusion. Also, it is a crucial question of multi-species population ethics whether the critical level for human beings is higher or lower than that for non-human beings. We articulated the moral implications of the two possibilities. Third, a species-lexical critical-level general utilitarianism is examined. It can avoid all negative conclusions, i.e., the animal repugnant/sadistic conclusions, and the human, ordinary repugnant conclusion.

Concerning multi-species critical-level general utilitarianism, our study (for now) inclines us to accept the animal sadistic conclusion, avoiding both the human repugnant conclusion and the animal repugnant conclusion. It may sound speciesist, which we do not deny. Moreover, the lexical approach proposed in this study also has the same flavor, although it can avoid the animal sadistic conclusion. Perhaps this upshot outlines a new kind of dilemma for multi-species population ethics; either a multi-species population principle entails an *implausible* conclusion (such as the repugnant conclusion), or it entails a *morally problematic* conclusion (such as speciesism).

Appendix A: technical materials

We prove Propositions 1 and 2 for a species-relative critical-level generalized utilitarian ordering. The case for multi-species critical-level utilitarianism is obtained as a special case.

Proof of Proposition 1. Assume that $c_\alpha > c_\beta$. We should distinguish two cases: (a) $c_\beta = 0$ and (b) $c_\beta > 0$. Assume that $c_\beta = 0$. Take any positive integer n, a positive number ξ, and a positive number ε less than ξ. Since g_β is increasing, $g_\beta(\varepsilon) - g_\beta(c_\beta)$ is positive. Thus, $m[g_\beta(\varepsilon) - g_\beta(c_\beta)] > n[g_\alpha(\xi) - g_\alpha(c_\alpha)]$ for a sufficiently large number m. Thus, there exists a sufficiently large positive integer m such that $(u_\emptyset, \varepsilon\mathbf{1}_m)P(\xi\mathbf{1}_n, u_\emptyset)$. The animal repugnant conclusion holds. Next, assume that $c_\beta > 0$. Take $\xi > c_\alpha$ and $\varepsilon > 0$ such that $c_\beta > \varepsilon$ holds. Since g_α and g_β are increasing, $g_\alpha(\xi) - g_\alpha(c_\alpha)$ is positive and $g_\beta(\varepsilon) - g_\beta(c_\beta)$ is negative. Obviously, $n[g_\alpha(\xi) - g_\alpha(c_\alpha)] > m[g_\beta(\varepsilon) - g_\beta(c_\beta)]$ for all integers n, m. Thus, $(\xi\mathbf{1}_n, u_\emptyset)P(u_\emptyset, \varepsilon\mathbf{1}_m)$. The animal repugnant conclusion is avoided.

Next, assume that $c_\alpha < c_\beta$. Let $\xi > c_\alpha$ and $\varepsilon > 0$ such that $\varepsilon < c_\alpha$. Because $c_\alpha < c_\beta$, it holds that $c_\beta > \varepsilon$. Since g_α and g_β are increasing, $g_\alpha(\xi) - g_\alpha(c_\alpha)$ is positive and $g_\beta(\varepsilon) - g_\beta(c_\beta)$ is negative. This implies that for all positive integers n, m, it holds that $(\xi\mathbf{1}_n, u_\emptyset)P(u_\emptyset, \varepsilon\mathbf{1}_m)$ because

$n[g_\alpha(\xi) - g_\alpha(c_\alpha)] > 0 > m[g_\beta(\varepsilon) - g_\beta(c_\beta)]$. Thus, the animal repugnant conclusion is avoided. ■

Proof of Proposition 2. First, consider the case where $c_\alpha > c_\beta$. Assume that $c_\beta = 0$. Take any $\varepsilon < 0$ and $\xi > 0$. Notably, $g_\alpha(\xi) - g_\alpha(c_\beta)$ is always positive because g_α is increasing and $\xi > c_\beta = 0$. Then, for any sufficiently large integer m, it holds that $m[g_\alpha(\xi) - g_\alpha(c_\beta)] > 0 > [g_\beta(\varepsilon) - g_\beta(c_\alpha)]$. Then $(u_\emptyset, \xi\mathbf{1}_m)P(\varepsilon\mathbf{1}_1, u_\emptyset)$. The animal sadistic conclusion is avoided. Then, assume that that $c_\beta > 0$. Take any $\varepsilon < 0$. There exists $\xi > 0$ such that $g_\beta(c_\beta) > g_\beta(\xi)$ since $c_\beta > 0$ and g_β is increasing. Then, for any sufficiently large integer m^*, $[g_\alpha(\varepsilon) - g_\alpha(c_\alpha)] > m^*[g_\beta(\xi) - g_\beta(c_\beta)]$ because the term $m[g_\beta(\xi) - g_\beta(c_\beta)]$ approaches negative infinity as m approaches to ∞. Moreover, $[g_\alpha(\varepsilon) - g_\alpha(c_\alpha)] > m[g_\beta(\xi) - g_\beta(c_\beta)]$ for all integers m greater than m^*. The animal sadistic conclusion holds.

Second, we turn to the case where $c_\alpha < c_\beta$. Take any $\varepsilon > 0$ and $\xi > 0$ such that $c_\beta > \xi$. Since g_β is increasing, $g_\beta(\xi) - g_\beta(c_\beta)$ is negative. Then, for a sufficiently large integer m^*, it holds that $[g_\alpha(\varepsilon) - g_\alpha(c_\alpha)] > m^*[g_\beta(\xi) - g_\beta(c_\beta)]$ because $m[g_\beta(\xi) - g_\beta(c_\beta)]$ approaches negative infinity as m approaches to ∞. For all integers m greater than m^*, $[g_\alpha(\varepsilon) - g_\alpha(c_\alpha)] > m[g_\beta(\xi) - g_\beta(c_\beta)]$ holds, and thus, the animal sadistic conclusion holds. ■

Proof of Proposition 4. Assume that R satisfies existence independence, species-relative anonymity, and weak existence of critical levels. Since the case for species β can be parallelly examined, we focus on the case for species α. By the first requirement of weak existence of critical levels, there exist a non-empty distribution $u_\alpha \neq u_\emptyset$ of species α and a utility level c_α such that $(u_\alpha, u_\emptyset)I((u_\alpha, c_\alpha), u_\emptyset)$. Let us take any distribution $v = (v_\alpha, u_\beta)$. Existence independence implies that

$$(u_\alpha, u_\emptyset)I((u_\alpha, c_\alpha), u_\emptyset) \iff ((u_\alpha, v_\alpha), v_\beta)I((u_\alpha, v_\alpha, c_\alpha), v_\beta).$$

This means that c_α is a critical level associated with $((u_\alpha, v_\alpha), v_\beta)$. Now, we can apply existence independence again to this. Then,

$$((u_\alpha, v_\alpha), v_\beta)I((u_\alpha, v_\alpha, c_\alpha), v_\beta) \iff (v_\alpha, v_\beta)I((v_\alpha, c_\alpha), v_\beta).$$

As a result, we obtain $(v_\alpha, v_\beta)I((v_\alpha, c_\alpha), v_\beta)$. Notably, (v_α, v_β) is chosen arbitrarily. Thus, $(v_\alpha, v_\beta)I((v_\alpha, c_\alpha), v_\beta)$ holds for all $(v_\alpha, v_\beta) \in \Omega$. This implies that c_α is a constant critical level for species α. ■

Proof of Proposition 5. First, we show that a species-lexical critical-level utilitarian ordering avoids the animal repugnant conclusion. Take a positive integer n, a positive number ξ

greater than c_α, and a positive number ε less than ξ. For all positive integers m, it holds that $(\xi\mathbf{1}_n, u_\emptyset)P(u_\emptyset, \varepsilon\mathbf{1}_m)$ because $n[g_\alpha(\xi)-g_\alpha(c_\alpha)] > 0$ and the first clause of species-lexical critical-level utilitarianism applies to this. This implies that the animal repugnant conclusion is avoided.

Next, we consider the statement on the animal sadistic conclusion. Take a negative number ε and a positive number ξ. Since $c_\alpha \geq 0$, it holds that $c_\alpha > \varepsilon$. the first clause of species-lexical critical-level utilitarianism applies to this, and thus, the only difference distribution of species α matters. Since $g_\alpha(\varepsilon) - g_\alpha(c_\alpha)$ is negative, we obtain $(u_\emptyset, \xi\mathbf{1}_m) > (\varepsilon\mathbf{1}_1, u_\emptyset)$ for all positive integers m. Thus, the animal sadistic conclusion is avoided. ■

A species-lexical critical-level generalized utilitarian ordering satisfies existence independence. Assume that $(u_\alpha, u_\beta)R(v_\alpha, v_\beta)$. Take any $w_\alpha = (w_{1\alpha}, \ldots, w_{s\alpha})$ and $w_\beta = (w_{1\beta}, \ldots, w_{t\beta})$. We now show that

$$((u_\alpha, u_\beta), (w_\alpha, w_\beta))R((v_\alpha, v_\beta), (w_\alpha, w_\beta)).$$

Since $(u_\alpha, u_\beta)R(v_\alpha, v_\beta)$, it holds that

$$\sum_{i=1}^{n}[g_\alpha(u_{i\alpha}) - g_\alpha(c_\alpha)] > \sum_{i=1}^{m}[g_\alpha(v_{i\alpha}) - g_\alpha(c_\alpha)]$$

or

$$\sum_{i=1}^{n}[g_\alpha(u_{i\alpha}) - g_\alpha(c_\alpha)] = \sum_{i=1}^{m}[g_\alpha(v_{i\alpha}) - g_\alpha(c_\alpha)]$$

and

$$\sum_{i=1}^{p}[g_\beta(u_{i\beta}) - g_\beta(c_\beta)] \geq \sum_{i=1}^{q}[g_\beta(v_{i\beta}) - g_\beta(c_\beta)].$$

First, assume that $\sum_{i=1}^{n}[g_\alpha(u_{i\alpha}) - g_\alpha(c_\alpha)] > \sum_{i=1}^{m}[g_\alpha(v_{i\alpha}) - g_\alpha(c_\alpha)]$. Then,

$$\sum_{i=1}^{n}[g_\alpha(u_{i\alpha}) - g_\alpha(c_\alpha)] + \sum_{i=1}^{s}[g_\alpha(w_{i\alpha}) - g_\alpha(c_\alpha)] > \sum_{i=1}^{m}[g_\alpha(v_{i\alpha}) - g_\alpha(c_\alpha)] + \sum_{i=1}^{s}[g_\alpha(w_{i\alpha}) - g_\alpha(c_\alpha)]$$

Then, $((u_\alpha, u_\beta), (w_\alpha, w_\beta))R((v_\alpha, v_\beta), (w_\alpha, w_\beta))$. Next, we assume that $\sum_{i=1}^{n}[g_\alpha(u_{i\alpha}) - g_\alpha(c_\alpha)] = \sum_{i=1}^{m}[g_\alpha(v_{i\alpha}) - g_\alpha(c_\alpha)]$ and $\sum_{i=1}^{p}[g_\beta(u_{i\beta}) - g_\beta(c_\beta)] \geq \sum_{i=1}^{q}[g_\beta(v_{i\beta}) - g_\beta(c_\beta)]$. It holds that

$$\sum_{i=1}^{n}[g_\alpha(u_{i\alpha}) - g_\alpha(c_\alpha)] + \sum_{i=1}^{s}[g_\alpha(w_{i\alpha}) - g_\alpha(c_\alpha)] = \sum_{i=1}^{m}[g_\alpha(v_{i\alpha}) - g_\alpha(c_\alpha)] + \sum_{i=1}^{s}[g_\alpha(w_{i\alpha}) - g_\alpha(c_\alpha)]$$

and

$$\sum_{i=1}^{p}[g_\beta(u_{i\beta}) - g_\beta(c_\beta)] + \sum_{i=1}^{t}[g_\beta(w_{i\beta}) - g_\beta(c_\beta)] \geq \sum_{i=1}^{q}[g_\beta(v_{i\beta}) - g_\beta(c_\beta)] + \sum_{i=1}^{t}[g_\beta(w_{i\beta}) - g_\beta(c_\beta)].$$

Then, $((u_\alpha, u_\beta), (w_\alpha, w_\beta))R((v_\alpha, v_\beta), (w_\alpha, w_\beta))$. Then, if uRv, then $(u, w)R(v, w)$. We can prove that uRv if $(u, w)R(v, w)$. Thus, existence independence is satisfied by the species-lexical critical-level generalized utilitarian ordering.

Appendix B: the flipped repugnant conclusion

In addition to the animal repugnant conclusion, Williamson (2021) criticized critical-level utilitarianism in means of what he called the "flipped repugnant conclusion":

> 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. (Williamson 2021, p. 411)

For us, this criticism is not as serious as it may seem because the argument depends on the existence of constant critical levels as long as one interprets it literally. In general, there are various population principles without constant critical levels. The direct interpretation of the flipped repugnant conclusion cannot apply to them. Therefore, the avoidance of the flipped repugnant conclusion is not a general normative requirement, although it helps understand a feature of critical-level utilitarianism.

This appendix attempts to reformulate the flipped repugnant conclusion as a general condition. Let us consider a single mayfly with a positive utility $\varepsilon > 0$, and a human population where all of n human-beings obtains a higher utility level, ξ. Our reformulated version of the flipped repugnant conclusion is entailed if and only if there *exist* ξ and ε such that the single mayfly is better than the human society for any sufficiently large n. That is formally stated as follows.

The flipped repugnant conclusion (reformulated version). There exist a positive number ε, a positive number ξ larger than ε, and a positive integer n^* such that for all positive integers n larger than n^*,

$$(u_\emptyset, \varepsilon \mathbf{1}_1) P (\xi \mathbf{1}_n, u_\emptyset),$$

where $(u_\emptyset, \varepsilon \mathbf{1}_1)$ is a single-individual distribution of species β and $(\xi \mathbf{1}_n, u_\emptyset)$ is a distribution of species α.

Notably, this condition can apply to any class of population principles. On the other hand, because of the existential quantifier, which is necessary for the general applicability of the flipped repugnant conclusion, it seems that the reformulated version is not very repugnant.

Now, we provide the proposition that identifies conditions under which the flipped repugnant conclusion holds.

Proposition 6. Assume that both c_α and c_β are non-negative. If an ordering R is multi-species critical-level generalized utilitarian or species-lexical critical-level generalized utilitarian, the flipped repugnant conclusion is avoided if and only if $c_\alpha = 0$.

Proof. (i) Let R be multi-species critical-level generalized utilitarian. Assume that $c_\alpha > 0$. Take $\varepsilon, \xi > 0$ such that $c_\alpha > \xi > \varepsilon$. Since g_α are increasing, $g_\alpha(\xi) - g_\alpha(c_\alpha)$ is negative. Then, there exists a large integer n^* such that $g_\beta(\varepsilon) - g_\beta(c_\beta) > n^*[g_\alpha(\xi) - g_\alpha(c_\alpha)]$. Thus, for all $n > n^*$, $(u_\emptyset, \varepsilon\mathbf{1}_1)P(\xi\mathbf{1}_n, u_\emptyset)$. The flipped repugnant conclusion holds. Assume that $c_\alpha = 0$. Take any $\varepsilon > 0$ and $\xi > \varepsilon$. It holds that $\xi > \varepsilon > c_\alpha$. Thus, $g_\alpha(\xi) - g_\alpha(c_\alpha)$ is positive. For any large integer n, $(\xi\mathbf{1}_n, u_\emptyset)P(u_\emptyset, \varepsilon\mathbf{1}_1)$ holds because $n[g_\alpha(\xi) - g_\alpha(c_\alpha)] > g_\alpha(\varepsilon) - g_\alpha(c_\alpha)$. The flipped repugnant conclusion is avoided.

(ii) Let R be species-lexical critical-level generalized utilitarian. Assume that $c_\alpha > 0$. Let $\varepsilon > 0$ and $\xi > \varepsilon$ such that $c_\alpha > \xi > \varepsilon$. Since g_α is increasing, $g_\alpha(\xi) - g_\alpha(c_\alpha)$ is negative. Since $0 > n[g_\alpha(\xi) - g_\alpha(c_\alpha)]$ for all integers n, $(u_\emptyset, \varepsilon\mathbf{1}_1)P(\xi\mathbf{1}_n, u_\emptyset)$ for all integers n. The flipped repugnant conclusion holds. Now, assume that $c_\alpha = 0$. For any positive number ξ, $n[g_\alpha(\xi) - g_\alpha(c_\alpha)] > 0$ for any number n. Thus, for all positive integers n, $(\xi\mathbf{1}_n, u_\emptyset)P(u_\emptyset, \varepsilon\mathbf{1}_1)$. The flipped repugnant conclusion does not hold. ■

Note that, for both classes of population principles, the only way to avoid the flipped repugnant conclusion is to make c_α to be equal to zero. This proposition implies that if $c_\alpha = 0$, then a species-lexical critical-level generalized utilitarian ordering avoids the flipped repugnant conclusion as well as the animal repugnant/sadistic conclusions. That seems to be a positive result. However, a well-known problem arises if the constant critical level for human beings is zero. As seen in Section 2, the repugnant conclusion happens in its most ordinary sense. Alternatively, we may assume that $c_\alpha > 0$. In that case, a species-lexical critical-level generalized utilitarian ordering avoids the repugnant conclusion, but it cannot avoid the animal sadistic conclusion.

References

[1] Arrhenius, G. (2000). An impossibility theorem for welfarist axiologies. Economics and Philosophy, 16, 247–266.

[2] Arrhenius, G. (forthcoming). Population Ethics: The Challenge of Future Generations. Oxford: Oxford University Press.

[3] Blackorby, C., and Donaldson, D. (1984). Social criteria for evaluating population change. Journal of Public Economics, 25, 13–33.

[4] Blackorby, C., Bossert, W., and Donaldson, D. (1995). Intertemporal population ethics: critical-level utilitarian principles. Econometrica, 63, 1303-1320.

[5] Blackorby, C., Bossert, W., and Donaldson, D. (1996). Leximin population ethics. Mathematical Social Sciences, 31(2), 115–131.

[6] Blackorby, C., Bossert, W., and Donaldson, D. (1997). Critical-level utilitarianism and the population-ethics dilemma. Economics & Philosophy, 13, 197–230.

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

[8] Blackorby, C., Bossert, W., Donaldson, D., and Fleurbaey, M. (1998). Critical levels and the (reverse) repugnant conclusion. Journal of Economics, 67, 1–15.

[9] Bossert, W., Cato, S., and Kamaga, K. (2021). Revisiting variable-value population principles. Institute of Social Science, University of Tokyo, Discussion Paper Series for Institute-Wide Joint Research Project "Methodology of Social Sciences: Measuring Phenomena and Values," E-21-007. Available at `https://web.iss.u-tokyo.ac.jp/methodology/en/dp/dp/202109/E-21-007.html`

[10] Bossert, W., Cato, S., and Kamaga, K. (forthcoming). Critical-level Sufficientarianism. Journal of Political Philosophy. DOI: https://doi.org/10.1111/jopp.12267

[11] Broome, J. (1993). Goodness is reducible to betterness: The evil of death is the value of life. In: Koslowski, P. and Shionoya, Y. (eds.). The Good and the Economical: Ethical Choices in Economics and Management, pp. 69–83. Berlin: Springer.

[12] Broome, J. (2004). Weighing Lives. Oxford: Oxford University Press.

[13] Brown, C. (2007). Prioritarianism for variable populations. Philosophical Studies, 134, 325–361.

[14] Carlson, E. (1998). Mere addition and two trilemmas of population ethics. Economics & Philosophy, 14(2), 283–306.

[15] Dorsey, D. (2010). Three arguments for perfectionism. Noûs, 44, 59–79.

[16] Hurka, T. (1983). Value and population size. Ethics, 93, 496–507.

[17] Ng, Y.-K. (1986). Social criteria for evaluating population change: an alternative to the Blackorby–Donaldson criterion. Journal of Public Economics, 29, 375–381.

[18] Ng, Y.-K. (1989). What should we do about future generations? Impossibility of Parfit's theory X. Economics & Philosophy, 5, 235–253.

[19] Parfit, D. (1976). On doing the best for our children. In: Bayles, M.D. (ed.). Ethics and Population, pp. 100–102. Cambridge, MA: Schenkman.

[20] Parfit, D. (1982). Future generations, further problems. Philosophy and Public Affairs, 11, 113–172.

[21] Parfit, D. (1984). Reasons and Persons. Oxford: Oxford University Press.

[22] Parfit, D. (2016). Can we avoid the repugnant conclusion?. Theoria, 82, 110–127.

[23] Sider, T. R. (1991). Might theory X be a theory of diminishing marginal value?. Analysis, 51(4), 265–271.

[24] Spears, D., and Budolfson, M. (2021). Repugnant conclusions. Social Choice and Welfare, 57, 567–588.

[25] Stewart, R.T. (forthcoming). Path independence and a persistent paradox of population ethics. Journal of Philosophy.

[26] Vallentyne, P. (1993). The connection between prudential goodness and moral permissibility. Journal of Social Philosophy, 24, 105–128.

[27] Vallentyne, P. (2009). Broome on moral goodness and population ethics. Philosophy and Phenomenological Research, 78, 739–746.

[28] Williamson, P. (2021). A new argument against critical-level utilitarianism. Utilitas, 33, 399–416.