Discussion Paper Series

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

A simple impossibility theorem in population ethics

Susumu Cato
University of Tokyo

Ko Harada
University of Tokyo

May 25, 2023

E-23-001

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

A simple impossibility theorem in population ethics*

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

Ko Harada
Graduate School of Economics
University of Tokyo
7-3-1, Hongo, Bunkyo-ku
Tokyo 113-0033
Japan
ko.harada11@gmail.com

This version: May 15, 2023

Abstract. This paper establishes a simple impossibility result for welfaristic evaluations when the population varies. We consider a weak version of the repugnant conclusion instead of the commonly used version. It is shown that if a population principle satisfying two reasonable properties avoids the sadistic conclusion, then the weak repugnant conclusion must hold. We use a general variable-population setting where the identities of individuals can matter. *Journal of Economic Literature* Classification Nos.: D31, D63.

Keywords: variable population, repugnant conclusion, sadistic conclusion, utilitarianism, identity problem.

* Financial support from KAKENHI through grants Nos. JP20H01446, JP20K01565, JP22K01387, JP22H05083, and JP22H05086 is gratefully acknowledged.

1 Introduction

The fundamental observation in population ethics is offered by Parfit (1984). In a variable-population setting, total utilitarianism ends up with the repugnant conclusion, according to which for any large population, say P, with very high utility levels, there is a much larger population, say Q, with hardly worth living such that Q is better than P; see Blackorby, Bossert, Donaldson, and Fleurbaey (1998) for its formal and general argument.[1] A solution for this problem is to use average utilitarianism or its variants. However, this modification leads to another negative consequence. Adding people with negative utilities is better than adding people with positive utilities. This is called the sadistic conclusion (Arrhenius, 2000). In the field of population ethics, it has been recognized that there is a tension between avoidance of the repugnant conclusion and avoidance of the sadistic conclusion, and many have offered impossibility results; see Arrhenius (2000) and Bossert, Cato, and Kamaga (2022a).[2]

This paper establishes a simple impossibility theorem associated with this tension. In particular, we show that a population principle satisfying two reasonable properties entails the sadistic conclusion or a weak version of the repugnant conclusion. The weak version states that adding a large population with very high utility levels is worse than adding a much larger population with hardly worth living. One of the two properties is egalitarian dominance, which is substantially weaker than the Pareto principle. This property is uncontroversial.[3] The other is the limit property, which essentially requires that a certain type of aggregation applies if the number of individuals increases. This axiom is satisfied by most principles; an exception is the maximin principle.[4] It is noteworthy that weakening the repugnant conclusion is necessary for obtaining the impossibility result. To show this point, we offer a counterexample (after our theorem), which offers a population principle that satisfies the two properties and avoids both the repugnant and sadistic conclusions.

Our theorem is closely related to several existing results. An earlier impossibility theorem is offered by Arrhenius (2000). This theorem essentially uses an axiom that requires inequality aversion, while this paper does not rely on such an axiom. A related result by Blackorby, Bossert, and Donaldson (2005) focuses only on the class of utilitarian principles. Bossert, Cato, and Kamaga (2022a) offer an impossibility result that uses the limit property, but they use the large but specific class of population principles. All of the aforementioned works use the original version of the repugnant conclusion, not the weak version. This version is introduced by Arrhenius (2009) in the framework where utility values are discrete, not continuous.[5] Some of Arrhenius's theorems do not use inequality aversion

[1] See also Blackorby, Bossert, and Donaldson (2005).

[2] For the significance of this type of impossibility, see Zuber et al. (2021). Cato and Ishida (2022) offer a new formal analysis related to the two conclusions in the framework with multi-species.

[3] See Cato (2023) for the logical equivalence of Paretian axioms.

[4] It is known that the maximin principle avoids the repugnant and sadistic conclusions; see Asheim and Zuber (2014), Zuber (2018), and Bossert, Cato, and Kamaga (2022a). See also Bossert (1990). A recent work by Bossert, Cato, and Kamaga (2022b) demonstrates that a sufficientarian principle can avoid both conclusions. For another direction, see Pivato (2020).

[5] This is originally introduced by Arrhenius's Ph.D.thesis: Future Generations—A Challenge for Moral Theory. Ph.D.thesis, Uppsala University (2000).

but employ different types of axioms (he calls them non-elitism and general non-extreme priority).

Finally, we stress that, to the best of our knowledge, this paper is the first paper that uses a general variable-population setting to establish the incompatibility of avoidance of the repugnant conclusion and avoidance of the sadistic conclusion. This general setting is developed by Blackorby, Bossert, Donaldson, and Fleurbaey (1998), who establish another paradox in population ethics and its variation. As an advantage of their setting, it does not impose anonymity in advance; thus, it can distinguish two populations of the same size. Our result shows that the impossibility can be established even in such a setting with simple, rather uncontroversial axioms.

The rest of this paper is organized as follows. Section 2 explains the setting and axioms that we use. Section 3 offers the main theorem and its proof.

2 Preliminaries

Let us consider a general setting with variable populations. Let $\mathbb{N}$ be the set of natural numbers and $\mathbb{R}$ the set of real numbers. Let $\mathcal{N}$ be the collection of nonempty and finite subsets of $\mathbb{N}$. Each element N of $\mathcal{N}$ represents a population (a set of individuals). For each $i \in \mathbb{N}$, $u_i \in \mathbb{R}$ represents individual i's utility level. The collection of distributions is given as follows:

$$\Omega = \bigcup_{N \in \mathcal{N}} \mathbb{R}^N$$

For each element of Ω, there exist the corresponding set N of individuals and a profile of their utility levels, $(u_i)_{i \in N}$. For simplicity, $(u_i)_{i \in N}$ is written as u_N. Let $\mathcal{N}_0$ be the set of all subsets including the emptyset. For convenience, let $\mathbb{R}^{\emptyset} = \{u_{\emptyset}\}$, where $u_{\emptyset}$ is the null utility vector.

A population principle R is a transitive and complete binary relation over Ω. Notably, $u_N R v_M$ means that distribution u_N for population N "is at least as good as" distribution v_M for population M. The asymmetric and symmetric parts of R are denoted by P and I, respectively.

We say that the *repugnant conclusion* holds for R if, for all $\xi \in \mathbb{R}_{++}$, for all $N \in \mathcal{N}$, and for all $\varepsilon \in (0, \xi)$, there exists $M \supsetneq N$ such that

$$\varepsilon \mathbf{1}_M P \xi \mathbf{1}_N.$$

Many variants of the repugnant conclusion have been considered; see, for example, Spears and Budolfson (2021). We consider one of them. The *weak repugnant conclusion* holds for R if there exist $L \in \mathcal{N}_0$ and $u_L \in \mathbb{R}^L$ such that, for all $\xi \in \mathbb{R}_{++}$, for all $N \in \mathcal{N}$ with $N \cap L = \emptyset$, and for all $\varepsilon \in (0, \xi)$, there exists $M \supsetneq N$ such that $M \cap L = \emptyset$ and

$$(u_L, \varepsilon \mathbf{1}_M) P (u_L, \xi \mathbf{1}_N).$$

If the repugnant conclusion holds, then the weak repugnant conclusion holds. Note that if a population principle avoids the weak repugnant conclusion, it avoids the repugnant conclusion.

We say that the *sadistic conclusion* holds for R if there exist $N, M, L \in \mathcal{N}$, $u_N \in \mathbb{R}^N_{++}$, $v_M \in \mathbb{R}^M_{--}$, and $w_L \in \mathbb{R}^L$ such that $L \cap N = \emptyset$, $L \cap M = \emptyset$, and

$$(v_M, w_L) P (u_N, w_L).$$

As mentioned earlier, this was originally proposed by Arrhenius (2000). A recent work by Franz and Spears (2020) examines the implications of the sadistic conclusion.

Now, we introduce two axioms. The first one requires that among egalitarian distributions, a distribution with higher utility levels is better than a distribution with lower utility levels.

Egalitarian Dominance. For all $N \in \mathcal{N}$, and for all $\mu, \nu \in \mathbb{R}$ with $\mu > \nu$,

$$\mu \mathbf{1}_N P \nu \mathbf{1}_N.$$

The second one is a property on aggregation. This axiom essentially requires that if a lot of people whose utility levels are equal are added, the betterness relation is sensitive to their utility levels.

Limit Property. For all $N \in \mathcal{N}$, for all $u_N, v_N \in \mathbb{R}^N$, for all $k \in \mathbb{R}$, and for all $\delta \in \mathbb{R}_{++}$, there exists a nonempty set $M^* \subseteq \mathbb{N} \setminus N$ such that, for all $M \subseteq \mathbb{N} \setminus N$ with $M \supsetneq M^*$,

$$(v_N, (k+\delta)\mathbf{1}_M) R (u_N, k\mathbf{1}_M) R (v_N, (k-\delta)\mathbf{1}_M).$$

An axiom similar to this axiom is originally introduced by Blackorby, Bossert, Donaldson, and Fleurbaey (1998); see also Spears and Budolfson (2021) for the use of its variant.

3 Theorem

We now establish our main theorem.

Theorem 1. *If a population principle R satisfies egalitarian dominance and the limit property, then the sadistic conclusion or the weak repugnant conclusion holds.*

Proof. Assume that R satisfies egalitarian dominance and the limit property. We show that the weak repugnant conclusion holds if the sadistic conclusion does not hold. Let $\alpha > 0$ and $u_1 = -\alpha$; notably, $u_1 = -\alpha \mathbf{1}_{\{1\}}$ holds. Let $\xi \in \mathbb{R}_{++}$, $\varepsilon \in (0, \xi)$, and $N \in \mathcal{N}$ such that $1 \notin N$. Because the sadistic conclusion does not hold, the completeness of R implies that

$$(\xi \mathbf{1}_{N' \cup \{1\}}, \xi \mathbf{1}_N) R (u_1, \xi \mathbf{1}_N)$$

for all $N' \in \mathcal{N}$ such that $N' \cap (N \cup \{1\}) = \emptyset$. Egalitarian dominance implies that

$$(\xi + 1)\mathbf{1}_{N' \cup \{1\} \cup N} P (\xi \mathbf{1}_{N' \cup \{1\}}, \xi \mathbf{1}_N)$$

for all $N' \in \mathcal{N}$ such that $N' \cap (N \cup \{1\}) = \emptyset$. By the limit property, there exists a nonempty set $M^* \subseteq \mathbb{N} \setminus (N \cup \{1\})$ such that, for all $M \subseteq \mathbb{N} \setminus (N \cup \{1\})$ with $M \supsetneq M^*$, it holds that

$$((\xi+2)\mathbf{1}_{M \cup N}, -\beta \mathbf{1}_{\{1\}}) R (\xi+1)\mathbf{1}_{M \cup \{1\} \cup N},$$

where $\beta > 0$. Now, we take $N^* \in \mathcal{N}$ such that

$$N^* \supsetneq M^* \text{ and } N^* \cap (N \cup \{1\}) = \emptyset.$$

By construction, we obtain

$$\begin{aligned}
&(\xi \mathbf{1}_{N^* \cup \{1\}}, \xi \mathbf{1}_N) R (u_1, \xi \mathbf{1}_N);\\
&(\xi+1)\mathbf{1}_{N^* \cup \{1\} \cup N} P (\xi \mathbf{1}_{N^* \cup \{1\}}, \xi \mathbf{1}_N);\\
&((\xi+2)\mathbf{1}_{N^* \cup N}, -\beta \mathbf{1}_{\{1\}}) R (\xi+1)\mathbf{1}_{N^* \cup \{1\} \cup N}.
\end{aligned}$$

Because the sadistic conclusion does not hold, we obtain

$$((\xi+2)\mathbf{1}_{N^* \cup N}, \frac{\varepsilon}{2}\mathbf{1}_{N'' \cup \{1\}}) R ((\xi+2)\mathbf{1}_{N^* \cup N}, -\beta \mathbf{1}_{\{1\}})$$

for all $N'' \in \mathcal{N}$ such that $N'' \cap (N^* \cup N \cup \{1\}) = \emptyset$; note that N^*, N, and $\{1\}$ are disjoint by assumption. By the limit property, there exists a nonempty set $K^* \subseteq \mathbb{N} \setminus (N^* \cup N \cup \{1\})$ such that, for all $K \subseteq \mathbb{N} \setminus (N^* \cup N \cup \{1\})$ with $K \supsetneq K^*$,

$$(-\alpha \mathbf{1}_{\{1\}}, \varepsilon \mathbf{1}_{K \cup N^* \cup N}) R ((\xi+2)\mathbf{1}_{N^* \cup N}, \frac{\varepsilon}{2}\mathbf{1}_{K \cup \{1\}}).$$

Let $L^* \in \mathcal{N}$ be such that

$$L^* \supsetneq K^* \text{ and } L^* \cap (N^* \cup N \cup \{1\}) = \emptyset.$$

By construction, we obtain

$$\begin{aligned}
&((\xi+2)\mathbf{1}_{N^* \cup N}, \frac{\varepsilon}{2}\mathbf{1}_{L^* \cup \{1\}}) R ((\xi+2)\mathbf{1}_{N^* \cup N}, -\beta \mathbf{1}_{\{1\}});\\
&(-\alpha \mathbf{1}_{\{1\}}, \varepsilon \mathbf{1}_{L^* \cup N^* \cup N}) R ((\xi+2)\mathbf{1}_{N^* \cup N}, \frac{\varepsilon}{2}\mathbf{1}_{L^* \cup \{1\}}).
\end{aligned}$$

Then, by transitivity, we obtain

$$(u_1, \varepsilon \mathbf{1}_{L^* \cup N^* \cup N}) P (u_1, \xi \mathbf{1}_N),$$

where we use $u_1 = -\alpha \mathbf{1}_{\{1\}}$. Thus, the weak repugnant conclusion holds. ■

We now show that there exists a population principle that satisfies egalitarian dominance and the limit property, and avoids both the sadistic and repugnant conclusions. Define $\hat{R}$ by letting

$$\begin{aligned}
&u_N \hat{P} v_M \text{ if } |N| = 1 \text{ and } |M| > 1;\\
&u_N \hat{R} v_M \Leftrightarrow \sum_{i \in N} u_i \geq \sum_{i \in M} u_i \text{ for all remaining population sizes } |N|, |M|.
\end{aligned}$$

Note that $\hat{R}$ is complete and transitive, and thus it is a well-defined population principle. It is easy to verify that egalitarian dominance and the limit property are satisfied, and that both the sadistic and repugnant conclusions are avoided. Thus, our result is not valid when the weak repugnant conclusion is replaced with the repugnant conclusion.

What is an additional axiom that yields the incompatibility between avoidance of the sadistic conclusion and avoidance of the original repugnant conclusion? We consider the following axiom, which is used by Blackorby, Bossert, and Donaldson (2005).

Existence Independence. For all $N, M, L \in \mathcal{N}$ such that $(N \cup M) \cap L = \emptyset$, for all $u_N \in \mathbb{R}^N$, for all $v_M \in \mathbb{R}^M$, and for all $w_L \in \mathbb{R}^L$,

$$u_N R v_M \iff (u_N, w_L) R (v_M, w_L).$$

In the presence of existence independence, the weak repugnant conclusion is replaced with the repugnant conclusion; the proof is very similar to that of Theorem 1.

Corollary 1. *If a population principle R satisfies egalitarian dominance, existence independence, and the limit property, then the sadistic conclusion or the repugnant conclusion holds.*

This result clarifies the difference between Theorem 1 and a recent theorem by Bossert, Cato, and Kamaga (2022a), which focuses on a certain class of population principles. They show that if a population principle in the class satisfies the limit property, then the sadistic conclusion or the repugnant conclusion holds. Notably, the principles examined by Bossert, Cato, and Kamaga (2022a) do not necessarily satisfy existence independence (egalitarian dominance is satisfied). In Corollary 1, the class of principles is not specified, but existence independence is imposed.

References

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

[2] Arrhenius, G. (2009). One more axiological impossibility theorem, in L. Johansson, J. Österberg, and R. Sliwinski (eds.), Logic, Ethics, and All That Jazz: Essays in Honour of Jordan Howard Sobel. Uppsala Philosophical Studies 57, 23–37. (Uppsala: Uppsala University)

[3] Asheim, G. B., and Zuber, S. (2014). Escaping the repugnant conclusion: Rank-discounted utilitarianism with variable population. Theoretical Economics, 9(3), 629–650.

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

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

[6] Bossert, W. (1990). Maximin welfare orderings with variable population size. Social Choice and Welfare, 7, 39–45.

[7] Bossert, W., Cato, S., and Kamaga, K. (2022a). Revisiting variable-value population principles. Economics & Philosophy, 1–17. doi:10.1017/S0266267122000268

[8] Bossert, W., Cato, S. and Kamaga, K. (2022b). Thresholds, critical levels, and generalized sufficientarian principles. Economic Theory https://doi.org/10.1007/s00199-022-01439-z

[9] Cato, S. (2023). When is weak Pareto equivalent to strong Pareto? Economics Letters, 222, article no. 110953.

[10] Cato, S., and Ishida, S. (2022). Multi-species population ethics with critical levels. Discussion Paper Series of Institute-Wide Joint Research Project "Methodology of Social Sciences." E-22-004.

[11] Franz, N., and Spears, D. (2020). Mere Addition is equivalent to avoiding the Sadistic Conclusion in all plausible variable-population social orderings. Economics letters, 196, article no. 109547.

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

[13] Pivato, M. (2020). Rank-additive population ethics. Economic Theory, 69(4), 861–918.

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

[15] Zuber, S. (2018). Population-adjusted egalitarianism. CES Working Paper, halshs-01937766.

[16] Zuber, S., Venkatesh, N., Tännsjö, T., Tarsney, C., Stefánsson, H. O., Steele, K., ... and Asheim, G. B. (2021). What should we agree on about the repugnant conclusion?. Utilitas, 33(4), 379–383.