

[Forthcoming in *Mind*.]

Review of Iwao Hirose, *Moral Aggregation*

Johan E. Gustafsson

Moral Aggregation, by Iwao Hirose. Oxford and New York: Oxford University Press, 2015. Pp. xiii + 234. H/b £37.99, \$55.00.

In a choice between saving five people or saving another person, is it better to save the five, other things being equal? According to utilitarianism, it would be better to save the five if the combined gain in well-being for them would be greater than the loss for the one. A standard objection is that adding up the gains or losses of different people in this manner is a problematic form of interpersonal aggregation. It is far from clear, however, what more precisely is supposed to be problematic about utilitarian aggregation. The aggregation critics—that is, among others, John Rawls, Robert Nozick, Thomas Nagel, John M. Taurek, and T. M. Scanlon—have not offered a clear criterion for what counts as a morally problematic form of aggregation and what does not. Hence it is hard to know what to make of this objection.

In *Moral Aggregation*, Iwao Hirose makes an admirable attempt to provide a precise account of the kind of aggregation that these critics have found problematic in utilitarianism and similar theories. Hirose's interpretation is mainly based on the following remark by Taurek:

It is not my way to think of them [the five] as each having a certain *objective value*, determined however it is we determine the objective value of things, and then to make some estimate of the *combined value* of the five as against the one. (Taurek 1977, p. 307 as quoted by Hirose, p. 22)

Let *interpersonal aggregation* be the kind of aggregation that aggregation critics such as Taurek have found morally problematic. According to Hirose, interpersonal aggregation is 'the combination of different people's morally relevant factors (i.e. good, well-being, happiness, pleasure, desire-satisfaction, claims, reasons, and so on) into a real value that represents the relation of sets of morally relevant factors.' (p. 24) More precisely, Hirose proposes that interpersonal aggregation is equivalent to the conjunction of the following four conditions:

Interpersonal comparability: For every pair of persons, it is possible to compare the morally relevant factor of one person with the morally relevant factor of the other. (p. 32)

Impartiality: Other things being constant, two alternatives are morally indifferent if they differ only with regard to the identities of people. (p. 36)

Pareto: If one alternative is *F*-er for some person than another alternative, and if it is at least as *F* for the other persons, then it is *F*-er than the other. (p. 38)

Continuity: For any set of morally relevant factors a_i , define $A(a_i) = \{a_j \mid a_j \text{ is at least as } F \text{ as } a_i\}$ and $B(a_i) = \{a_j \mid a_i \text{ is at least as } F \text{ as } a_j\}$. Then, $A(a_i)$ and $B(a_i)$ are closed (in other words, $A(a_i)$ and $B(a_i)$ contain their own boundaries, for any a_i in the set of morally relevant factors). (p. 40)

Based on this interpretation of the aggregation critics, Hirose goes on to discuss whether interpersonal aggregation defined in this manner should be rejected in ethics. His answer in the end is a qualified no. The qualification is that he only defends what he calls *formal aggregation* where the morally relevant factors for individuals are undetermined outside of the aggregation process. Hirose contrasts formal aggregation with *substantive aggregation* where the morally relevant factors for individuals are determined outside of the aggregation process, as in hedonistic utilitarianism where the well-being of an individual only depends on the individual's hedonic states. In the following, I shall assume that individual well-being is determined in this substantive manner.

Hirose's support for his interpretation of the aggregation critics is mainly that these four conditions are satisfied by utilitarianism, which aggregation critics typically find problematic, and Maximin and Leximin, which aggregation critics typically find unproblematic, violate one of the conditions. Maximin and Leximin are defined as follows, where a state of affairs is represented by a vector of the well-being level w_i for each individual i that exists in the state:

Maximin: A state of affairs $x = (w_1, w_2, \dots, w_n)$ is at least as good as another state of affairs $y = (w'_1, w'_2, \dots, w'_n)$ if and only if $\min(w_1, w_2, \dots, w_n) \geq \min(w'_1, w'_2, \dots, w'_n)$. (p. 28)

Leximin: For all $x = (w_1, w_2, \dots, w_n)$ and $y = (w'_1, w'_2, \dots, w'_n)$, if there exists a position k in $N = \{1, 2, 3, \dots, n\}$ such that: (1) the well-being level in k is strictly higher in x than y ; and (2) the well-being level of every position $j < k$ is the same in x as in y , then x is strictly better than y . Otherwise, x and y are equally good. (p. 29)

Maximin does not involve interpersonal aggregation on Hirose's account, because it violates Pareto. It violates Pareto, because an improvement for someone who is not among the worst off in a state of affairs does not make the state better overall. Leximin, on the other hand, satisfies Pareto, but it still does not involve interpersonal aggregation on Hirose's account, because, instead of Pareto, it violates Continuity. To see that Leximin violates Continuity, consider the two-person distribution $(0, 2)$ and the series of two-person distributions $(1/n, 1)$ for $n = 1, 2, 3, \dots$. According to Leximin, each distribution in this series is better than $(0, 2)$; yet the series converges to $(0, 1)$, which is worse than $(0, 2)$.

My worry about Hirose's account of interpersonal aggregation is that it fails to capture what the aggregation critics typically find morally problematic about utilitarian aggregation. For each of the last three conditions, it seems that there is a variant of utilitarianism which violates the condition yet relies on a form of aggregation that would be objectionable to aggregation critics in the same way as the standard utilitarian aggregation. For our discussion, we shall, following Hirose, take '*F*-er' in these conditions to be 'better' and 'at least as *F*' to be 'at least as good'.

To see that Impartiality is unnecessary for interpersonal aggregation, consider a partial variant of utilitarianism where, as on some religious views, there is a person who is special because of his or her identity; and the well-being of this special person has more weight in the aggregation than the well-being of others. This theory violates Impartiality, yet it favours saving five ordinary people over saving another ordinary person if the combined benefit for the five outweighs the harm for the one. This partial variant seems to involve a form of aggregation which would be objectionable to aggregation critics in the same way as standard utilitarian aggregation. My point is not that this partial variant of utilitarianism is particularly plausible; rather, my point is that one of its controversial properties is that it weighs the combined claims different people against the claims of others. Note that there is no mention of partiality in the quote from Taurek. He seems to be worried about weighing the combined claims of the five against the claim of the one. He would, I presume, also be opposed to weighing combinations of claims where the claims of some individuals are given slightly more weight than those of others.

To see that Pareto is unnecessary for interpersonal aggregation, consider instead a desert-sensitive variant of utilitarianism where no weight is given to the well-being of sinners. This desert-sensitive variant of utilitarianism violates Pareto, because things do not get better overall if the well-being of a sinner is improved, other things being equal. Hence this variant avoids interpersonal aggregation in the same way as Maximin on Hirose's account. But it handles the comparison between saving five

saints and just saving another saint in the same way as standard utilitarianism. Hence it should be objectionable in the same way to aggregation critics. The sole argument Hirose offers for Pareto being necessary for interpersonal aggregation is the remark that,

other things being equal, the increase of one factor should make the value of the combined factors strictly greater. If the increase of one factor does not make any difference, then this is perfectly consistent with Taurek's position and therefore does not elucidate what Taurek is opposing. (p. 25)

But, as we have seen, the desert-sensitive variant of utilitarianism violates Pareto, yet it is clearly incompatible with Taurek's position.

Finally, to see that Continuity is unnecessary for interpersonal aggregation, consider a combination of utilitarianism and Leximin according to which a state of affairs x is at least as good as a state of affairs y if and only if

either the sum total of well-being is greater in x than in y
or the sum total of well-being is the same in x as in y and Leximin ranks x at least as high as y .

This variant is lexical just like Leximin. Hence, just like Leximin, it rules out Continuity. To see this, consider the three-person distribution $(0, 2, 2)$ and the series of three-person distributions $(1/n, 1, 3 - 1/n)$ for $n = 1, 2, 3, \dots$. According to the lexical variant of utilitarianism, each distribution in this series is better than $(0, 2, 2)$; yet the series converges to $(0, 1, 3)$, which is worse than $(0, 2, 2)$.

Since this lexical variant of utilitarianism violates Continuity, it avoids interpersonal aggregation on Hirose's account. But, if standard utilitarianism involves interpersonal aggregation, then surely this lexical variant does so too; it aggregates just like standard utilitarianism except that, if there is a tie, it takes distribution into account as a tiebreaker. Hirose admits that

It is not entirely clear what [Taurek] means by objective value. I take it to be the *numerical* value that is comparable between two individuals or two groups of individuals. (p. 22)

The reason Hirose takes Continuity to be necessary for interpersonal aggregation is that it is required for representing the aggregate value by a continuous real-valued function. Yet it seems that Taurek would also be opposed to weighing the combined value of a group of individuals against the combined value of another group even if the ordering of these objective values were not representable by a continuous real-valued function, as in the above lexical variant of utilitarianism.

Thus the last three conditions all seem unnecessary for interpersonal aggregation. Hirose's account, it seems, fails to capture the kind of aggregation that is supposed to be morally problematic according to Taurek and other aggregation critics. Hence it is still hard to know what to make of the surprisingly influential objection that utilitarianism involves a morally problematic form of aggregation.

In this review, I have focused on the first of the book's two parts. The second part contains an extensive critical overview of the literature on the number problem, which is very helpful. It should be of interest to anyone working on the number problem and, in particular, F. M. Kamm's argument for best outcomes.

I wish to thank Gustaf Arrhenius, Richard Yetter Chappell, Iwao Hirose, and Martin Peterson for valuable comments.

References

Taurek, John M. (1977) 'Should the Numbers Count?', *Philosophy & Public Affairs* 6 (4): 293–316.