Erschienen in: Economics and Philosophy. Cambridge University Press. 2016, 32(1), pp. 149-156
https://doi.org/10.1017/S0266267115000437

Moral Aggregation, Iwao Hirose. Oxford University Press, 2015, xiii + 234 pages.

In *Moral Aggregation*, Iwao Hirose defends a weak form of interpersonal aggregation that he refers to as *formal aggregation*. The book is divided into two parts. In the first, Hirose lays out what he means by formal aggregation. In the second, he demonstrates the flexibility of formal aggregation by using a formally aggregative approach to solve the so-called

* Department of Philosophy, University of Sheffield, 45 Victoria Street, Sheffield S6 Sheffield S3 7QB. Email: j.lenman@sheffield.ac.uk. URL: http://www.jimmylenman.com/.

Number Problem, i.e. the problem of what a moral agent should do when she can save at most one of two differently sized groups of people from substantial harm. I will argue that Hirose's solution to the Number Problem is interesting and plausible, but that his defence of formal aggregation is unconvincing.

1. FORMAL AGGREGATION AS A WEAK CONSTRAINT

A moral theory is *formally aggregative* in Hirose's sense if and only if it allows us to express the overall goodness of any state of affairs in terms of a strictly increasing, real-valued function that takes as its arguments the 'personal goodness values' of all the individuals who figure in the state of affairs in question.

The formal aggregation that Hirose defends is weak for two reasons. First, Hirose is concerned exclusively with how a moral theory assesses and compares states of affairs. As long as it does this in a way that is consistent with formal aggregation, it is formally aggregative in Hirose's sense, no matter how loose the connection between how the theory assesses states of affairs and how it implies we should deliberate about the morally right thing to do. With regard to deontological constraints, Hirose (24–5) puts the point as follows: '...aggregation may well be constrained by some deontic constraints. For example, even if a state of affairs is judged to be strictly better than another, it may well be wrong to bring it about because bringing it about violates some people's basic rights. In such a case, aggregation takes place, but the rightness or wrongness of an act is not derived directly from this aggregation.' Formal aggregation also leaves room for agent-centred prerogatives: Hirose (113–4) concedes that we may sometimes look after our own self-interest instead of being morally required to try and bring about the best possible state of affairs.

Second, formal aggregation is not concerned with the overt procedure that a moral theory employs to assess the goodness of states of affairs. Hirose does not express himself very clearly here, but he aims to distinguish formal aggregation from a more restrictive type of aggregation that he refers to as *substantive*. As I understand his distinction, a substantively aggregative moral theory (i) locates what is morally relevant about states of affairs in factors that are indisputably tied to individuals,[1] and (ii) assesses the overall goodness of a state of affairs by combining these factors into an overall goodness value. Classical utilitarianism is a prime example of a substantively aggregative moral theory (52). By contrast, a theory can be formally aggregative without making overt use of an aggregative method in its assessment of states of affairs. For a theory to count as formally aggregative, it merely has to be possible to represent

[1] Well-being or desire satisfaction are examples for such factors.

whatever method it employs in terms of a mapping from a vector of 'personal goodness values' to a scalar overall goodness value. To give an example, it is a commonly held view that inequality is impersonally bad, so that where inequality is present, it makes states of affairs worse without necessarily being bad for anyone. While this view is incompatible with substantive aggregation, it need not thereby be non-aggregative. As long as it is formally possible to assign the disvalue of inequality to specific individuals – e.g. to those who get less than their equal share of the good in question – the view is formally aggregative in Hirose's sense.

2. FORMAL AGGREGATION'S BITE

Because formal aggregation is weak in the ways just described, many theories which would not usually be characterized as aggregative still count as formally aggregative in Hirose's sense. In fact, formal aggregation may appear as too weak a constraint on moral theories to have any bite at all. In Chapter 3, Hirose discusses Thomas Scanlon's World Cup Case (WCC) as an intuitively compelling counterexample to formal aggregation. Hirose contends that several hidden assumptions of the WCC bias us in favour of Scanlon's non-aggregative conclusion. He moreover remarks that *any* constraint on moral theories will have counterintuitive implications in certain extreme cases, so that we should not reject constraints based on scattered counterexamples. What Hirose fails to notice is that Scanlon's WCC is actually fully consistent with formal aggregation. Here is how Scanlon describes his case:

> **The World Cup Case.** 'Suppose that Jones has suffered an accident in the transmitter room of a television station. Electrical equipment has fallen on his arm, and we cannot rescue him without turning off the transmitter for fifteen minutes. A World Cup match is in progress, watched by many people, and it will not be over for an hour. Jones's injury will not get any worse if we wait, but his hand has been mashed and he is receiving extremely painful electrical shocks. Should we rescue him now or wait until the match is over? Does the right thing to do depend on how many people are watching – whether it is one million or five million or a hundred million? It seems to me that we should not wait [. . .]' (Scanlon, quoted in Hirose, 42–3).

Suppose we agree with Scanlon that it would be wrong to wait. Further suppose that we derive this judgement from the belief that waiting would lead to morally worse consequences.[2] A simple way to capture this belief is by distinguishing trivial from substantial pleasures and pains in such a

[2] To think of the WCC as a counterexample to formal aggregation, Hirose must assume that in the WCC, we judge it wrong not to bring about the best possible consequences.

	Jones' Personal Good (PG)	Each Viewer's PG
(1) Save Jones Immediately	−10	$-(1/n)$
(2) Wait	−20	$1/n$

TABLE 1. The WCC Decision Situation

way that no amount of the former can outweigh any amount of the latter.[3] This will for example be the case if:

- A state of affairs' overall goodness is defined as the simple sum of the personal goods of all individuals concerned;
- The pleasures that an individual experiences add to her personal good, while the pains that she experiences detract from it;
- Substantial pleasures and pains contribute to an individual's personal good with an absolute value that is always strictly greater than 2;
- Trivial pleasures and pains contribute to an individual's personal good with an absolute value that never exceeds $1/n$, where n is the number of individuals whose personal good contributes to the overall goodness of the states of affairs under comparison.

Now suppose that for the World Cup viewers, only trivial pleasures and pains are at stake, whereas for Jones, different amounts of substantial pain are at stake. We can then represent the WCC decision situation as follows (see Table 1):

From Table 1, it can easily be seen that *no matter how many TV viewers there are*, the option of saving Jones without delay can never have consequences worse than is expressed by the badness value of −11, whereas the option of waiting can never have consequences better than −19. The judgement that it is always better to save Jones right away thus poses no challenge to formal aggregation.

Some might want to argue that it is inadmissibly ad hoc to introduce a factor such as n into the calculation of an individual's personal good. But it is at least arguable whether the introduction of n would constitute an ad hoc move. More importantly, Hirose (105) makes it clear that he is not opposed to ad hoc fixes that serve to accommodate what would otherwise be non-aggregative ideas into a formally aggregative framework.

Given that even our 'non-aggregative' intuitions about the WCC turn out to be compatible with formal aggregation, we might start to wonder whether *any* approach to morally assessing states of affairs is clearly

[3] While this view has problems of its own, this need not concern me here, as I am merely suggesting that our intuitions about the WCC are consistent with formal aggregation.

ruled out by formal aggregation. Hirose (29–31, 74) thinks that Leximin qualifies. Leximin is a rule for ordering states of affairs that expresses a special concern for the worst off. With Leximin, improvements for the worst off always have a categorically more significant impact on a state of affairs' overall goodness than do improvements for those who are already better off. Leximin ranks states of affairs in a manner that violates the technical condition of *continuity*.[4] If an ordering violates continuity, the existence of a function that represents it cannot be guaranteed. It follows that Leximin is not always consistent with formal aggregation.[5]

But there are in fact numerous rules for the ordering of states of affairs that are *clearly* inconsistent with formal aggregation, namely those that violate *completeness*. Such rules regrettably find no mention in Hirose's book. If a rule violates completeness, it implies that there are some states of affairs S and S' such that it is neither true that S is morally better than S', nor that S' is morally better than S, nor that they are equally good. To endorse an incomplete rule is to affirm that it is not always possible to meaningfully compare different states of affairs, e.g. because the values that they instantiate are too disparate.

Continuity and completeness are without a doubt properties that it will often be *useful* for an ordering to have. They help ensure that we can represent the ordering in functional form, so that we can describe and explore it with the potent tools that mathematics has to offer. But this does not give us sufficient reason to reject all those moral theories which work with discontinuous or incomplete orderings. To give just one example, a moral theory according to which we only occasionally have a duty to try and bring about the best possible consequences may cope fine with an incomplete ordering. Moreover, even where continuity and completeness would clearly be useful, this does not provide us with a reason to think that the correct or morally most appropriate ordering will therefore satisfy them.

While Hirose (15) proclaims that the aim of his book is to 'defend' formal aggregation, he never discusses what reason we might have to reject the moral theories that violate it.[6] Much to my surprise, he does not even lament the formal intractability that I have just suggested may sometimes be a problem for these theories. Yet if this is the case, then

[4] A complete and transitive order $\succsim$ on some set X is continuous iff for every $y \in X$, the sets $\{x \in X: x \succsim y\}$ and $\{x \in X: x \precsim y\}$ are closed.

[5] Hirose (39–40) makes the stronger claim that discontinuous orderings can never be represented by real-valued functions and are thus by definition inconsistent with formal aggregation. This is incorrect. If—as Hirose (23, fn. 3) assumes – only countably many states of affairs need to be ranked, it is possible to represent *any* complete and transitive order with a real-valued function.

[6] In light of the fact that Hirose defends formal aggregation as a *constraint* on moral theories, I take it that he wishes to reject the theories that violate it.

what does Hirose's defence of formal aggregation amount to? A major problem with Hirose's book is that this remains unclear throughout. As things stand, I take it that Hirose pursues the following strategy: he tries to demonstrate that formal aggregation is flexible enough to accommodate most of our firmly held judgements about particular cases, so that we have little reason to be opposed to it. I believe that Hirose wishes to illustrate the flexibility of formal aggregation in the second part of his book, where he discusses the Number Problem.

3. THE NUMBER PROBLEM

To solve the Number Problem is to explain and justify how a moral agent should act when she can save at most one out of two differently sized groups of people from substantial harm. Consider the following cases, and what Hirose takes to be our intuitive judgements about them:

1. **John Taurek's Original Case.** In this case, you can save either one or five individuals from certain death. You cannot save all six. Hirose thinks that in Taurek's Original Case, you ought to save the five.
2. **Frances Kamm's Sore Throat Case.** In this case, you can either save one individual from certain death, or you can save another individual from certain death *and* cure a third individual's sore throat. In this case, Hirose thinks it plausible that you ought to flip a coin to decide between your two alternatives.
3. **Kamm's Large-Scale Rescue Case.** In this case, you can save either 1000 or a different 1001 individuals from certain death. In this case, Hirose again thinks it plausible that you ought to flip a coin.

In an interesting discussion, Hirose looks at some of the solutions to the Number Problem that have been put forward in the literature. Especially informative is his treatment of Thomas Schelling's otherwise little discussed 'probabilistic argument' (168) in favour of saving the greater number. After invariably rejecting the solutions he considers, Hirose moves on to propose a solution of his own. Unlike the existing proposals he considers, his novel proposal is clearly consistent with formal aggregation.

Hirose assumes that a moral agent in a Number Problem-type case ought to bring about the best possible state of affairs. He moreover assumes that the goodness of any state of affairs is a strictly increasing function of the personal good of the individuals who figure in the state of affairs in question. Lastly, he assumes that an individual's personal good is negatively affected not only when she suffers a setback to her well-being, but also when she *suffers unfairness*. Hirose then draws upon ideas defended by John Broome to argue that in cases where a moral agent can

save only one out of two groups from substantial harm, an individual threatened with substantial harm suffers unfairness if the moral agent assigns less of a chance to helping her than she assigns to helping others. This setup allows us to see Number Problem-type cases as involving trade-offs between the moral goods of fairness and welfare. In Taurek's Original Case, if you behave unfairly towards the one, you can save five people for sure as opposed to an expected value of only three. In Kamm's Sore Throat Case, the expected payoff of behaving unfairly is much smaller: it amounts to only 50% of the moral value of curing a sore throat. Finally, in Kamm's Large-Scale Rescue Case, the expected payoff of behaving unfairly amounts to 50% of the moral value of saving a human life. This is substantial. But compared to the other two cases, the unfairness involved here is multiplied by a factor of 1000. These brief remarks hopefully make it sufficiently clear that for a large range of functional specifications of personal and overall goodness values, Hirose's setup yields conclusions that are in line with, and can thus explain, our considered judgements.

Hirose's solution to the Number Problem strikes me as superior to the existing solutions he discusses. I find it an intuitively compelling idea that in cases where we can save only one of two differently sized groups from substantial harm, we are facing a trade-off between treating equally pressing claims equally, and effectively promoting well-being. Hirose moreover shows that this trade-off between fairness and well-being can be captured in a way that is fully consistent with formal aggregation. He thereby clarifies that even considerations of procedural fairness – which we would not usually think of as aggregative – can be incorporated into the kind of maximizing framework that formal aggregation supplies.

But beyond a demonstration of its surprising flexibility, Hirose's defence of formal aggregation does not amount to much. Hirose repeatedly asserts that formal aggregation 'helps us to understand the structure of ethical arguments' (221) because it 'clarifies the structural relationship between all-things-considered judgements and various morally relevant factors' (220). But I do not think that formal aggregation helps us *understand* the structure of ethical arguments; as I see it, it rather *prescribes* a structure that is sufficiently weak for many different ethical arguments to be squeezed into it. Take the argument that when you can save only one out of two groups from certain death, there is value to giving the same chance of survival to both groups, as this expresses an appropriately equal concern for all those whose lives are at stake. Though this idea can be made consistent with formal aggregation, we do not in the process of making it consistent discover its basic structure. Quite to the contrary: only after we have determined what is essential to it are we able to investigate how to best incorporate it into a formally aggregative framework. In light of his welcoming stance towards ad hoc moves in

the context of this process, it is surprising that Hirose does not see things this way. While ad hoc moves are admissible when for pragmatic reasons, we seek to represent an independently plausible ethical argument in functional form, they are quite at odds with the enterprise of uncovering the argument's basic structure.

In sum, what makes this book well worth reading is its treatment of the Number Problem. Hirose's discussion of existing solution proposals is concise and informative, and his own proposal, which is plausible in its simplicity, deserves serious consideration. It is a good thing that the second part of the book can be read largely independently of the first.

Susanne Burri*

ACKNOWLEDGEMENTS

I thank Gustaf Arrhenius and Alex Voorhoeve for helpful comments.

BIOGRAPHICAL INFORMATION

Susanne Burri is an Assistant Professor in the Philosophy Department at the London School of Economics and Political Science. At the time of writing this review, she was a postdoctoral researcher in philosophy at the University of St. Gallen in Switzerland. Her research interests include normative ethics and philosophy of economics.

* Department of Philosophy, Logic and Scientific Method, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK. email: s.burri@lse.ac.uk. URL: www.susanneburri.ch.