Fairness and Aggregation

A. C. PASEAU
Wadham College, Oxford

BEN SAUNDERS
University of Southampton

Sometimes, two unfair distributions cancel out in aggregate. Paradoxically, two distributions each of which is fair in isolation may give rise to aggregate unfairness. When assessing the fairness of distributions, it therefore matters whether we assess transactions piecemeal or focus only on the overall result. This piece illustrates these difficulties for two leading theories of fairness (proportionality and shortfall minimization) before offering a formal proof that no non-trivial theory guarantees aggregativity. This is not intended as a criticism of any particular theory, but as a datum that must be taken into account in constructing a theory of fairness.

Suppose some people have claims on a divisible good that cannot all be satisfied. A natural question is what a *fair* distribution of the good among the claimants should be. Different theories of fairness offer different answers. Our aim is not to offer our own answer to this question, but to illustrate a problem that affects virtually all theories. Two transactions, each of which is fair in isolation, may produce an aggregate result which would be judged as unfair had it resulted from a single distribution.

For simplicity, we take money as our example of a divisible good (the case of other divisible goods is parallel). Take the schematic case in which Debtor, D, owes money to two Creditors, C_1 and C_2. He has no other obligations but, come the time to repay these debts, he does not have enough to repay C_1 and C_2 fully. Suppose that he owes d_1 to C_1 and d_2 to C_2, but he has only m, where $m < d_1 + d_2$. (Monetary amounts are non-negative quantities throughout.) How should he divide m between C_1 and C_2?

Various theories of fairness offer solutions to this question. In this article, we show that all non-trivial theories of fairness suffer from a problem of aggregation; that is, any non-trivial theory of fairness will, paradoxically, sometimes return different answers in the following two cases:

© Cambridge University Press 2015
Utilitas Vol. 27, No. 4, December 2015
doi:10.1017/S0953820815000230

CrossMark
https://doi.org/10.1017/S0953820815000230 Published online by Cambridge University Press

Case 1: D owes d_1 to C_1 and d_2 to C_2 but has only m (where $m < d_1 + d_2$). A second Debtor, D^*, owes d_1^* to C_1 and d_2^* to C_2 but has only m^* (where $m^* < d_1^* + d_2^*$).

Case 2: D owes $d_1 + d_1^*$ to C_1 and $d_2 + d_2^*$ to C_2 but has only $m + m^*$ (where $m + m^* < d_1 + d_1^* + d_2 + d_2^*$).

Intuitively, one might think that the amount repaid to each of C_1 and C_2 should be the same in either case, since each is owed the same amount ($d_1 + d_1^*$ and $d_2 + d_2^*$, respectively) and there is as much money available to repay them ($m + m^*$). But this expectation cannot be satisfied in general: a simple mathematical argument demonstrates that no acceptable theory of fairness is aggregative, in a sense that will be clarified shortly.

Some writers seem to have a passing awareness of the aggregation phenomenon,[1] while others have discussed mathematical aspects of fairness,[2] but no philosopher, so far as we know, has explicitly discussed the simple result presented here. We do not mean that our result is wholly original, since there is an extensive and sophisticated economics literature in this area, which appears to imply our result.[3] This article's originality lies rather in its isolation of the result in a way that we believe has not previously been done, in its particularly accessible proof and, most importantly, in its relating the result to philosophical theories of fairness. Certainly, the result's significance has not hitherto been appreciated by philosophers working on fairness. Before giving the general argument, we illustrate it for two prominent theories of fairness.

FAIRNESS AS PROPORTIONALITY

Though we consider alternative accounts below, we begin with the idea that fairness demands proportionality. Many writers draw this connection,[4] but its best-known development is

[1] E.g. Geoffrey Cupit, '*Fairness*, by Nicholas Rescher [book review]', *Mind* 113 (2002), pp. 387–90, at 389.

[2] E.g. Steven J. Brams and Alan D. Taylor, *Fair Division: From Cake-cutting to Dispute Resolution* (Cambridge, 1996).

[3] A seminal paper on the bankruptcy (or conflicting claim) problem is R. Aumann and M. Maschler, 'Game Theoretic Analysis of a Bankruptcy Problem from the Talmud', *Journal of Economic Theory* 36 (1985), pp. 195–213. For a survey of the ensuing literature, see W. Thomson, 'Axiomatic and Game-theoretic Analysis of Bankruptcy and Taxation Problems: A Survey', *Mathematical Social Sciences* 45 (2003), pp. 249–97.

[4] E.g. Mark Reiff, 'Proportionality, Winner-take-all and Distributive Justice', *Politics, Philosophy & Economics* 8 (2009), pp. 5–42, at 11; Alexander Brown, 'Principles of Stakes Fairness in Sport', *Politics, Philosophy & Economics* 14 (2015), pp. 152–86.

https://doi.org/10.1017/S0953820815000230 Published online by Cambridge University Press

in the work of John Broome,[5] which has been much discussed elsewhere.[6]

According to Broome, fairness consists in the proportionate satisfaction of claims. Note that Broome distinguishes fairness from what we should do all things considered. Since we focus on the former, claims about 'what should be done' should be read with the implicit qualifier 'as a matter of fairness'. Here, C_1 has a claim to d_1 and C_2 has a claim to d_2. Proportionality implies that, if these claims are of equal strength, C_1 will receive $m.d_1/(d_1 + d_2)$ and C_2 will receive $m.d_2/(d_1 + d_2)$; for instance they may each receive half of what they are owed. If C_2's claim is twice the strength of C_1's, then Broome holds that C_2 should receive twice the satisfaction (proportionately), e.g. if C_1 has a claim to £20 of strength 5 and C_2 has a claim to £40 of strength 10, then a fair distribution of £20 would be to give £4 to C_2 (20% satisfaction) and £16 (40% satisfaction) to C_2. We focus on cases where claims are of equal strength for simplicity since showing that no theory is aggregative when dealing with claims of equal strength shows that no theory is aggregative in general.

That Broome's theory is non-aggregative can be demonstrated by example. Suppose $d_1 = 80$, $d_2 = 40$ and $m = 60$. Here D can only repay half his debts, so (according to Broome) the fair thing for him to do, if the claims are equal, is to pay 40 to C_1 and 20 to C_2, giving each creditor half of what she was owed. Now introduce a second debtor, D^*. Suppose $d_1{}^* = 40$, $d_2{}^* = 80$ and $m^* = 90$. Broome's theory would have D^* pay 30 to C_1 and 60 to C_2, giving each three-quarters of what they are owed. The aggregate result of these two transactions is that C_1 was owed 120 $(d_1 + d_1{}^*)$ and received 70, while C_2 who was also owed 120 $(d_2 + d_2{}^*)$ received 80. Thus, C_1 receives greater satisfaction than C_2, though *ex hypothesi* both were owed the same amount and had equal claims.

This inequality between C_1 and C_2 seems unfair. Consider the case where $D = D^*$. Here this single individual owes a total of 120 $(d_1 + d_1{}^*)$ to C_1 and 120 $(d_2 + d_2{}^*)$ to C_2 and has 150 $(m + m^*)$ with which to repay these debts. In this case, the fair thing to do is to repay 75 to each of C_1 and C_2, reflecting their equal claims to equal amounts. The surprise is that we get a different solution where $D \neq D^*$. In other words, Broome's

[5] John Broome, 'Fairness', *Proceedings of the Aristotelian Society* 91 (1990), pp. 87–101.

[6] E.g. Brad Hooker, 'Fairness', *Ethical Theory and Moral Practice* 8 (2005), pp. 329–52; Ben Saunders, 'Fairness between Competing Claims', *Res Publica* 16 (2010), pp. 41–55; Patrick Tomlin, 'On Fairness and Claims', *Utilitas* 24 (2012), pp. 200–13; Hugh Lazenby, 'Broome on Fairness and Lotteries', *Utilitas* 26 (2014), pp. 331–45; and James R. Kirkpatrick and Nick Eastwood, 'Broome's Theory of Fairness and the Problem of Quantifying the Strengths of Claims', *Utilitas* 27 (2015), pp. 82–91.

https://doi.org/10.1017/S0953820815000230 Published online by Cambridge University Press

theory is non-aggregative. It focuses on the distribution of particular goods on particular occasions, what we might call a 'narrow' view, but – as our example shows – the outcome of several fair transactions may be unfair, and vice versa, when we take a wider, overall view.

A response might be that the nature of the claims determines how they should be met. If C_1 is owed money by D and D^* separately, then they should repay separately, whereas if the money is owed by D and D^* together then they should repay collectively. But this response appears unsatisfactory.

First, it is intuitively unfair if C_1 receives greater satisfaction than C_2 though they have equally strong claims to the same amount (albeit against different debtors), so an adequate theory of fairness should be able to account for this. Any theory that tells us this situation is fair is, at best, an incomplete theory of fairness as it fails adequately to account for fairness in such cases.

Second, Broome defined claims as duties *owed to* particular individuals,[7] but he did not specify that they must be *owed by* particular individuals. Thus, if C_1 has a claim to £80 and a claim to £40, then C_1 does have claim to £120, even if no single agent owes this amount to her. We do not mean that C_1 has a *single* claim to this amount, but only that she has claim to this amount. In our debtor/creditor example, it is natural to separate C_1's claims into claims against D and D^*. However, this strategy would not necessarily work for all claims, such as claims based on need,[8] which are not necessarily claims against any particular individual. (One could suggest that claims of need are directed against specific individuals, for instance whoever is best-placed to meet that need. But this appears to run into problems if that individual does not respond to the claim.) In these cases, a fair response to an individual's claims depends on whether they are treated piecemeal or collectively.

Third, nothing in our schematic example specifies whether $D = D^*$ or $D \neq D^*$. It is possible that D and D^* are parts of some larger organization, such as the state, or even the same person, perhaps with two separate debts incurred in different currencies or at different times. This person could act fairly in settling one debt at t_1 and fairly in settling another debt at t_2 but, once again, we would probably want to know whether his actions were fair in aggregate.

Our example suggests that two acts that are both fair in isolation may nonetheless result in unfairness when aggregated. As we shall show, this feature is not unique to Broome's theory of fairness. We will

[7] Broome, 'Fairness', p. 92.
[8] Broome, 'Fairness', p. 93.

https://doi.org/10.1017/S0953820815000230 Published online by Cambridge University Press

not repeat the points in this paragraph and the previous three, but they should be understood as applying to all our subsequent discussion.

FAIRNESS AS SHORTFALL MINIMIZATION

An alternative theory of fairness states that D should divide m between C_1 and C_2 so as to minimize the largest shortfall suffered by either creditor.[9] This has been termed 'shortfall minimization', though it might more accurately be called 'minimax shortfall'. If D owes 900 to C_1 and 600 to C_2 but has only 500 then, on this theory, he should pay 400 to C_1 and 100 to C_2, leaving both creditors with a shortfall of 500.

Abstracting from this example, assume without loss of generality that $d_1 \geq d_2$. Two types of cases must be distinguished. If $m \leq d_1 - d_2$ then the shortfall minimization theory would have D give all his money to C_1 and none to C_2. Alternatively, if $m \geq d_1 - d_2$ then D should give C_1 amount $(m + d_1 - d_2)/2$ and C_2 amount $(m + d_2 - d_1)/2$; since $m \geq d_1 - d_2$ and $d_1 \geq d_2$ both amounts are non-negative. The formulas return the same answer in the overlapping case where $m = d_1 - d_2$.

Again, this theory can be shown to be non-aggregative by example. Suppose that a second debtor, D^*, owes 600 to C_1 and 900 to C_2 but has only 200. This theory directs him to give all 200 to C_2, since this is the most he can do to reduce C_2's shortfall, which remains larger than C_1's. The aggregate result is that C_1 receives 400 from D and 0 from D^*, leaving her with a total shortfall of 1,100, while C_2 receives 100 from D and 200 from D^* leaving her with a total shortfall of 1,200. However, had D and D^* acted as a collective agent distributing $m + m^*$ then both C_1 and C_2 would have received 350, leaving each of them with a shortfall of 1,150. Once again, if each agent acts in isolation, applying the theory of fairness only to his own distributive problem and without regard to what others will do, the outcome differs from when they act in concert, even though they act as the theory directs.

THE GENERAL CASE

Similar aggregation problems arise for any non-trivial theory (as defined below), as we now prove. Suppose that a theory of fairness would have D repay C_1 amount $F_1(m, d_1, d_2)$ and C_2 amount $F_2(m, d_1, d_2)$, where F_1 and F_2 are functions of the stated variables. By stipulation, $m < d_1 + d_2$. Also,

$$F_1(m, d_1, d_2) + F_2(m, d_1, d_2) \leq m \quad \text{for all } m, d_1 \text{ and } d_2$$

[9] Shortfall minimization is discussed in Nicholas Rescher, *Fairness* (New Brunswick, NJ, 2002), pp. 52–3, and endorsed by Cupit, 'Fairness [review]'.

https://doi.org/10.1017/S0953820815000230 Published online by Cambridge University Press

and in particular

$$F_1(m, d_1, d_2) \leq m \quad \text{for all } m,\ d_1 \text{ and } d_2$$

$$F_2(m, d_1, d_2) \leq m \quad \text{for all } m,\ d_1 \text{ and } d_2$$

The first inequality, from which the second and third follow (since the amounts in question are non-negative), expresses the fact that D has no resources beyond the amount m. Note that these last two inequalities imply the equations

$$F_1(0, d_1, d_2) = 0 \qquad \text{for all } d_1 \text{ and } d_2 \tag{Eq1}$$

$$F_2(0, d_1, d_2) = 0 \qquad \text{for all } d_1 \text{ and } d_2 \tag{Eq2}$$

since the values of F_1 and F_2 are non-negative.

Two further facts about the functions F_1 and F_2 are needed.

$$F_1(m, 0, d_2) = 0 \qquad \text{for all } d_2 \text{ and } m \tag{Eq3}$$

$$F_2(m, d_1, 0) = 0 \qquad \text{for all } d_1 \text{ and } m \tag{Eq4}$$

(Eq3) says that if D owes C_1 nothing – C_1 has no claim on D – then the fair thing for D to do vis-à-vis C_1 is not to pay her anything; similarly for (Eq4) if D owes C_2 nothing.

Finally, we define precisely two items of terminology introduced earlier. A function is said to be *trivial* if it is identically zero, i.e. it outputs 0 for any inputs:

$$F(m, d_1, d_2) = 0 \qquad \text{for all } m,\ d_1 \text{ and } d_2.$$

To introduce the second definition, recall that, intuitively, repayment functions aggregate over individual cases if the amount C_1 should receive in Case 1 is the same as she receives in Case 2, and likewise for C_2. More formally, we say that the function F_1 is *aggregative* just when, for all m, m^*, d_1, d_2, d_1^*, d_2^*,

$$F_1\left(m + m^*, d_1 + d_1^*, d_2 + d_2^*\right) = F_1(m, d_1, d_2) + F_1\left(m^*, d_1^*, d_2^*\right)$$

and similarly for F_2. With this terminology and conditions in place, the result will now follow.

The general proposition is that if (Eq1) and (Eq3) are satisfied, then F_1 is aggregative only if it is trivial. Similarly for F_2 given (Eq2) and (Eq4). The proof proceeds simply by noting that if F_1 is aggregative then

$$F_1(m, d_1, d_2) = F_1(m, 0, d_2) + F_1(0, d_1, 0)$$

https://doi.org/10.1017/S0953820815000230 Published online by Cambridge University Press

By (Eq3), $F_1(m, 0, d_2) = 0$. By (Eq1), $F_1(0, d_1, 0) = 0$. Hence for all m, d_1, d_2

$$F_1(m, d_1, d_2) = 0 + 0 = 0.$$

A parallel argument can be run for F_2 using (Eq2) and (Eq4).

It is worth briefly pointing out two further generalizations of this already general result. The first generalization is to the case of M debtors and N creditors, which follows immediately from the two-Debtor, two-Creditor case. If a fairness function (which outputs the amount the i^{th} Debtor should repay the j^{th} Creditor) is aggregative in the M-Debtor, N-Creditor case then it must also be aggregative when restricted to any two of these Debtors and any two of these Creditors (setting all other debts equal to 0). Our result tells us it must be trivial for this pair, and since this is true of any pair then by aggregation the general function must also be trivial.

The second generalization is to the case in which claims differ not only in size but also in strength. As noted, we have so far assumed equal strength of claims. Broome holds that stronger claims should receive more satisfaction, but weaker claims should still receive some satisfaction.[10] Another approach would be to say that claims should be satisfied in order of strength, so a claim of strength X receives no satisfaction until all stronger claims are fully met.[11] This view is obviously non-aggregative (the result of D and D^* acting independently will differ from that of their acting in concert if e.g. all of the claims on D^* are stronger than those on D). Our proof does not incorporate strength of claims, so it directly concerns only cases where claims are equal in strength. Nonetheless, our result applies in this more general setting too, since a counterexample to a restricted version of a theory (in this instance, in which all claims are of equal strength) also serves as a counterexample to the general theory. (This holds on the natural assumption that, on a single occasion, D's owing £M to C with strength X and £N to the same creditor C with equal strength X is equivalent to D's owing £$M + N$ to C with strength X.) Aggregation thus fails for any (non-trivial) theory of fairness that takes strength of claims into account. This applies also to a fairness function including any other parameters, assuming these can be set to the same value.

[10] Broome, 'Fairness', p. 95.

[11] Peter Stone, 'Lotteries, Justice and Probability', *Journal of Theoretical Politics* 21 (2009), pp. 395–409, at 398.

https://doi.org/10.1017/S0953820815000230 Published online by Cambridge University Press

ILLUSTRATIONS

Although we have concentrated on Debtor–Creditor cases for simplicity, our discussion has much greater application. Most obviously, as mentioned earlier, the good in question can be any divisible good whatsoever (time, amount of a divisible commodity, amount of care, quality, etc.) or the monetary equivalent. Further, the agents in question need not be individuals; they may also be groups, corporations, nation states, etc. Moreover, as we now briefly illustrate, aggregation can occur in various ways, for instance we may aggregate over agents, over time and over goods.

First, aggregation might occur across different agents. This was the case in our earlier example, where two debtors each acting fairly produce an aggregate outcome that seems unfair. To give another example, if Mother and Father each treat their children fairly, but do so independently, the result may differ from what would happen if they pooled their resources and divided the total fairly.

Second, aggregation might also occur over time. For instance, one agent may respond fairly to two claims on Monday and then respond fairly to two claims on Tuesday, but the overall result may be different from what would have occurred had she made one single distribution of goods. This case straightforwardly parallels that of different agents since, in effect, we can regard the agent on Monday and the agent on Tuesday as if they were different agents.

Third, aggregation may occur across different goods to be distributed. For instance, suppose a debtor has debts in both dollars ($) and pounds (£). Dividing the dollars and the pounds that she has fairly between her creditors may result in a different outcome than would arise were she to change all her funds to a single currency, and then divide the whole fairly between her creditors.

COMMENTS

Our result is entirely general, and rests only on the mildest assumptions on the 'fairness functions'. We may paraphrase it by saying that non-trivial functions are not aggregative. Its import, however, must not be exaggerated. The claim is not that, for any non-trivial theory of fairness, aggregation cannot be achieved in any circumstances whatsoever. It is rather that, for any non-trivial theory of fairness, aggregation fails in at least some cases. Indeed, for Broome's theory, aggregation fails in almost all cases. (Specifically, when the denominators in the fractions below are non-zero, aggregation holds for Broome's proportionality theory in the two-Debtor, two-Creditor case just when either $m/m^* = (d_1 + d_2)/(d_1^* + d_2^*)$ or

https://doi.org/10.1017/S0953820815000230 Published online by Cambridge University Press

$d_1/d_2 = d_1^*/d_2^*$. If m, d_1, d_2, m^*, d_1^*, d_2^* take values in the positive real numbers, then the equations define two surfaces each of six-dimensional volume 0 in a space with infinite six-dimensional volume.) In sum, aggregated individual fair allocations and collective fair allocations can be identical; what our result shows is that the range of cases for which this holds must necessarily fall short of *all* the cases, for any non-trivial theory of fairness satisfying the innocuous conditions set out above.

Second, our result should not be confused with the more familiar point that, sometimes, an act that would be deemed unfair if taken in isolation is needed in order to produce a fair outcome overall. Our result supports this point, but is distinct from it, as well as the argument that usually supports it. The usual argument is that an unfair action taken in isolation is often required to redress some previous *unfair* actions; for instance, affirmative action programmes may give goods to those with weaker claims, in order to redress prior unfairness. Our point is that a series of individually fair actions may give rise to an outcome that would be judged unfair had it resulted from a single action. Thus, an action that is unfair (taken in isolation) may be required to preserve an overall pattern of fairness, even when all previous actions were *fair* (taken in isolation). This argument for the necessity of redress is much less familiar and interestingly different from the usual one.

CONCLUSION

Our result is a datum that any comprehensive theory of fairness must respect. Our proof demonstrates that this problem cannot be escaped or evaded, for any acceptable (non-trivial) theory. Since no theory of fairness we know of has dealt with the problem of non-aggregativity and every (non-trivial) theory is subject to it, they are all to that extent problematic, or at least incomplete. Although it is not our ambition to articulate in detail how various theorists of fairness should respond, it is worth briefly stressing three morals that may be drawn for theories of fairness.

What we have shown is that individually fair actions must in some circumstances inevitably lead to collectively unfair actions. Every theory of fairness must clearly articulate the level of collectivity at which it is supposed to apply in any given case, for different choices will lead to different outcomes. Relatedly, theorists of fairness must broaden their range to take in dynamic (i.e. multi-period) problems as well as static ones, which tend to be the focus of current theories. As demonstrated, the aggregation of solutions to consecutive single-period distribution problems is in general not equal to the solution of the temporally aggregated distribution problem (except for the trivial

https://doi.org/10.1017/S0953820815000230 Published online by Cambridge University Press

theory). Finally, even if each of us treats everyone else fairly, there may nonetheless be aggregate unfairness in the resulting states of affairs, so we may wish to create institutions that correct for this.[12] Tackling the problem through institutional means is, we tentatively suggest, probably less demanding and less objectionable than each of us trying to assess the effects of our actions on overall outcome fairness. With the proper institutional correctives in place, we need only ensure that we treat others fairly. All these morals apply to any (non-trivial) theory of fairness.[13]

alexander.paseau@philosophy.ox.ac.uk
b.m.saunders@soton.ac.uk

[12] John Rawls, *Political Liberalism* (New York, 1993), pp. 265–71.

[13] The authors share authorship and credit equally for this work. A predecessor of this article was presented at the Universities of Stirling (September 2012), St Andrews (September 2012) and Oxford (January 2013). The authors thank these audiences, along with John Broome, Hugo Dixon, Hugh Lazenby, Patrick Tomlin and two anonymous referees for the journal for their comments.

https://doi.org/10.1017/S0953820815000230 Published online by Cambridge University Press