Group epistemic value

Jeffrey Dunn[1,2]

Abstract Sometimes we are interested in how groups are doing epistemically in aggregate. For instance, we may want to know the epistemic impact of a change in school curriculum or the epistemic impact of abolishing peer review in the sciences. Being able to say something about how groups are doing epistemically is especially important if one is interested in pursuing a consequentialist approach to social epistemology of the sort championed by Goldman (Knowledge in a social world. Oxford University Press, Oxford, 1999). According to this approach we evaluate social practices and institutions from an epistemic perspective based on how well they promote the aggregate level of epistemic value across a community. The aim of this paper is to investigate this concept of group epistemic value and defend a particular way of measuring it.

1 Introduction

Sometimes we are interested in how groups are doing epistemically *in aggregate*. Consider the following two examples.

Case 1 In 2014 the Indiana State Board of Education approved new education standards for several key subjects, including math and English. These new standards led educators to shift their curricula in certain ways. A natural question to ask: when we compare the Indiana students who were educated before the change to those who

✉ Jeffrey Dunn
    jeffreydunn@depauw.edu

1   University of Agder, Kristiansand, Norway

2   DePauw University, Greencastle, USA

were educated after the change, which group is doing better? There are, of course, several things we might mean by 'better' here. We might wonder which group is having more success at college admission, which group is making more money, which group is satisfied with their life, etc. But there is, it seems, a particularly *epistemic* question that is of central importance: which group is doing better with respect to whatever matters epistemically? That is, which group is doing better with respect to knowledge, accuracy, understanding, or the like?

Case 2 Suppose that peer review in academic research is abolished, perhaps along the lines suggested by Heesen and Bright (2020). Is this a good idea? Such a move would of course have many consequences and the decision could be evaluated in many ways. But an epistemic question is central here. How are we to evaluate such a change in journal policy with respect to knowledge, accuracy, understanding, or the like? Indeed, this is how Heesen & Bright frame the issue themselves: "We evaluate these changes [in peer review] in terms of their expected effect on the ability of the scientific community to produce scientific knowledge in an efficient manner."

Examples like these could be multiplied, but these are enough to show that sometimes we want to know how a group is faring epistemically. In addition to specific cases, there are also more theoretical reasons to be interested in the epistemic evaluations of groups, especially if one takes a consequentialist or teleological approach to epistemology. According to such approaches, epistemic value–whether this be accuracy, knowledge, understanding, or something else–comes first. Our epistemic norms are then derived by considering how to best promote this value. Such views are structurally similar to consequentialist or teleological approaches in ethics (such as classical utilitarianism).[1] One instance of this approach is so-called *accuracy-first epistemology*, initiated by Joyce (1998) and most notably developed by Pettigrew (2016). For Pettigrew (and others), accuracy is the fundamental epistemic value. We score belief states based on their accuracy, and then formulate epistemic norms by determining which belief-forming strategies are best to adopt with the goal of gaining as much accuracy as possible.[2]

More germane to the examples above is a different sort of consequentialist approach to epistemology, championed by Alvin Goldman (1999) in the area of social epistemology. Like the accuracy-first epistemologists, Goldman endorses the thesis that accuracy is the sole fundamental epistemic value. He then argues that we

---

[1] For discussion of such views in epistemology see Ahlstrom-Vij and Dunn (2018).

[2] Other representative work in this area includes Greaves and Wallace (2006), Gibbard (2008), Joyce (2009), Leitgeb and Pettigrew (2010a,b), Greaves (2013), Konek and Levinstein (2019), and Carr (2017).

should epistemically evaluate social practices and institutions based on how well they promote the aggregate level of this value across an entire community.[34]

It is important to any consequentialist approach to epistemology that one can, at least in principle, measure the kind of epistemic value that is to be promoted. For instance, it is important to Pettigrew-inspired accuracy-first epistemology that we can measure the accuracy of belief states. And indeed this is an issue that has received significant attention.[5] For the same reason, it is important for Goldman-inspired social epistemology that there is some way to measure, at least in principle, the aggregate level of epistemic value for a group. Very little attention has been paid to this issue. But there is, I think, something to Goldman's idea that we should evaluate social practices and institutions based on how well they promote the aggregate level of this value across an entire community. And as the two examples above demonstrate, there are particular cases where we want to know about the aggregate level of epistemic value even if we aren't explicitly committed to Goldman-inspired social epistemology.

That, then, is the focus of this paper: how should we understand the notion of *aggregate epistemic value* that motivates Goldman's social epistemology and is also made salient by the examples above? To be clear, the focus is not on whether Goldman's approach to social epistemology is tenable. The focus, instead, is whether there is a way to understand aggregate epistemic value where we can think of this value as something that is to be promoted. If there is *no* plausible way to understand this notion, then Goldman's approach is untenable. But even if there is, this does not settle the complicated question of whether we *should* promote such aggregate epistemic value.

In what follows I'll show that there are two main approaches to thinking about aggregate epistemic value, only one of which is suitable to Goldman's idea that such value is to be promoted. There are, however, many instances of this kind of approach to aggregate epistemic value and at first glance, all such instances have

---

[3] Heesen and Bright ([2020]), mentioned above, explicitly place their work in this Goldman-inspired approach to social epistemology.

[4] Here is what Goldman says in full: "Many social practices aim to disseminate information to multiple agents, and their success should be judged by their propensity to increase the V-values of many agents' belief states, not just the belief states of a single agent. Thus, we should be interested in the *aggregate* level of knowledge of an entire community (or a subset thereof)." (p. 93) 'V-value' is Goldman's term for veritistic epistemic value, that is, accuracy. And though he speaks of increasing the aggregate level of *knowledge*, he is clear that he is using a "thin" sense of knowledge according to which it only refers to true belief.

[5] One part of this literature is the debate over which *scoring rule* is the appropriate one for measuring the accuracy of an individual credence. The Brier score is one such score. It says that a credence in P of n is given a score of $(1 - n)^2$ when P is true. For representative work in this area see Pettigrew ([2016]), chapters 3 and 4, Levinstein ([2012]), Fallis and Lewis ([2016]), and Dunn ([2019]). Another part of this literature concerns the question of how a score for an entire doxastic state should be composed from scores for individual beliefs or credences. One key question here is how to think about cases where a believer adds or removes credences or beliefs. The first paper to squarely deal with this issue is Carr ([2015]); for other work see Pettigrew ([2018]), Pérez Carballo ([2018]), and Talbot ([2019]). I will say more about this literature in Sect. 2.2 and how it relates to the project of this paper.

intolerable consequences. To make this more precise, I prove two impossibility theorems for measures of aggregate epistemic value, but then argue that one particular way of measuring such value can be defended. This I interpret as a partial defense of Goldman's project. It shows that there is a defensible notion of aggregate epistemic value that it would make sense to promote. Whether such value should really be promoted is a question I leave unanswered in this paper.

2 Two approaches to group epistemic value

As already noted, there has been very little work on how to understand aggregate epistemic value. Perhaps surprisingly, Goldman himself does not give much attention to the question. He does give one example, however, and presenting that example is a nice way to illustrate the two different main approaches to the question. In the example Goldman (1999, pp. 93–4) gives, we are to imagine two groups interested in the question whether P (we assume P is true), where the group members each have their own degrees of belief with respect to P:

Group 1	Group 2
$c(P) = 0.4$	$c(P) = 0.7$
$c(P) = 0.7$	$c(P) = 0.9$
$c(P) = 0.9$	$c(P) = 0.6$
$c(P) = 0.2$	$c(P) = 0.8$

Goldman proposes that aggregate epistemic value can be determined by taking the arithmetic mean of the degrees of belief of the group members with respect to P. Group 1's mean degree of belief in P is 0.55. Group 2's mean degree of belief in P is 0.75. Since P is true, Group 2 has more aggregate epistemic value.

But what are we doing when we take the mean of these members' degrees of belief? There are two ways of interpreting this example. According to one way, in taking the arithmetic mean of the degrees of beliefs of the members, we are constructing the *group's* degree of belief. Group 1, on this interpretation, has a group belief in P of 0.55 while Group 2 has a group belief in P of 0.75. The epistemic value of those *group* belief states are then measured according to their closeness to 1. A different way of interpreting his example is that each member's degree of belief in P corresponds to the epistemic value for that very degree of belief since (in this simple example) the accuracy of a degree of belief increases as the degree of belief approaches 1. According to this interpretation what we are really doing is first calculating the epistemic value for each member's degree of belief and then combining the epistemic value scores for each member by taking the arithmetic mean.

These are two different ways of thinking about group epistemic value. According to the first approach we first combine members' beliefs to yield a group belief state, which is itself evaluated for epistemic value. According to the second approach, we first score the members' belief states for epistemic value and then combine these

scores to yield the level of group epistemic value. In Goldman's simple example, these two methods happen to yield exactly the same result. In general, however, that is not the case. I turn now to considering the different approaches in turn.

2.1 Combine-then-score

According to the *combine-then-score* approach, to determine group epistemic value, you first construct or discover the *group* belief state and then score this group belief state for epistemic value. There are several different ways you might construct (or discover) a group belief state. Judgment aggregation rules provide us with one possible method.[6] Suppose we have an agenda of propositions, and group members believe some of the propositions in this agenda. A judgment aggregation rule takes as input the sets of propositions corresponding to the group members beliefs with respect to that agenda and outputs a single set of propositions that is to represent the aggregate judgment of the group over the agenda. A simple and familiar judgment aggregation rule is the simple majority rule. Suppose the agenda of propositions consists of P, Q, R and their negations. We might have three group members with the following beliefs:

	Beliefs		
Member 1	P	Q	R
Member 2	$\sim P$	$\sim Q$	$\sim R$
Member 3	P	Q	$\sim R$

The simple majority rule would issue the following aggregate judgement for the group: $\{P, \sim Q, \sim R\}$. Of course, there are more complicated judgment aggregation rules that one could adopt.

One potential limitation for this particular approach to group epistemic value is that judgment aggregation rules are not typically designed to handle cases where group members do not have beliefs with respect to the same set of propositions. In the context of group epistemic value, this limitation may be problematic. Consider two groups. The first:

	Beliefs	
Member 1	A	B
Member 2	A	B
Member 3	A	B

---

[6] On judgment aggregation, see List and Pettit (2011).

The second:

	Beliefs		
Member 1	*A*	*B*	*C*
Member 2	*A*	*B*	*D*
Member 3	*A*	*B*	*E*

Suppose that all of *A, B, C, D,* and *E* are true. In such a case the second group has more aggregate epistemic value than the first, but traditional judgment aggregation rules won't deliver this. Consider the simple majority rule. Both groups end up with the same aggregate belief set: {*A*, *B*}. So on a combine-then-score approach, these two groups have the same level of aggregate epistemic value.

There are, however, different ways to think about group belief states. Belief merging or belief revision is a different option, which gets around the particular problem raised above.[7] The basic idea behind belief merging is that we try to construct the maximal consistent subset of all the beliefs that the members hold. In the example above, the second group would have the aggregate belief set: {*A*, *B*, *C*, *D*, *E*} compared to the aggregate belief set for the first group, which would be just: {*A*, *B*}. Consider a slightly more interesting example. Suppose there are two members in a group. One believes {*A, A → B*}. The other believes {∼*B, A, C*}. The set of all these beliefs together is: {*A, A → B,* ∼*B, C*}. The maximal consistent subset of these is either {*A, A → B, C*} or {*A → B,* ∼*B, C*} or {*A,* ∼*B, C*}. Different approaches to belief merging have different rules about how to select the final group belief state.

I've briefly considered two formal approaches to constructing group belief states. It is worth noting that there is nothing about the combine-then-score approach that requires a *formal* method for combining belief states. We could, for instance, say that the group believes *P* just in case were the group members to pool all of their information the group would reach consensus with respect to *P*. We then score that group belief state for epistemic value. This would still be an instance of the combine-then-score approach.

What should we think about the general combine-then-score approach to aggregate epistemic value? Some version of such an approach might be the best way to understand what it means to say one group is better informed than another when there is a difference in what we might call *coverage* by one group compared to another. Consider a simple example. Suppose that both the Oxford/AstraZeneca and NIH/Moderna vaccine labs have very knowledgeable staff members. In particular, both labs have 10 members, each of whom has the same high amount of epistemic value. However, whereas the 10 members of the Oxford lab specialize in distinct areas, the 10 members of the NIH lab completely duplicate the expertise of the other (that is, each member of the NIH lab is an epistemic duplicate of the others). In such

---

[7] Early work on belief merging is in Baral et al. (1991). Early work on belief revision can be found in Alchourrón et al. (1985) and Gärdenfors (1988). For a philosophical introduction to some of this work, see Pigozzi (2016).

a case, we might say that the Oxford lab knows[8] more about vaccines than does the NIH lab. Though standard judgment aggregation rules do not handle this case well, a suitably worked up version of the belief merging approach could vindicate such a claim.

That said, the combine-then-score approach isn't going to work well for Goldman-inspired social epistemology. To see this, suppose that we have 10 very simple agents in a group. They are simple because they are only capable of one of two belief states. Each agent can believe there is a predator nearby (P) or can believe there is not a predator nearby ($\sim P$). Suppose one way of structuring this simple group leads 9 out of 10 to believe P (in a situation where P is true). Suppose a different way of structuring this simple group leads 8 out of 10 to believe P in the same situation. If epistemic value is just accuracy, then the aggregate level of epistemic value is greater for the first group than for the second. Put another way, suppose a certain social institution, such as a system of warning calls, leads to the first group having 9 of 10 agents believing P whereas some alternative system of warning calls leads to the second group having 8 of 10 agents believing P. The first system of warning calls is the one that Goldman should recommend because it leads to more aggregate true belief.

But the combine-then-score approach cannot regularly give such a verdict. In this simple case, there are three options for what the combined group belief state might be: the group believes P, the group believes $\sim P$, or the group has no belief with respect to P. In a case like this, it is natural to say that both the first and the second groups believe P. But then their aggregate scores will be the same, contrary to what was desired. Of course, it might be that the 9/10 group is assigned the group belief that P whereas the 8/10 group is assigned no group belief. But in that case, simply change the example so that we are comparing the 8/10 group with a 7/10 group. Presumably they would both be groups that are assigned no group belief, so would get the same score. But, again, there is a difference in the aggregate level of true belief.[9]

Let me be clear about the claim in this section. I do not claim that no version of the combine-then-score approach is an interesting measure of *something*. What I claim is that it is not a good measure for the aggregate level of epistemic value within a group that can be paired with the kind of social epistemology that Goldman suggests.

2.2 Score-then-combine

According to the previous approach, we first combine the epistemic states of the group members and then score this single state. The second approach inverts this. On this *score-then-combine* approach we first score the epistemic state of each

---

[8] To stay neutral between different views about fundamental epistemic value, I am using the word 'know' in an ambiguous way to possibly refer to merely true belief or possibly refer to something more robust.

[9] And if one objects that the 7/10 group could be assigned the belief that $\sim P$ in this case, we can simply run the same argument with a 7/10 and 6/10 group.

group member and then combine these scores in some way. We are trying to determine, to put it roughly, how much epistemic value is distributed amongst a group. That is, we want a measure of group epistemic value that is analogous to measures of welfare for populations. We want, then, a function that takes as input the level of epistemic value for each member of the group and gives us a level of epistemic value for the group as a whole. The analogy to population axiology is clear, where the concern is to determine the total welfare of a population given the welfare of the individuals that make up that population.[10]

One downside of this approach is that it misses differences in content coverage as exemplified by the Oxford lab and NIH lab example above. On the score-then-combine approach, it is hard to see how we could treat the two labs differently. For each member is scored independently of the others and by construction they are all doing equally well. Combining equal numbers of equally good scores, however, is bound to produce the same aggregate score.

However, even if the score-then-combine approach cannot handle such cases, it seems that it might capture something interesting. Such an approach can easily say, for instance, that the group where 9 of 10 members get it right has more epistemic value than the group where 8 of 10 members get it right. For this reason, such an approach seems to be on the right track for how to understand the idea of aggregate epistemic value.

But it turns out that many proposals for how to devise a score-then-combine measure of group epistemic value face severe problems. The puzzle here is very similar to, though in some ways distinct from, the puzzle about how to measure the welfare of populations. It seems obvious that we must be able to make such comparisons and yet every plausible move seems blocked by a counterintuitive consequence. Score-then-combine measures of group epistemic value are similar in this. In the next several sections, I will consider plausible score-then-combine measures of group accuracy and show the problems they encounter. After this, I will defend one view in particular.

The discussion to follow is related in interesting ways to the literature on what Jennifer Carr (2015) has called epistemic expansions and contractions (see footnote 5). Suppose S has beliefs or credences over a set of propositions at one time. At a later time S may have beliefs or credences over a set of propositions that includes all the propositions from the earlier time and more (so be an expansion), or S may have beliefs or credences over a set of propositions that lacks some of the propositions from the earlier time (and so be a contraction), or it may be a combination of both. The key question here is how to (and whether we can) compare these different belief states for accuracy. How one answers this question depends on how one thinks a

---

[10] The classic work in this area is Parfit's (1984) *Reasons and Persons*. The analogy to population axiology raises an interesting aside. Could one pursue the combine-then-score approach in population axiology? Such a view doesn't get discussed because, I think, it is not clear how you could combine lives. You could imagine stringing a series of lives together to make one long, time-extended life but this would distort a variety of things such as the importance of relationships as well as order effects that might be present when it comes to welfare. This, then, seems to be one disanalogy between population axiology and social epistemology axiology: the combine-then-score approach of the previous subsection isn't a genuine option in population axiology.

score for an entire doxastic state should be composed from scores for individual beliefs or credences. In addressing this, analogies to population axiology have been explored, particularly by Pettigrew (2018) and Talbot (2019).

There are structural similarities between this literature and the project in this paper. Here we are asking how to compare different groups of believers for epistemic value. How one answers *this* question depends on how one thinks a score for a group should be composed from scores for individual believers. And as just noted, there are some fruitful analogies to population axiology here, too.

But there are also some important differences between these two projects. Most obviously, the considerations about what makes a group have a certain level of epistemic value may be different from considerations about what makes an individual accurate. This is for several reasons. For one, in the group case we can have multiple group members all of whom have the same doxastic states. The analogous thing doesn't happen in the individual case. For that to occur would be for one individual to have multiple beliefs or credences in the same proposition (e.g., an individual who has three distinct beliefs that P). Second, when it comes to individual believers, there is no distinction between score-then-combine approaches and combine-then-score approaches. In the case of a single believer we are only dealing with one doxastic state, and so combine-then-score makes no sense. The only option there is to score each belief or credence and then combine those scores somehow.

In the next several sections we will consider particular score-then-combine proposals for measuring the aggregate epistemic value of groups. Before that, two quick assumptions should be stated. First, the arguments below work no matter how you think of epistemic value (e.g., whether it corresponds to accuracy, to understanding, to knowledge, or some combination of these). However, the presentation is smoother if we focus on one of these, so I will assume veritism in what follows: that epistemic value consists solely in accuracy. Second, I will assume that we have some way of measuring the level of epistemic value for individuals. Given the first assumption, this will be a measure of accuracy for individuals, which I will call their *individual accuracy scores*.[11] This is, in essence, to assume that the debate about epistemic expansions and contractions has already been decided. Notice two important things about this second assumption. First, it is consistent with the assumption that there are individual accuracy scores that certain propositions are weighted more heavily when such scores are calculated. Thus, this assumption is

---

[11] In essence, this means that I am assuming the debates over epistemic expansions and contractions just discussed, as well as the debate over the correct scoring rule for individual beliefs or credences, have all been settled. Just to fix ideas, one individual accuracy score simply sums the number of true beliefs and subtracts the number of false beliefs. If we think that beliefs come in degrees, another way of thinking of individual accuracy is via a scoring rule, such as the Brier score (see footnote 5). It is important to note, though, that while I am assuming these debates are settled in what follows, this doesn't mean that theorizing about the group case is held hostage to such debates. Indeed, as I will point out later in the paper, successfully answering some objections to certain proposals for group accuracy might place some constraints on what we say about individual accuracy (see footnote 21).

consistent with the idea that being right about important matters contributes more to one's individual epistemic value than does being right about trivial matters.[12] Second, notice that this assumption is also consistent with the idea that in some contexts we only care about group members' individual accuracy scores *with respect to some special set of propositions*. So, this assumption doesn't preclude the idea that sometimes we may want to score a group's accuracy with respect to, say, the 2020 US election.

With these assumptions stated our task is thus clear: we want to parlay measures of individual accuracy for group members into a measure of group accuracy.

3 Average Accuracy

There are two score-then-combine proposals for how to measure group accuracy that immediately suggest themselves. The first is:

Total Group Accuracy The accuracy of a group, X, is the sum of the individual accuracy scores for each member of X.

The second natural proposal is:

Average Group Accuracy the accuracy of a group, X, is the sum of the individual accuracy scores for each member of X, divided by the number of members in X.

These proposals give the same rankings when the groups being compared are of the same size.[13] But they can give different verdicts when groups are not the same size. Both these accounts seem to run into immediate problems. We start with Average Group Accuracy.

The first problem with this view arises when we consider the following two groups:

Group A   2 people each with a high individual accuracy score of h.
Group B   1000 people each with a high individual accuracy score of h, 1 person with an individual accuracy score slightly less than h.

According to Average Group Accuracy A is more accurate than B. But that seems to be mistaken: if we are promoting accuracy, we should choose an outcome like group B over group A. There is also a mirror image of this problem when we consider very *inaccurate* groups. Let group A now contain two members who are each entirely inaccurate in their beliefs. The natural way to represent this is to assign each person a *negative* individual accuracy score, say -l. Let group B now have 1000 members who are similarly inaccurate with individual accuracy scores of $-l$, and one member who is doing slightly better, perhaps because in addition to all her

---

[12] See Pérez Carballo (2018) for more on the topic of weighting propositions in this way for individual accuracy scores.

[13] This is because for $n \geq 0$, $x/n > y/n$ iff $x > y$.

mistaken beliefs, she has one true belief. Her individual accuracy score is $-l + \varepsilon > -l$. According to Average Group Accuracy, group A is more inaccurate than B. But again this seems wrong. It seems that in this case B is more inaccurate since B contains so *many* people who are so mistaken.

The second problem for Average Group Accuracy is that it violates an epistemic version of what is sometimes called the Mere Addition Principle in population ethics. In the literature on population ethics, we have a mere addition when we add an individual to a population who has positive welfare, and where adding this person doesn't affect the welfare of any of the other people in that population. The **Mere Addition Principle** says that a population doesn't get *worse* by adding such an individual. When it comes to accuracy, the principle says that adding a member with positive accuracy should not make a group less accurate. Average Group Accuracy violates this principle, since adding a member that has a positive accuracy lower than the average of the group lowers the average and hence makes the group less accurate.

These first two problems are relatively well known in the context of population ethics. As far as I know, a third problem for Average Group Accuracy is novel. The problem is that Average Group Accuracy violates a plausible principle about combining groups. Suppose that we have divided a set of people into four distinct groups, *X, Y, Z, W*, in such a way that X is more accurate than Y and Z is more accurate than W. The plausible principle says that if you instead think of the set of people as divided into *two* groups, one that is a combination of X and Z (the more accurate subgroups) and one that is the combination of Y and W (the less accurate subgroups), then the X and Z group is more accurate than the Y and W group. More carefully, the principle is:

> **No Simpson:** If Group X is more accurate than Group Y and Group Z is more accurate than Group W, then Group $X \cup Z$ is more accurate than $Y \cup W$.[14]

Violating No Simpson is, I think, an intolerable result. A violation means that the way you carve up groups can affect whether or not a particular intervention increased or decreased accuracy. For instance, suppose that there is a school district containing two schools. The district implements a new policy and it turns out that each school improves in accuracy after the change. Nevertheless, if our measure of accuracy violates No Simpson, it could be that the district as a whole *decreases* in accuracy. No plausible measure of group accuracy should have this result.

Average Group Accuracy, however, violates No Simpson. To see this consider:

Group C   85 members, each with 100 accuracy
Group D   100 members, each with 99 accuracy
Group E   100 members, each with 50 accuracy
Group F   100 members, each with 49 accuracy

---

[14] The principle is so-called because any violation of it is an instance of what is called Simpson's Paradox. The classic illustration of Simpson's Paradox concerns admissions data from UC Berkeley in the 1970s. Most departments admitted a higher percentage of women than men and yet the admissions data for the university as a whole showed a higher rate of acceptance for men.

According to Average Group Accuracy, C is more accurate than D and E is more accurate than F. But consider:

Group $C \cup E$ 85 members with 100 accuracy, 100 members with 50 accuracy. Average: 73

Group $D \cup F$ 100 members with 99 accuracy, 100 members with 49 accuracy. Average: 74

So, Average Group Accuracy violates No Simpson.

I think that No Simpson is an intuitively plausible constraint on group accuracy measures. But because it will play a role in further arguments (and one of the theorems in the Appendix) it is worth saying more in its defense. No Simpson says that combining two groups that are more accurate than two other groups doesn't make the first combined group somehow less accurate than the second. The mere combination shouldn't have this kind of power. Consider an analogy. Suppose that David Justice was a better hitter than Derek Jeter in the 1995 baseball season. And suppose that Justice was also a better hitter than Jeter in the 1996 season. Then of course Justice is a better hitter than Jeter over the course of the *two* seasons. That is the essence of what No Simpson says.

But you might object here. It is after all intuitively plausible that if Justice has a better batting average than Jeter in 1995, and if Justice has a better batting average than Jeter in 1996, then Justice has a better batting average over the two seasons combined. But *that* is in fact false.[15] So, you might say, our intuitions are a poor guide here to whether No Simpson is a genuine constraint.

In response, I grant that we do get confused in such cases. It *is* surprising to learn that Justice can have a better average than Jeter in each of two seasons, and yet have a worse average over the two seasons combined. But I don't think that this impugns No Simpson. For it is still plausible that if player A is a *better hitter* than player B in season X, and if player A is a *better hitter* than player B in season Y, then player A is a *better hitter* than player B over the two seasons combined. What we learn–and what is surprising–is that batting average alone is therefore not a good measure for excellence in hitting. It doesn't undermine our confidence that betterness in hitting over each of two seasons must result in betterness over the combination of seasons.

I think something similar is true of No Simpson. No Simpson is making a claim about the epistemic betterness of groups: it places a constraint on when one group can be epistemically better than another. It turns out that average accuracy can

---

[15] In 1995 Derek Jeter had a batting average of .250, which was less than David Justice's batting average in 1995 of .253. In 1996, Jeter batted .314, which was again less than Justice's batting average in 1996 of .321. Nevertheless, over the two seasons, Jeter batted .310 while Justice batted .270. This happened because most of Jeter's at-bats were in the 1996 season when he batted .314, whereas most of Justice's at-bats were in the 1995 season when he batted .253. Here is the total breakdown:

	1995	1996	Combined
Jeter:	48 (.250)	183/582 (.314)	195/630 (.310)
Justice:	104/411 (.253)	45/140 (.321)	149/551 (.270)

violate this constraint and so–perhaps surprisingly–cannot be the correct measure of accuracy. But the existence of some purported measures of accuracy that result in Simpson's paradoxical scenarios does not undermine the plausibility of the No Simpson constraint.

A distinct objection to No Simpson is to say that it gets its plausibility from thinking about only a particular kind of case, but that this kind of case is unrepresentative. No Simpson says that if group X is more accurate than group Y and Z is more accurate than W, then the combination of X and Z is more accurate than the combination of Y and W. This might bring to mind the following kind of case. A certain school district has two middle schools: West and East. For an entire year there is no change in the students enrolled at either school. The students at West Middle School in the winter (group X) are more accurate than they were in the fall (group Y). And the students at East Middle School in the winter (group Z) are more accurate than they were in the fall (group W). In this case, one might think, the district middle schoolers must be more accurate in the winter (groups X and Z) than they were in the fall (groups Y and W). No Simpson *does* say that this must be the case. But this is not the kind of case where Average Group Accuracy violates No Simpson. Average Group Accuracy does not violate No Simpson when the X and Y groups are the same size and the Z and W groups are the same size. So, one might object, the plausibility of No Simpson comes from thinking of cases where this is the case even though No Simpson extends beyond this, to cases where the group sizes differ.

In response, it may be that some of No Simpson's plausibility comes from thinking about cases like the middle school one above. But I think it is plausible even in the cases where the groups are of different sizes. The main idea guiding our thoughts in the middle school case, I think, is that we should give the same verdict about increases or decreases in accuracy for a single period of time independent of how we choose to divide up the people over that time into groups. And this guiding idea applies even when the groups are of different sizes. Suppose that West Middle School has a net loss of students from fall to winter. Suppose that East Middle School has a net gain of students from fall to winter. Still, if both schools increase in accuracy over that time, surely the middle schoolers as a whole increase in accuracy over that time. The key idea here is that once you have the accuracy of each student at each time, and which students left and joined the schools, *how you choose to divide them into groups* shouldn't matter for how accurate we think they are. That guiding idea does drive our verdict in the simple middle school case in the paragraph above. But it also drives our verdict in more complicated cases where the groups are of different sizes. And that lends support to No Simpson in full generality, and not just in the special case mentioned above.

Now, you might have a different response to thinking about these kinds of cases. This might make you skeptical that there is a correct measure of group accuracy at all. We can talk about average accuracy, you might say, and we can talk about total accuracy, but none has the claim to being *the* measure of accuracy. Analogy: we can talk about batting averages over different time periods, we can talk about the number of hits over different time periods, but there is no measure of batting excellence overall. But note that this is not to object to No Simpson, and is in fact

perfectly compatible with it. No Simpson simply says one thing that is required to be a legitimate measure of group accuracy. One might agree with this and also think that there is in fact no measure of group accuracy.

At any rate, I maintain that the three problems discussed in this section warrant the rejection of Average Group Accuracy as a measure of aggregate epistemic value.

4 Total accuracy

We just presented three problems for Average Group Accuracy. Total Group Accuracy remedies all three of these problems because it says that the accuracy of a group is the sum of the individual accuracy scores of each group member. The first problem for Average Group Accuracy concerned groups A and B where B has many more members than A with only slightly different accuracy profiles. Total Group Accuracy can say that B is more accurate than A when both groups have relatively accurate members and can say that B is more inaccurate than A when both groups have relatively inaccurate members. The second problem for Average Group Accuracy concerned the addition of positive-accuracy members. Total Group Accuracy says that the addition of a group member with positive accuracy always improves the accuracy of the group since it always increases the total accuracy of that group. Finally, Total Group Accuracy does not violate No Simpson, since for any four numbers, a, b, c, d, if $a > b$ and $c > d$, then $a + c > b + d$.

But Total Group Accuracy faces its own challenge. The main worry for this proposal is an epistemic version of Parfit's (1984) repugnant conclusion. Consider:

Group G n members all with very high individual accuracy
Group H $n + m$ members each with just slightly greater than 0 individual accuracy

With sufficiently large m, H will be more accurate than G according to Total Group Accuracy. But this may seem wrong. Every member of G, we can suppose, is perfectly accurate about every proposition in their belief set. Every member of H, on the other hand, is doing quite a lot worse.

So, the two most obvious proposals run into trouble. To put my cards on the table, I will end up advocating for Total Group Accuracy, so I think that there is a response to the epistemic repugnant conclusion. However, that doesn't mean that it isn't a *prima facie* worry for Total Group Accuracy. And we will be better placed to see the response after considering some alternative proposals for measuring group accuracy.

5 Two-dimensional accuracy

A natural remedy to the problems in the previous section is to find some way of combining the Total and Average views. The simplest way to do this is to say that if both proposals agree that one group is more accurate than another, then we go with that recommendation; otherwise, the groups are incomparable.

Letting $\text{Tot}(X)$ be the sum of the accuracy of group X, and $\text{Avg}(X)$ be the average accuracy of group X, the view in question is:

2D Group Accuracy: If $\text{Tot}(X) > \text{Tot}(Y)$ and $\text{Avg}(X) > \text{Avg}(Y)$, then Group X is more accurate than Group Y; otherwise, X is incomparable/on a par[16] with Y.

This view runs into problems, however. The first thing to note is that it is implausible that whenever two groups differ in their ordering in terms of total accuracy and average accuracy, those groups are either on a par or incomparable. The first case comparing groups A and B from Sect. 3 can illustrate this. In that case, $\text{Avg}(A) > \text{Avg}(B)$ and $\text{Tot}(B) > \text{Tot}(A)$, yet this is not a case where we feel comparisons cannot be made. In fact, this kind of case is similar to cases that Ruth Chang (1997) calls *nominal/notable comparisons*. Chang uses this term to refer to several different kinds of comparisons, but what I have in mind are two groups where there is a nominal difference between them in terms of one dimension of evaluation but a notable difference in terms of the other dimension of evaluation. In such a case 2D Group Accuracy says that the groups are incomparable, even though they don't in fact seem to be incomparable. Here is an example:

Group I   total accuracy: 100,000; average accuracy of each member: 100
Group J   total accuracy: 1000; average accuracy of each member: 100.1
2D Group Accuracy says that I and J are incomparable or on a par. But that seems wrong. Pretty clearly, group I has better aggregate accuracy than J; the small difference in average accuracy doesn't overcome the large difference in total accuracy.

Perhaps, however, some won't agree that these nominal/notable verdicts are problematic. There is however another problem with 2D Group Accuracy: it too violates No Simpson.

Group K   10 members, each with 100 accuracy
Group L   10 members, each with 99 accuracy
Group M   11 members, each with 50 accuracy
Group N   10 members, each with 49 accuracy

---

[16] I've left it open whether the 2D Group Accuracy says that groups are *incomparable* or *on a par*. The latter term is from Chang (1997), who argues that in addition to the three standard relationships that comparable things can stand in to one another (better than, worse than, equal to), there is a fourth relation: on a par. Things stand in this fourth relationship to each other when they are comparable, but do not stand in any of the traditional three comparative relationships. As a heuristic, one can think of *on a par* as something like "rough equality". There are some differences between saying that two options are incomparable versus saying they are on a par, but it turns out that these differences won't matter for the objections considered here.

Notice that $\text{Avg}(K) > \text{Avg}(L)$ and $\text{Tot}(K) > \text{Tot}(L)$. Similarly, $\text{Avg}(M) > \text{Avg}(N)$ and $\text{Tot}(M) > \text{Tot}(N)$. So 2D Group Accuracy says that K is more accurate than L and M is more accurate than N. However, $\text{Avg}(K \cup M) < \text{Avg}(L \cup N)$. So, 2D Group Accuracy does not say that $K \cup M$ is more accurate than $L \cup N$. No Simpson is thus violated.

Finally, it is worth noting that many groups will be incomparable in terms of accuracy according to this account. This is not a welcome result if one is formulating an account of group accuracy to work with the kind of approach to social epistemology championed by Goldman. For if many groups are incomparable in terms of accuracy, we will often not get any verdicts about whether a certain policy or institution raised or lowered the aggregate epistemic value of a group.

6 Variable value views

The next option is to come up with a measurement that weights both total and average in different ways in different contexts. These are variable value views. The basic idea is that when we are dealing with a small group, accuracy is dictated primarily by total accuracy. However, when we are dealing with very large groups, accuracy is dictated primarily by average accuracy.

6.1 Sider's geometric value view

One way to pursue this idea is explored by Sider (1991) in the context of population axiology. Sider's view (which he presents but does not endorse) is built on the idea that each additional person we add to a population adds diminishing value to the population. This has the effect that for low total numbers of members, Geometric Group Accuracy approximates Total Group Accuracy. As the group gets larger, however, it more closely resembles Average Group Accuracy. When this change occurs depends on the choice of a particular parameter.[17] It can be proven that this view avoids the repugnant conclusion.

---

[17] I leave the technical details out of the main text. Here is the view stated precisely.

Geometric Group Accuracy (GGA): The accuracy of a group is given by its Geometric Value: $\sum_i u_i r^{(i-1)} + \sum_j v_j r^{(j-1)}$.
where u_i is the accuracy for the ith member in ordered set P (P: $< 1, ..., i, ..., n >$, containing the members with positive accuracy, ordered by descending accuracy), v_j is the accuracy for the jth member in ordered set N (N: $< 1, ..., j, ..., m >$, containing the members with negative accuracy, ordered by ascending accuracy), and where r is some number less than 1 and greater than 0.
For a simple example, let $r = 0.99$ and let each member have either the same positive accuracy, p, or the same negative accuracy, $-p$. With this choice, GGA says that a group with 300 positive members and 30 negative members is less accurate than a group with 250 positive members and 10 negative members. This is in line with what Average Accuracy says, and against Total Accuracy. On the other hand, GGA says that a group with 30 positive members and 3 negative members is more accurate than a belief state with 25 positive members and 1 negative member. This is in line with what Total Accuracy says, and against Average Accuracy.

  Though it escapes the repugnant conclusion, this view has been rejected in population ethics because it violates intuitions about fairness. In particular, Geometric Group Accuracy violates the following principle (the name comes from Ng, 1989):

> **Non-Antiegalitarianism:** if groups X and Y have the same number of members, every member of X has the same accuracy score, and the total accuracy of X is greater than the total accuracy of Y, then X is more accurate than Y.

To satisfy Non-Antiegalitarianism is to endorse the idea that equal spreading of resources–in this case accuracy–across a group is itself a good thing. For this reason, violations of Non-Antiegalitarianism can seem particularly objectionable when it comes to population axiology. As Sider himself notes, the view "generates rather extreme results with respect to distributive justice" (1991, p. 270). It is unclear, however, whether this is a problem for the view as applied to group accuracy. While inequality in *welfare levels* very well may be a negative feature of a population, it does not seem to be the case that inequality in *accuracy levels* is in and of itself an epistemically negative feature of a group.

  Nevertheless, I believe there is a more serious problem for the view: Geometric Group Accuracy also violates No Simpson. Here's is a specific case:

Group O	100 accurate members, with individual accuracy 9 each
Group P	50 accurate members, with individual accuracy 9 each
	1 even more accurate member, with individual accuracy 10
	49 low (but positive) accuracy members, with individual accuracy 1 each
*Group O**	Just like Group O
*Group P**	Just like Group P
*Group O ∪ O**	200 accurate members, with individual accuracy 9 each
*Group P ∪ P**	100 accurate members, with individual accuracy 9 each
	2 even more accurate member, with individual accuracy 10
	98 low (but positive) accuracy members, with individual accuracy 1 each

If we set the parameter $r = 0.99$ (see footnote 17), then although O is more accurate than O^* and P is more accurate than P^*, $O \cup O^*$ is *less accurate* than $P \cup P^*$. It is worth noting that this shows that Geometric Group Accuracy violates not only No Simpson, but also a significantly weaker version of it:

> **Weak No Simpson:** If Group X is more accurate than Group Y *and these groups are of the same size*, and Group Z is more accurate than Group W *and these groups are of the same size*, then Group $X \cup Z$ is more accurate than $Y \cup W$.[18]

---

[18] As noted in Sect. 3, Average Group Accuracy does *not* violate Weak No Simpson. It does, however, violate No Simpson even when the groups are *close to the same size*.

6.2 Ng's variable value view

A different way to pursue a variable value view comes from Yew-Kwang Ng. In his (1989) Ng proposes (but does not endorse) a possible response to Parfit's repugnant conclusion. Ng's proposal, applied to the epistemic case, works as follows. We take the accuracy of a group to be the product of the average accuracy of the group members *and* a concave function of the total number of group members. If we did not apply this concave function to the number of group members, we would simply have a view that gave the same verdicts as Total Group Accuracy (since the product of average accuracy and the number of group members just is the total accuracy). But, if we apply a concave function to the number of group members, we get something different. For small groups, this value looks very similar to Total Group Accuracy; for larger groups, this value looks similar to Average Group Accuracy.

If N is the number of group members, we let $f(N)$ be the concave function. Ng gives an illustrative example of how this concave function might work:

$$f(N) = 1 + \alpha + \alpha^2 + \ldots + \alpha^{(N-1)}, \text{ for } 0 < \alpha < 1.$$

If we let $\alpha = 0.99$, then $f(100) \approx 63.4$. So, this function has a dampening effect. Furthermore, this particular function approaches a limit, so there is no way to get arbitrarily high accuracy scores by adding more and more group members. As the total number (N) of group members increases, the contribution of each additional one to the overall accuracy is less and less important. Hence, we escape the repugnant conclusion.

Whatever the merits of Ng's view in population axiology, it is flawed as a measure of group accuracy. The first problem is that on Ng's view there are situations where it is better for group accuracy to add members with *negative* accuracy rather than members with low but *positive* accuracy.[19] This is highly counterintuitive. The situation arises because at a certain point $f(N)$ does not change much with N. Hence, adding 100 members does not alter $f(N)$ much more than adding 10 members. However, adding 100 members with very low but positive accuracy can more significantly affect the average accuracy than adding 10 members with negative accuracy.

Second, Ng's view, like Geometric Group Accuracy, violates Weak No Simpson. Here is one case that demonstrates this:

Group Q	100 members, each with accuracy level 3
Group R	50 members, each with accuracy level 3.01
*Group Q**	Just like Group Q
*Group R**	Just like Group R
*Group Q ∪ Q**	200 members, each with accuracy level 3
*Group R ∪ R**	100 members, each with accuracy level 3.01

---

[19] This is noted, in the context of population ethics, by Arrhenius (2000). It is also worth noting that the point stands even if one doesn't countenance *negative* accuracy. In that case, the objectionable point is that it can sometimes be better for group accuracy to add some members with near zero accuracy rather than some members with much higher (but still low) accuracy.

If we set $\alpha = 0.9$, then Q is more accurate than R, Q^* is more accurate than R^*, and yet $Q \cup Q^*$ is less accurate than $R \cup R^*$.

In looking at both Sider's proposal and Ng's proposal, we see that variable value views violate not just No Simpson but Weak No Simpson. This is because though the details of these variable value views are different, they are similar in that they diminish the impact of additional group members once a group reaches a certain size. This leaves them open to violations of Weak No Simpson. For this reason, variable value views are not a promising way to measure group accuracy.

7 Giving up?

After seeing the problems encountered by otherwise plausible views for measuring group accuracy, it is tempting to think that perhaps there is no measure to be had. In the Appendix I present two impossibility theorems that formalize this intuition. These impossibility theorems show that if you accept several seemingly plausible constraints on group accuracy, then there is no way to construct a measure of group accuracy. Perhaps that *is* the correct conclusion to draw, which would show that Goldman's approach to social epistemology is simply misguided.

But there is some reason to think that such a conclusion should be resisted. For we do seem to compare groups in terms of accuracy or epistemic value. One explanation for this is that while there are simple cases where such comparisons can be made, there is no account for how to do this in full generality. But in the interest of searching for something stronger than this, in the remainder of this paper, I argue for a particular way of measuring group accuracy in full generality, and respond to objections.

8 In favor of total group accuracy

In this section I argue in favor of Total Group Accuracy. Recall that the objection to Total Group Accuracy is that it leads to the epistemic repugnant conclusion. I will first argue that views avoiding the epistemic repugnant conclusion are more objectionable than simply accepting the conclusion itself. After this I will argue in a more positive way that there is a way of understanding epistemic value so that the epistemic repugnant conclusion is actually not repugnant at all.

8.1 The epistemic repugnant conclusion

I've already alluded to the epistemic repugnant conclusion, but it will be useful here to state it with a bit more precision:

> **Epistemic Repugnant Conclusion:** For any group with n members all of whom have very high accuracy, h, there is a group with $n+m$ members that is more accurate than this group even though all of the members have very low (but positive) accuracy, l^*.

Notice that the epistemic repugnant conclusion says something about a comparison between groups of *different* sizes (one with n members and one with $n + m$). I will first argue that *if* the epistemic conclusion is genuinely problematic, then this commits one to some claims about comparisons of groups that are of the *same* size. This is important because cases where groups are the same size are the relatively easy cases.[20] In part this is because in such cases, both Average Group Accuracy and Total Group Accuracy agree.

Suppose, then, that you think the epistemic repugnant conclusion is something to be avoided. Then you must think that the epistemic value of the n members with high accuracy h is more valuable than the epistemic value of the $n + m$ members with low but positive accuracy $l*$. That is: $nh > (n + m)l*$.

Now, consider two equally-sized groups:

Group S n members with very high individual accuracy a
              $n + m$ members with very low, but positive individual accuracy z
Group T $2n + m$ members with very low, but positive individual accuracy y such
              that $y > z$

In changing a group from S to T, n people lose a lot of accuracy: $(a − y)$. And $n + m$ people gain a little accuracy: $(y − z)$. Let $h = (a − y)$. Let $l* = (y − z)$. Then n people lose h and $n + m$ people gain $l*$. From the assumption that the epistemic repugnant conclusion is problematic, we know that the loss by the n people is a bigger loss than the corresponding gain by the $n + m$ people. So, if the epistemic repugnant conclusion is a problem, then group S is more accurate than group T. Notice that these two groups have the same number of members and that we can pick values of *n, m, a, y, and z* such that S is doing less well than T in terms of *both* total accuracy and average accuracy. So there is some pressure to say that group S is not more accurate than group T.

What kinds of views could avoid saying that group S is not more accurate than group T? First, any view that denies we can compare S and T for accuracy will say that S is not more accurate than T. But such views are implausible. Certainly *some* pairs of groups can be compared for accuracy. Moreover, when it comes to groups S and T we are dealing with groups of the same size. This is exactly the kind of case where accuracy should be comparable. An analogy with population axiology may be helpful: the hard cases in population axiology are thought to be the cases where we add or subtract members from a population and so are comparing populations of different sizes. The cases where the populations are the same size, in contrast, are not thought to be the ones that raise problems. So, we can reject views that say that S and T are incomparable in terms of accuracy.

The rest of the views to consider avoid saying that S is not more accurate than T by making the positive claim that S *is* more accurate than T. The first class of views that yields this result is the class of views according to which only the accuracy of certain group members matter. If, for instance, we compare groups for

---

[20] This is true not just with regard to group accuracy, but also in population axiology. Comparing two equally-sized populations for welfare is not completely straightforward, but such cases are not the hard ones in population axiology.

accuracy by only looking at each group's most accurate member, then we get the result that S is more accurate than T. But these sorts of views are implausible, since they are not in fact representing the accuracy of the group, but rather the accuracy of one, or several, members of the group. Such measures of group accuracy are analogous to dictatorship aggregation rules in the field of judgment aggregation. Dictatorship rules are thought to be undesirable because they do not represent a genuine aggregation of the group members' judgments. Similarly, these sorts of group accuracy measures do not genuinely represent the accuracy of the group members.

Another kind of view that yields the result that S is more accurate than T is a view according to which we measure accuracy by the difference between the most accurate and the least accurate members. But this view has fairly obvious problems. First, it is a kind of dictatorship accuracy measure since it does not take into account the accuracy of all the members. But more than this, it yields highly suspect verdicts: it says that any group where members have unequal individual accuracy is more accurate than any group where every member has the same individual accuracy.

Let us, then, consider measures of group accuracy that yield the result that S is more accurate than T *and* that take into account the individual accuracy scores of all the group members. There is only one other class of views of which I am aware that yield this result. These are views according to which diminishing value is given to further additions of people in such a way that the higher accuracy levels are privileged. That is, each additional member's contribution to the group's accuracy is diminished, but with this more strongly affecting those with lower rather than higher accuracy. Sider's Geometric Value View is an instance of this sort of view since we first put people in order from highest accuracy to lowest accuracy, and then apply our function that diminishes the contribution of each member. This view says that group S is more accurate than T in a way that is similar to how dictatorship accuracy measures say this. However, whereas dictatorship accuracy measures completely ignore certain members, the Geometric Value View pays less attention to the low-accuracy members.

The problem with diminishing value views, however, are twofold. First, while they can say that in *some* specific instances group S is more accurate than group T, they do not do so universally since this depends on the size of the groups in question and the choice of parameter r. Second, as we've already seen, diminishing value views violate No Simpson. The Geometric Value View, for instance, violates Strong No Simpson.[21] The reason that diminishing value views of this sort violate No Simpson can be seen fairly simply. Such views discount the contribution of each member of a group according to how far down the ranked list of accuracy that member is. For smallish groups, the discounting is not too great even for the last member, so a group all with members at some middling level of accuracy, m, can exceed the accuracy of a group with several high accuracy individuals, h, and the

---

[21] We saw that Ng's view violates No Simpson, too, but Ng's view does not have the right structure to say that S is more accurate than T. This is because Ng's view agrees with Total Group Accuracy when the groups being compared are of the same size.

rest with low accuracy, l. However, if we make these groups larger and larger but keep the same distribution of members, the discounting is so great once we reach the l-members in the second group, that they don't end up being such a liability. Hence, the unequal group, when made larger and larger, ends up being more accurate than the equal group where everyone has accuracy, m.

While I don't have a proof that we have canvassed every group accuracy view that could say that S is more accurate than T, I don't know of any other kind of view that could say this. This, together with the fact that T has both a better average and total accuracy score than S is a reason to think that group T really is more accurate than group S. Thus, I tentatively conclude that group S is not more accurate than group T. But above I argued that if the epistemic repugnant conclusion is a problem, then group S is more accurate than group T. So, this constitutes an argument for the claim that the epistemic repugnant conclusion is not a problem, that is, for accepting the epistemic repugnant conclusion.

8.2 Total group accuracy reconsidered

So far I've argued views according to which the epistemic repugnant conclusion is problematic are themselves implausible. This gives some reason to think favorably about Total Group Accuracy, since the objection to that view was that it leads to the epistemic repugnant conclusion. However, more must be said. For even if views that avoid the epistemic repugnant conclusion themselves have intolerable features, it doesn't follow that Total Group Accuracy is in the clear. It is consistent with all this that Total Group Accuracy's embrace of the epistemic repugnant conclusion is also intolerable. So, I must say something positive about why the epistemic repugnant conclusion is not actually so problematic.

My preferred answer to this challenge is to adapt a suggestion in the literature on population axiology from Mackie (1985), Tännsjö (1992) and Ryberg (1996). The solution is to be more careful in explaining what it is to have an accuracy score that is low but just barely positive. This is analogous to having a welfare level that is low, but just barely positive. In population axiology, such a welfare level is supposed to correspond to a life that is just barely worth living. This doesn't sound so great. But those cited above have claimed that a life that is just barely worth living may be one that contains quite a bit of welfare and so may not be as bad as it initially sounds. Hence, the repugnant conclusion is not repugnant. A very large population full of people who have lives just barely worth living is not one with a vast number of mostly miserable people, but instead, one with a vast number of perfectly happy people.

One may not find this plausible in the case of welfare. There are at least two objections one might bring. First, one could simply push back against the idea that a life that is barely worth living is a good life to lead. One might do this by reflecting on one's own case: my life, for instance, could get quite a lot worse before I think it would be no longer worth living. And if my current life were to get quite a lot worse, I don't think it would be the life of a very happy person.

The second objection comes from Arrhenius (2011). He points out that even if the repugnant conclusion is tolerable, many views that accept it (including views

analogous to Total Group Accuracy) lead to what he calls the *very repugnant conclusion*. This is the conclusion that a population of many, many people with lives barely worth living (with small positive value) and some people with very wretched lives (with extreme negative value) can be better than a population containing only people with very high welfare levels. Even if one is comfortable accepting the repugnant conclusion, one might hesitate to accept the very repugnant conclusion.

In response, I think that in the epistemic case things look better than in the welfare case. In response to the first objection, there is more freedom in specifying what it takes to have a positive accuracy score, that is, to say what it takes to be epistemically well off. Perhaps the naïve idea is that if you have just one more true belief than false beliefs then you have a positive individual accuracy score. But such individuals are performing just slightly better than chance at determining the truth. There is accordingly no reason to construct our individual accuracy scores to say that this corresponds to positive individual accuracy. We can consistently maintain that a positive individual accuracy score requires that one is doing quite well with respect to accuracy.[22] If we can get a vast number of people to a level of accuracy that is quite good, then this very well might be better epistemically than a case where a smaller number of people are fantastically well-informed. I do not maintain that it is *obvious* that the epistemic repugnant conclusion should be accepted; rather, I maintain that this move makes it no longer *repugnant* to accept.

What about the very repugnant conclusion? The response is similar: the very repugnant conclusion is not nearly as repugnant in the epistemic case. The epistemic very repugnant conclusion says that it can sometimes be better to get very many people to a satisfactory epistemic level, even if the cost is a few who have negative individual accuracy. It seems to me that whether this is correct is very different when we consider the case of population axiology, especially if just barely positive individual accuracy might correspond to a quite good epistemic situation. It does seem perverse to say it is acceptable to get a large population to lead satisfactory lives at the cost of causing a few extreme torture. But in the epistemic case, things are far less clear. First, those with negative accuracy are epistemically deficient in some way, but they need not be hopelessly mistaken about everything. Moreover, to be epistemically deficient, even quite severely, need not result in a life of pain and misery. Consider an example. Suppose that having a free press ultimately leads many people to be reasonably well informed. A disappointing side-effect, however, is that it allows some small number of people to be highly misinformed having been taken in by conspiracy theories and misinformation. Though this is a regret-

---

[22] I noted at the beginning of this paper that almost all the arguments are consistent with any approach to individual accuracy. Here is one case where this is not true. If one adopts, for instance, the naïve account of individual accuracy (according to which having even one more true belief than false beliefs results in a positive individual accuracy score), then the epistemic repugnant conclusion does perhaps look genuinely repugnant. This suggests that *if* we want a reasonable account of group epistemic value, we need to think carefully about our account of *individual* epistemic value.

table side-effect, we do not tend to think that this would show that having a free press is epistemically repugnant.[23]

9 Conclusion

This paper began with the observation that in certain cases we seem to care about the aggregate epistemic value possessed by a group. I also argued that getting clear on this notion of aggregate epistemic value is important if one takes a consequentialist or teleological approach to social epistemology, according to which the guiding idea is that such aggregate epistemic value is to be promoted. This, as noted, is the approach that Alvin Goldman (1999) takes towards social epistemology and it is a *prima facie* simple and attractive approach. But there is a genuine question about whether sense can be made of such a notion of aggregate epistemic value. I have argued that the answer to this question is *yes*. On the assumption that we can measure epistemic value for individuals, we can also measure aggregate epistemic value of a group made up of those individuals. This doesn't settle the question of whether such aggregate epistemic value is really something that we should aim to promote. But it does, I think, show that there is a defensible notion of aggregate epistemic value that it would make sense to promote. And so it shows that such an approach to social epistemology is worth pursuing.

---

[23] Richard Pettigrew (2018) argues for a measure of individual accuracy that scores credences in individual propositions and then sums up those scores to get the score for the total credence function. The resulting view is structurally similar to Total Group Accuracy, and also runs into an epistemic version of the repugnant conclusion. Pettigrew's response to this is to debunk our intuitions. He admits that we intuitively think that a credence function with 1 million very accurate credences is better than a credence function with 1 trillion barely more accurate than inaccurate credences. However, he claims that this reaction is because we get confused between epistemic and pragmatic evaluation. Whatever its merits in the individual case, this debunking argument won't work for Total Group Accuracy. This is because the groups we evaluate for accuracy are often not the sort of cohesive groups that are in the position to make group decisions or be evaluated pragmatically at all. Perhaps we want to evaluate for accuracy users of Wikipedia, or high school graduates in the year 2010, or the citizens of a certain country. These are not the kinds of groups that are in the position to make decisions, nor does anyone think they are, and so we could not be confusing pragmatic evaluation with epistemic evaluation in such cases.

Appendix: Impossibility Theorems

In this appendix, I prove several impossibility theorems related to different measures of group accuracy and note how the defender of Total Group Accuracy would respond.

Consider the following principles:

Nominal-Notable: Let group A contain 1 individual with accuracy level h. Then, there is a group, B, which contains $k>1$ individuals with accuracy level h- such that h- is only very slightly less than h and where group B is more accurate than group A.[24]

Transitivity: If group A is more accurate than group B and group B is more accurate than group C, then group A is more accurate than group C.

No Repugnance: If A is a group where all members have very high accuracy over h, and B is a group with more members than A, where all the members of B have very low but positive accuracy l or below, then A is not less accurate than B.

No Simpson: If Group A is more accurate than Group B, and Group C is more accurate than Group D, then Group $A \cup C$ is more accurate than $B \cup D$.

Partial Additivity[25]**:** If group A and group B are of the same size and the sum of the individual accuracy scores for the members of A is greater than the sum of the individual accuracy scores for the members of B, then A is more accurate than B.

Partial Dominance: If group A is at least as large as group B and every member of A has a higher individual accuracy score than every member of B, then A is more accurate than B.

Theorem 1 *No measure of group accuracy satisfies Nominal-Notable, Transitivity, No Repugnance, and No Simpson.*

Proof

1. Let A be a group with n members all with very high accuracy, h. Let Z be a group with $n + m$ members all with very low but positive accuracy, l. No Repugnance requires that A is not less accurate than Z.

2. Let A1 contain 1 member with accuracy h. Let Bk contain k individuals with accuracy level h- where $h > h$- in a way that satisfies Nominal-Notable. By Nominal-Notable, Bk is more accurate than A1.

---

[24] This principle is so-called because it involves a nominal change in average accuracy, from h to h-, but a notable change in total accuracy because of the larger group size. This is inspired by Ruth Chang's work on options that seem, at first glance, to be incomparable.

[25] I adapt this constraint from Pettigrew (2018).

3. By an application of No Simpson, Bk ∪ Bk (which I'll refer to as 'B2k') is more accurate than A1 ∪ A1 (which I'll refer to as 'A2').
4. By n applications of No Simpson, then, Bnk is more accurate than An = A.
5. Now, let B1 contain one individual with accuracy level h-. Let Ck contain k individuals with accuracy level h-- where h- > h-- in a way that satisfies Nominal-Notable. By Nominal-Notable, Ck is more accurate than B1.
6. By nk applications of No Simpson Cnkk is more accurate than Bnk.
7. By Transitivity from 4 and 6, Cnkk is more accurate than A.
8. By continued applications of this strategy we reach a point where some group Z has $n + m$ members each with accuracy level l such that $h > h$- > h-- > ... > l, such that Z is more accurate than A, which contradicts line 1.

Denying Transitivity is not a plausible way out of this problem, and, for reasons explained in earlier sections, neither is denying No Simpson. Nominal-Notable is extremely plausible, though it would be denied by defenders of Average Group Accuracy. This leaves No Repugnance as the only principle to deny, which is what Total Group Accuracy does by embracing the epistemic repugnant conclusion.

Theorem 2 *No group accuracy measure satisfies Transitivity, No Simpson, No Repugnance, Partial Additivity, and Partial Dominance.*

Proof Consider the following five groups:

A_n: n members each with high level of accuracy, h
B_n: n members each with negative level of accuracy, l
C_n: n members each with high level of accuracy, $h + \epsilon$
D_q: q members ($q > n$) each with negative level of accuracy, $l + \epsilon$
E_q: q members ($q > n$) each with very low positive accuracy, s

1. By Partial Additivity, $C_n > A_n$.
2. By Partial Dominance, $D_q > B_n$.
3. From 1, 2, and No Simpson, $C_n \cup D_q > A_n \cup B_n$.
4. Pick $n, q, l, h,$ and s such that:

$$nl + qs > n(h + \epsilon) + q(l + \epsilon)$$

By the Archimedean property of the reals, there are such numbers.[26] Notice that the sum of the personal accuracies of $B_n \cup E_q$ is $nl + qs$ and the sum of the personal accuracies of $C_n \cup D_q$ is $n(h + \epsilon) + q(l + \epsilon)$. Since $B_n \cup E_q$ and $C_n \cup D_q$ are the same size ($n + q$), by Partial Additivity, $B_n \cup E_q > C_n \cup D_q$.

---

[26] To see this, note that the inequality can be rewritten as $q(s$-$(l + \varepsilon)) > n(h + \varepsilon$-$l)$. The quantity s-$(l + \varepsilon)$ is positive since ($l + \varepsilon$) is negative and s is positive, and $n(h + \varepsilon$-$l)$ is positive since l is negative and all the rest are positive. Hence, this inequality is equivalent to $qx > y$ for positive reals x and y.

5. From 3, 4, and Transitivity, $B_n \cup E_q > A_n \cup B_n$.
6. By No Repugnance, $A_n \geq E_q$.
7. Since $B_n = B_n$, No Simpson and 6 imply that $A_n \cup B_n \geq B_n \cup E_q$, which contradicts line 5.

Total Group Accuracy gets out of this contradiction, again, by rejecting No Repugnance, and accepting the epistemic repugnant conclusion.

References

Ahlstrom-Vij, K., & Dunn, J. (Eds.). (2018). *Epistemic consequentialism*. Oxford University Press.

Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet functions for contraction and revision. *Journal of Symbolic Logic, 50*, 510–530.

Arrhenius, G. (2000). An Impossibility Theorem for Welfarist Axiologies. *Economics and Philosophy, 16*, 247–266.

Arrhenius, G. (2011). The impossibility of a satisfactory population ethics. In H. Colonius & E. Dxhafarov (Eds.), *Descriptive and normative approaches to human behavior, advanced series on mathematical psychology.* World Scientific Publishing Company.

Baral, C., Kraus, S., & Minker, J. (1991). Combining multiple knowledge bases. *IEEE Transactions on Knowledge and Data Engineering, 3/2*, 208–220.

Carr, J. (2015). Epistemic expansions. *Res Philosophica, 92*(2), 217–236.

Carr, J. (2017). Epistemic Utility Theory and the Aim of Belief. *Philosophy and Phenomenological Research, 95*, 511–534.

Chang, R. (1997). Introduction. In R. Chang (Ed.), *Incommensurability, incomparability and practical reason.* Harvard University Press.

Dunn, J. (2019). Accuracy, verisimilitude, and scoring rules. *Australasian Journal of Philosophy, 97*(1), 151–166.

Fallis, D., & Lewis, P. J. (2016). The brier rule is not a good measure of epistemic utility (and other useful facts about epistemic betterness). *Australasian Journal of Philosophy, 94*(3), 576–590.

Gärdenfors, P. (1988). *Knowledge in flux: Modeling the dynamics of epistemic states*. MIT Press.

Gibbard, A. (2008). Rational credence and the value of truth. In T. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology.* (Vol. 2, pp. 143–164). Oxford: Oxford University Press.

Goldman, A. (1999). *Knowledge in a social world*. Oxford University Press.

Greaves, H., & Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind, 115*(459), 607–632.

Greaves, H. (2013). Epistemic decision theory. *Mind, 122*, 915–952.

Heesen, R. & Bright, L. K. (2020). Is peer review a good idea?, *British Journal for the Philosophy of Science.* https://doi.org/10.1093/bjps/axz029.

Joyce, J. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science, 65*(4), 575–603.

Joyce, J. (2009). Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In F. Huber & C. Schmidt-Petri (Eds.), *Degrees of belief.* (pp. 263–297). Dordrecht: Springer.

Konek, J., & Levinstein, B. (2019). The foundations of epistemic decision theory. *Mind, 128*, 69–107.

Leitgeb, H., & Pettigrew, R. (2010a). An objective justification of Bayesianism I: Measuring inaccuracy. *Philosophy of Science, 77*(2), 201–235.

Leitgeb, H., & Pettigrew, R. (2010b). An objective justification of Bayesianism II: The consequences of minimizing inaccuracy. *Philosophy of Science, 77*(2), 236–272.

Levinstein, B. (2012). Leitgeb and Pettigrew on accuracy and updating. *Philosophy of Science, 79*, 413–424.

List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. University Press.

Mackie, J. L. (1985). Parfit's population paradox. In J. Mackie & P. Mackie (Eds.), *Persons and values.* Oxford University Press.

Ng, Y. (1989). What should we do about future generations: Impossibility of parfit's theory X. *Economics & Philosophy, 5*, 235–253.

Parfit, D. (1984). *Reasons and persons*. Oxford University Press.

Pérez Carballo, A. (2018). Good questions. In K. Ahlstrom-Vij & J. Dunn (Eds.), *Epistemic consequentialism.*Oxford University Press.

Pettigrew, R. (2018). The population ethics of belief. In search of an epistemic theory X. *Nous, 52*(2), 336–372.

Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford University Press.

Pigozzi, G. "Belief Merging and Judgment Aggregation", The Stanford Encyclopedia of Philosophy (Winter 2016 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2016/entries/belief-merging/.

Ryberg, J. (1996). Is the repugnant conclusion repugnant? *Philosophical Papers, 25*, 161–177.

Sider, T. (1991). Might theory X be a theory of diminishing marginal value? *Analysis, 51*, 265–271.

Talbot, B. (2019). Repugnant accuracy. *Nous, 53*(3), 540–563.

Tännsjö, T. (1992). Who are the beneficiaries? *Bioethics, 6*, 288–296.