Measuring Group Accuracy

Jeffrey Dunn

DRAFT of 10/4/2017

1. Introduction

Consequentialist approaches to epistemology have lately received increased attention.[1] According to such an approach, epistemic value comes first, and from this we derive our epistemic norms. One specific version of this approach, which we might call 'Epistemic Utility Theory', has been very influential within formal epistemology. Epistemic Utility Theory takes its lead from James Joyce's (1998) seminal work where he presents an argument for having probabilistically coherent degrees of belief based on the fact that incoherent degrees of belief are provably less accurate than coherent ones. According to Epistemic Utility Theory, accuracy is the only final epistemic value. We first score belief states based on their accuracy, and then figure out the epistemic norms by determining which belief-forming strategies are best to adopt with the goal of gaining as much accuracy as possible.[2]

So far, Epistemic Utility Theory has been deployed with an eye towards norms for individual agents. So, for instance, we get arguments attempting to show that an agent's degrees of belief should be probabilistically coherent, or that an agent should update her belief state according to conditionalization. But the structure of Epistemic Utility Theory makes it particularly amenable to certain questions in social epistemology. If at the individual level, we seek to maximize accuracy for a person, the Epistemic Utility Theorist seems compelled to say that at the group level, we seek to maximize accuracy for a group. Further, Epistemic Utility Theory has a structure analogous to decision theory. In decision theory, we attempt to determine what actions I should perform by determining what will (or what can be expected to) maximize my personal welfare. It's not a far step from this to some version of utilitarianism where we ask what actions are morally best by determining what will (or what can be expected to) maximize the welfare of those in one's group, one's society, or all those in existence. Just as decision theory leads naturally to this kind of utilitarianism, so it is natural for the Epistemic Utility Theorist to turn from the individual to the group and consider what social structures and institutions are epistemically best by determining what will (or what can be expected to) maximize the accuracy of a group, a society, or all those in existence.

In addition, this kind of approach to social epistemology has a precursor outside of Epistemic Utility Theory. Alvin Goldman (1999), in his influential book on social epistemology explicitly endorses veritism--the view that true beliefs and only true beliefs have final epistemic value and that false beliefs and only false

[1] See, for instance, Berker (2013a,b), Ahlstrom-Vij & Dunn (2014), and Goldman (2015).
[2] For representative work in this area see Greaves & Wallace (2006), Gibbard (2008), Joyce (2009), Leitgeb & Pettigrew (2010a,b), and Pettigrew (2016).

beliefs have final epistemic disvalue. He then epistemically evaluates social practices and rules in terms of their promotion of this veritistic value. As Goldman says, social practices are to be evaluated by how well they raise 'the aggregate level of [true belief] of an entire community' (93).

So there is good reason to be interested in a kind of consequentialist epistemic assessment of social practices and institutions. But for a project like this to succeed, there must be something that corresponds to the accuracy of a group as a whole. If there is no such thing, this approach will fail. In this paper, drawing inspiration from work in population ethics, I prove an impossibility theorem for any measure of group accuracy and then argue for a particular response to this theorem.

2. Group Accuracy: Two Proposals

There are several things one might mean by "group accuracy". We might want to know how accurate a group would be about a topic, were they to pool their information with each other. Alternatively, we might somehow construct a group belief state out of the beliefs of the members and then evaluate the group belief state for accuracy.[3] But if our approach to social epistemology is to be structurally similar to utilitarianism in ethics and if we are to follow Goldman's veritism, then we don't want either of these as our measure of group accuracy.[4] Rather, what we want is a function that takes as input the accuracy of each member of the group and gives us an accuracy for the group as a whole. This is clearly analogous to the central question in population ethics, which concerns the proper way to determine the total welfare of a population given the welfare of the individuals that make up that population.[5] I'll assume, then, that we have some way of measuring the accuracy for an individual, what I'll call a *personal accuracy score*. Our task is to parlay this into a measure of group accuracy.[6]

There are two 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 personal accuracy scores for each member of X.

The second natural proposal is:

---

[3] For early work on how to construct a group belief state out of the belief states of the members, see List & Pettit (2011). For more recent work on how to measure the accuracy of such states, see Pettigrew (*forthcoming a*).

[4] Which isn't, of course, to say that such measures are uninteresting. They're just not what we want for *this* project.

[5] The classic work in this area is Parfit's (1984) *Reasons and Persons*.

[6] If we represent individuals as having all-or-nothing beliefs, personal accuracy might be the sum of the true beliefs minus the false beliefs, or it might be the sum of the true beliefs divided by the total number of beliefs. If we represent individuals as having degrees of belief, then this might be the sum of local proper score of each particular degree of belief in a proposition, or it might be the sum of local proper score of each particular degree of belief in a proposition divided by the total number of propositions to which the agent has degrees of belief.

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

As is well-known, the Total and Average proposals will give the same rankings when the number of group members is the same between all groups being ranked.[7] But they can differ in their rankings when comparing groups of unequal size. And it is in considering such cases that both proposals seem to run into counterintuitive verdicts.

Start with Total Group Accuracy. It runs into an epistemic version of Parfit's (1984) "repugnant conclusion". Suppose group A has *n* members all with very high personal accuracy. Let group B have *n+m* members where all *n+m* members of B have just slightly greater than 0 personal accuracy. With sufficiently large *m*, group B will be ranked as more accurate than group A. But this may seem implausible. Every member of A, we can suppose, is perfectly accurate about every proposition in their belief set. Every member of B, we can suppose, is barely more accurate than inaccurate (working with all-or-nothing beliefs, perhaps each member of B has just one more true than false belief). If we countenance negative individual accuracy scores--perhaps had by someone with more false beliefs than true beliefs--we get something even worse. In that case, Total Group Accuracy leads to a version of what has been called (Arrhenius, 2011) the "very repugnant conclusion". Let group A be as before, but now suppose that *n* members of B have very negative accuracy scores. Perhaps every one of their beliefs is false. If the *m* members of B have just slightly greater than 0 accuracy, then with sufficiently large *m*, group B will be ranked as more accurate than group A.

In a related discussion, Richard Pettigrew (*forthcoming b*) defends something similar to Total Group Accuracy. However, Pettigrew is not arguing for a measure of *group* accuracy, but instead a measure of *personal* accuracy for credence functions. He opts for a measure of accuracy that scores credences in individual propositions and then sums up those individual scores to get the personal accuracy score for the total credence function. The resulting view, which he calls 'Total Epistemic Utilitarianism', is structurally similar to Total Group Accuracy. In response to repugnant-conclusion-like consequences of such a view, Pettigrew offers a debunking argument. He admits that we intuitively think that a credence function with 1 million very accurate credences is doing 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. Suppose that Phoebe has the first sort of credence function and Daphne has the second sort. Pettigrew writes:

> For a huge number of propositions, Daphne will go wrong in very many decisions that turn on her attitudes to those propositions. And for many others, she will not go right much more often than if she were simply to toss a coin. Phoebe, in contrast, will go right very often when the decision turns

[7] This is because for $n \geq 1$, $x/n > y/n$ iff $x > y$.

on a proposition to which she assigns a credence. For these reasons, I submit, we judge that Phoebe's credences are better than Daphne's in terms of their pragmatic utility. (p. 29)

Can a similar debunking response be made on behalf of Total Group Accuracy in the face of its repugnant conclusion? An analogous reply would maintain that we judge Group A to be more accurate than Group B, because we are confusing epistemic and pragmatic evaluation. Group A, with its highly accurate members, the reply goes, will go right very often when a group decision turns on propositions to which the group assigns credence. Group B, on the other hand, will not do so well when faced with a group decision.

Even if one finds Pettigrew's debunking response compelling in the case of a single credence function, it fails in the group case. Why? Because when it comes to social epistemology, we often are evaluating groups for accuracy when they are not the sort of cohesive group that is in the position to make group decisions or 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 here. There simply aren't any group actions--whether actual or counterfactual--that are candidates for pragmatic evaluation. So I conclude that the repugnant conclusion (and related worries) are a significant problem for Total Group Accuracy.

Average Group Accuracy does better with respect to the repugnant conclusion. For you cannot raise the average by adding many members with just slightly positive personal accuracy scores; to raise the average accuracy of a group you must add individuals with higher personal accuracy than the current group average. But Average Group Accuracy has its own set of counterintuitive consequences. First, consider group A, which consists of one person, with a very negative personal accuracy score. Compare this to group B, which consists of 100 people. 99 of these members have the same very negative personal accuracy scores and 1 member has a negative personal accuracy score, but just slightly higher than the 99. Average Group Accuracy says that group B is more accurate than group A. But this may seem wrong. Although there is, in a sense, more accuracy in group B than group A (after all, that one person in B is doing better than everyone in A), this might seem counterbalanced by all the inaccuracy in group B. Average Group Accuracy misses this. Here is another concern with this proposal. Suppose Group A and Group B initially have the same number of members, and every member of B, b_i, can be paired with someone in A, a_i, such that a_i is more accurate than b_i (and not vice versa). Let v be the average individual accuracy of members of B and assume this level is quite high. Suppose, now that we add m group members to Group A who have individual accuracy, v-, just slightly below v. At some level of m, Group A will become less accurate than Group B. But this may seem incorrect. For Group A was initially more accurate than Group B, and it seems that adding more members to A that are all themselves highly accurate (recall that *v*- may be a very high level of accuracy), somehow makes Group A less accurate than Group B.

So, two very natural and initially plausible accounts of group accuracy turn out not to have some counterintuitive consequences. In some sense this shouldn't be too surprising given work in population ethics showing that many plausible measures of group welfare have counterintuitive consequences. In fact, with inspiration from that literature, we can prove an impossibility theorem about measures of group accuracy. This will give us a sense of the possible moves.

3. Impossibility

I'll consider five independently plausible constraints on a measure of group accuracy and then prove that they are nevertheless jointly inconsistent. Before giving the constraints, a word about notation: capital letters refer to groups, and subscripts tell us the number of members. So, 'A_n', refers to the *n*-membered group A. I'll use '$A_n > B_m$' to say that the *n*-membered group A is more accurate than the *m*-membered group B. Here, then, are the constraints[8]:

Transitivity: For any three groups, A, B, and C, if $A \geq B \geq C$, then $A \geq C$ (with strict inequality if one of the initial inequalities is strict).

This is an extremely plausible constraint when we are considering the measurement of one particular quantity, in this case, accuracy. If accuracy is not transitive in this way, it is not clear what it means to say that we ought to maximize (expected) accuracy, so consequentialists have a strong reason to accept this constraint.

Partial Additivity[9]: If the sum of the personal accuracy scores for the members of A_n is greater than the sum of the personal accuracy scores for the members of B_n, then $A_n > B_n$.

Notice that Partial Additivity is restricted to comparisons between groups of equal size. In this case, both Average and Total Group Accuracy entail Partial Additivity. Note, further, that Partial Additivity prohibits certain kinds of spurious measures of group accuracy, such as measures that say the accuracy of the group is simply equal to the personal accuracy of some one member. Such measures are spurious because they aren't considering the belief states of all the group members.

[8] It is perhaps worth noting that the impossibility theorem I will prove bears a resemblance to similar impossibility results in population ethics (for example Arrhenius (2011), Arrhenius (2000)), and an impossibility result in epistemology by Pettigrew (*forthcoming b*). Nevertheless, though the approach is similar--show that some set of constraints are not mutually satisfiable--the set of constraints I list here is novel as is the theorem.

[9] I adapt this constraint from Pettigrew (*forthcoming b*).

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

Partial Dominance would be controversial if we didn't include the stipulation that group A is at least as large as group B. For in that case, Total Group Accuracy would deny it: there could be cases where a small group of very accurate individuals possesses less total accuracy than a larger group of slightly less accurate individuals. But with the restriction, this is again a constraint to which both Average and Total Group Accuracy are committed. And it is plausible in its own right.

The final two constraints are slightly more complex. The first, is:

No Simpson: If $A_n \geq A_m$ and $B_j \geq B_k$ then $A_n \cup B_j \geq A_m \cup B_k$ (with strict inequality if both $A_n > A_m$ and $B_j > B_k$).

Here's a way to think about what No Simpson says. Suppose you have two groups and you can break each group into subsets such that each subset of the first group can be paired off uniquely with one subset of the second that it is more accurate than. Then, the first group is more accurate than the second. This is a pretty plausible constraint. It implies, for instance, that if each subset of a group gets more accurate over time, then the group as a whole gets more accurate over time. Nevertheless, it is violated by Average Group Accuracy.[10]

Finally, our last constraint:

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 *s* or below, then $A \geq B$.

This constraint, as the name suggests, is meant to rule out an epistemic version of the repugnant conclusion. It rules out a very large group where each member is minimally accurate being more accurate than a smaller group where each member is very accurate. Note two things about No Repugnance. First, it rules out Total Group Accuracy, since the sum of many low positive numbers can always exceed the sum of fewer

[10] That Average Group Accuracy violates No Simpson is an instance of what is called Simpson's Paradox (hence the name of the constraint). 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 much higher rate of acceptance for men. Another, slightly simpler, illustration of the paradox can be seen with batting averages in baseball (Ken Ross, 2004). 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 2-year period from '95-'96, Jeter batted .310 while Justice batted .270. This happened because most of Jeter's at-bats were in the '96 season when he batted .314, whereas most of Justice's at-bats were in the '95 season when he batted .253. Here is the total breakdown:

	1995	1996	Combined
Jeter:	12/48 (.250)	183/582 (.314)	195/630 (**.310**)
Justice:	104/411 (**.253**)	45/140 (**.321**)	149/551 (.270)

and yet higher positive numbers. Second, No Repugnance is similar to Partial Dominance, but distinct. Partial Dominance says that if the larger group is universally more accurate than the smaller, then the larger group has a higher accuracy. No Repugnance says that in certain situations the smaller group is more accurate; namely, when the smaller group contains members all with very high accuracy and the larger group contains members with barely positive accuracy.

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

To establish this, 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, $h + \varepsilon$

D_q : *q* members ($q > n$) each with negative level of accuracy, $l + \varepsilon$

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:

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

By the Archimedean property of the reals, there are such numbers.[11] 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 $n(h + \varepsilon) + q(l + \varepsilon)$. Since $B_n \cup E_q$ and $C_n \cup D_q$ are the same size ($n + q$), by Partial Additivity, $B_n \cup E_q > C_n \cup D_q$.

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. *QED*

4. Response to the Theorem

[11] 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*.

It is hard to see any plausible way to give up Transitivity, Partial Additivity, or Partial Dominance. That leaves the options of rejecting No Repugnance or No Simpson. If we give up No Repugnance, and hence accept the repugnant conclusion, it opens up the possibility of Total Group Accuracy. If we give up No Simpson, and hence Simpson-paradoxical scenarios, it opens up the possibility of Average Group Accuracy.

I've already given some initial arguments against Average Group Accuracy, but there is a further reason not to accept this way out of the impossibility theorem. Average Group Accuracy ends up yielding odd results when we pair it with either an average-based approach to personal accuracy or a total-based approach to personal accuracy. To see this, consider the following. Start with group A. Add a member, *j*, to group A who has a set of beliefs **F**, and where *j*'s personal accuracy (the average accuracy of each of her beliefs) is greater than the group average, but less than individual *i* who is in A. According to Average Group Accuracy, group A + *j* is more accurate than A, since the addition of *j* increases the average accuracy. But now consider group A again. Instead of adding a new member to A, take the existing member, *i*, and let *i* acquire exactly the set of beliefs **F** that *j* had. Since *j*'s personal accuracy was lower than *i*'s personal accuracy, this addition to *i* makes *i* less accurate. So, the accuracy of this group is lower than A. But this is implausible. We've added the same set of beliefs, **F**, to group A and gotten different results.

Suppose now that we adopt Total Personal Accuracy, which says that a personal accuracy score is the sum of the accuracy score for each proposition a person believes. Start with group A. Add a member, *j*, to group A who has a set of beliefs **F**, and where *j*'s personal accuracy (total) is lower than everyone in the group but still positive. Given average group accuracy, the accuracy of the group A + *j* is less than the accuracy of A. But now, instead of adding a new member, take an existing member, *i*, and let *i* acquire exactly the set of beliefs **F** that *j* had. Since *j*'s personal accuracy was positive, this addition to *i* makes *i* more accurate. So, the accuracy of this group is greater than A. I think these results give us good reason to reject Average Group Accuracy as *the* measure of group accuracy.

The natural option, then, is to endorse Total Group Accuracy. Total Group Accuracy rejects No Repugnance, and thus gets us out of the impossibility theorem. But I find the repugnant conclusion as repugnant in epistemology as in population ethics. And further, I think there is a better way forward.

We can see this better way by noting that both No Repugnance and No Simpson tacitly assume that every group is comparable with every other group, including comparisons between groups of different sizes. If we reject that some groups are comparable in this way, we could have a group accuracy measure, which doesn't land us with the counterintuitive results we get with both Total and Average Group Accuracy, and which escapes the purview of the theorem.

The simplest way to pursue this idea is to deny that groups of different sizes can be compared for accuracy. Since we know that Total and Average Group Accuracy agree in these cases, and since the controversial verdicts concern comparisons between different-sized groups, this may seem like an attractive option. However, such a view denies comparisons that we should be able to make. Suppose we have a group with 3 members, where all three members have very high personal accuracy. Suppose we compare this to a group with 2 members, where both members have very low personal accuracy. It seems clear that the 3-membered group is more accurate than the 2-membered group, but since they are of different sizes, the simple proposals says they cannot be compared.

Note, however, that in this case, both Total and Average Group Accuracy agree that the 3-membered group is more accurate than the 2-membered group. This suggests the following proposal:

> **Two-Component Group Accuracy (TCGA):** Let a_X be the Average Group Accuracy of Group X and let t_X be the Total Group Accuracy of Group X. For any groups, A and B, If $a_A \geq a_B$ and $t_A \geq t_B$, then $A \geq B$ (with strict equality if either of the inequalities in the antecedent are strict); otherwise, A and B are not comparable in terms of accuracy.

This gets the comparison between the 3-membered and 2-membered group correct. The idea behind Two-Component Group Accuracy is that while accuracy is the sole epistemic value, there are two ways to value it. When a group increases its total amount of accuracy, this is a kind of improvement with respect to accuracy. In Goldman's terms, it is to raise the aggregate level of true belief of an entire community. But similarly, when a group increases its average amount of accuracy, this is another, different kind of improvement with respect to accuracy. It, too, is to raise the aggregate level of true belief of an entire community. But, it turns out that these two ways of improving with respect to accuracy, though related, can sometimes come apart from each other. And when they do, according to TCGA, there is no epistemic way to evaluate trade-offs between an improvement in one of these and a diminishment in the other.[12]

For instance, consider the following two groups:

Group 1				
	Member 1	Member 2	Member 3	
True Beliefs	100	100	100	
False Beliefs	25	25	25	
Group 2				
	Member 1	Member 2	Member 3	Member 4
True Beliefs	90	90	90	90
False Beliefs	25	25	25	25

[12] This leaves open that there might be pragmatic or moral reasons to prefer one sort of improvement to another.

Which group is more accurate here? It is not at all obvious. Group 2 has overall more true belief, and in addition has a greater difference between true and false belief. But, on the other hand, Group 1 has a better ratio of true to false belief. It seems to me correct to say that in this case, Groups 1 and 2 cannot be compared in terms of accuracy. In one respect Group 1 is better; in another respect Group 2 is better.

There are three key questions with respect to TCGA that I'd like to address. First, I want to consider a disanalogy between social epistemology and population ethics, which helps to explain why it is not problematic to have groups that are not comparable in terms of accuracy, although it is problematic to have populations that are not comparable in terms of welfare. Second, I want to consider more carefully how the TCGA responds to the impossibility theorem. Finally, and importantly, I'd like to consider whether it makes sense to say that this is a purely veritistic measure--that is, that accuracy is the sole epistemic value--if both the average and the sum of a group's accuracy are doing important work in ranking groups.

4.1 Disanalogy to Population Ethics

It seems that one important ethical question concerns the existence of future persons, in the sense that it is at least sometimes (and perhaps always) an ethical matter whether a person is brought into existence or not. Given this, no plausible ethical theory can remain silent when it comes to decisions about whether a future person is to exist or not. But, if one is an ethical consequentialist, the right verdict about whether a future person should exist or not depends on comparing (future) populations with different people almost certainly of different sizes. So it just isn't an option for the ethical consequentialist to say that there are certain populations of different sizes that we cannot compare. To do so is to be silent about an important ethical question.

But in social epistemology, things are different. Note, for one, that while it is an ethical matter whether to increase or decrease the size of the population, it is not obviously an epistemic matter whether to increase or decrease the size of a group. So, at least in one way, there is a disanalogy to population ethics.

That said, we still may want to compare groups of different sizes in terms of their level of accuracy. For instance, suppose we have two possible educational interventions that can be done among high school students in a particular school. One intervention will target 50 students; the other will target 100 students. In evaluating these two interventions, we will want to compare the accuracy of the students in both cases. But this seems to require us to compare a group of 50 students to a group of 100 students, and opens up the possibility that they are incomparable because the average and total scores will point in opposite directions. There may be difficult cases here, but note that there are still plausible ways to make the comparison. The comparison, for instance, could be made by comparing the group of 50 students who have the first intervention *plus* 50 more who do not have the intervention, to the group of 100 students who have the second intervention. Or, somewhat more naturally, we could compare the accuracy of the total set of students at the school under both interventions. Since when we have the same number of members in each group, the total and average scores are guaranteed to agree, we can reach a verdict. This is not the sort of thing one can do in population ethics, however. If a given population has, say, 50 people in it, we cannot sensibly add more members to it to compare it to a population of 100. So, there is a further disanalogy between population ethics and group accuracy.

4.2 The Two Component Response to the Impossibility Theorem

TCGA satisfies Transitivity, Partial Additivity, and Partial Dominance. What it rejects are both No Simpson and No Repugnance. But it rejects these constraints in a very important way. It says that No Repugnance is false, not because there is some A where all members have very high accuracy over *h*, and some B with more members than A, where all the members of B have very low but positive accuracy *s* or below, and yet $A < B$. Rather, it says that when this happens, it is is not the case that $A \geq B$ because in fact these groups are not comparable. In such a case, though group B may have more total accuracy, group A will have more average accuracy. The Two Component proposal says something similar in its rejection of No Simpson. We don't get Simpson-paradoxical cases, rather, we get cases where the two groups are not comparable.

One might think this is a stronger response to the theorem than required. After all, to get out of the impossibility results requires rejection of only one of the constraints. However since TCGA implies Partial Additivity, Partial Dominance, and Transitivity the theorem tells us that if we also adopt No Simpson, then we are stuck with the claim that $B_n \cup E_q > A_n \cup B_n$. Recall that E_q is a group with many individuals of all very low (but positive accuracy). A_n is a group with less members all that are highly accurate. The claim, then that $B_n \cup E_q > A_n \cup B_n$ strikes me as bad as the repugnant conclusion.

If, on the other hand, one embraces No Repugnance together with TCGA, then the theorem tells us that we must reject No Simpson. We can reject it either because in Simpson-paradoxical cases the groups are not comparable, or because in such cases, we get Simpson's Paradox. But once you have embraced No Repugnance and so embraced a variety of comparisons between groups of different sizes, it is hard to motivate the claim that in Simpson-paradoxical cases the groups are not comparable. So, you will get Simpson-paradoxical scenarios, that is, situations where all the subsets of one group are more accurate than the subsets of the second group and yet the second group is more accurate. This, however, strikes me as intolerable.

4.3 Value Monism?

I've argued that the best way to measure group accuracy is with TCGA. According to this proposal, there are cases where groups are not able to be compared. These are cases where one group has a higher Total Group Accuracy score and the other group has a higher Average Group Accuracy score. One might argue that this shows that this is not a view according to which accuracy is the sole epistemic value. For, the objection continues, if there is only one kind of value, then there are never cases of incomparability.

In response, first note that Ruth Chang (1997, pp. 16-7) has argued that value monism does not imply that all outcomes are comparable. Chang points out that something can have more amount of value than something else, and yet not necessarily be better with respect to that value. For instance, someone can possess more friendliness than another, and yet not be a better friend; it is possible to be too friendly. Given this, it is possible that different quantities of some single value are not comparable. So the move from incomparability between options to pluralism about value is too quick.

Further, I propose that though total accuracy and average accuracy are important epistemic values, they are not distinct final epistemic values. Rather, they represent two different ways to value the same one final value. That is, what epistemically matters is accuracy; but both its amount and its distribution. When the total accuracy of a group goes up we have more total amount accuracy than before, but if the average accuracy simultaneously goes down, this group is not epistemically better. So, is the group epistemically worse? No, because total accuracy goes up. Similarly, when total accuracy goes down we have less total amount accuracy than before, but if the average accuracy goes up, this is not a worse state of affairs. Is it a better state of affairs? No, because total accuracy goes down.

Consider an analogy. We can rank basketball players both in terms of the total points they score per game as well as the average points they score per shot taken. These are both important ways in which to evaluate players, but the fact that we care about both doesn't show that there are two kinds of values with respect to offensive play in basketball. There is just one value: points, which can be valued in different ways. Accuracy, I say, is similar.

5. Conclusion and Future Work

There are open questions and future work related to TCGA. First, supposing that TCGA is a good measure of group accuracy, it seems to have implications for measures of personal accuracy. For when it comes to an individual's belief state, we can both take the total accuracy in each proposition believed, or we can take the average accuracy across all propositions believed. The arguments here may tell in favor of a similar Two Component proposal at the level of personal accuracy.

Second, the Two Component proposal will face cases that Chang (1997) calls 'nominal-notable comparisons'. These are cases where there is a huge difference in one value and only a nominal difference in the other one. For instance, suppose we have a group and there is change to this group such that the Average Group Accuracy score decreases a small amount in exchange for a very big gain in the Total Group Accuracy score. Such cases can happen if we add many members to the group who have personal accuracy scores that are high, but slightly below the group average. TCGA says that we cannot compare the group before and after the change, but this might seem implausible. These are important cases to consider, though as a partial response we can note that if we accept comparability here, this is exactly the kind of initial step that eventually gets us to the repugnant conclusion.

Finally, if we embrace a purely veritistic, consequentialist, social epistemology that uses TCGA as its measure of accuracy, we will (of course) get certain cases where options are incomparable. This raises a question about how choice between these options are to be determined. That is, suppose we are facing a decision where the outcome will either be Group 1 or Group 2 from the table in section 4. Since these groups are not comparable in terms of accuracy, it is unclear what the social epistemologist should recommend. Are both options permissible? Neither? Is there some way to nevertheless determine which decision is correct? These are important questions to address, about the relation between our deontic assessments among options some of which are incomparable to each other..

Despite this further work, I find TCGA to be a promising way to measure group accuracy in light of the impossibility theorem from section 3. TCGA opens up the possibility of a purely veritistic, consequentialist, social epistemology. Some may be more pessimistic, however, and use the theorem to show that no measure of group accuracy is plausible and so to doom the entire project of a purely veritistic, consequentialist, social epistemology. I'm more hopeful, but the theorem sets the parameters for such a debate.

Works Cited

Ahlstrom-Vij, K. & Dunn, J. (2014). "A Defence of Epistemic Consequentialism," Philosophical Quarterly, 64(257): 541--551.

Arrhenius, G. (2000). "An Impossibility Theorem for Welfarist Axiology", Economics and Philosophy, 16: 247--266.

Arrhenius, G. (2011). "The Impossibility of a Satisfactory Population Ethics" in H. Colonius & E. Dzhafarov (Eds.) Descriptive and Normative Approaches to Human Behavior, Advanced Series on Mathematical Psychology, World Scientific Publishing Company.

Berker, S. (2013a). "Epistemic Teleology and the Separateness of Propositions," Philosophical Review, 122(3): 337--393.

Berker, S. (2013b). "The Rejection of Epistemic Consequentialism," Philosophical Issues, 23: 363--387.

Change, R. (1997). "Introduction", in R. Chang (ed.), *Incommensurability, Incomparability, and Practical Reasoning*, Harvard University Press.

Gibbard, A. (2008). "Rational Credence and the Value of Truth," in T. Gendler & J. Hawthorne (Eds.), Oxford Studies in Epistemology, vol. 2, Oxford University Press.

Goldman, A. (1999). Knowledge in a Social World. Oxford University Press.

Goldman, A. (2015). 'Reliabilism, Veritism, and Epistemic Consequentialism'. Episteme, 12(2): 131--143.

Greaves, H. & Wallace, D. (2006). "Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility," Mind, 115(459): 607--632.

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 , Springer.

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.

List, C. & Pettit, P. (2011). Group Agency: The Possibility, Design, and Status of Corporate Agents. Oxford University Press.

Parfit, D. (1984). Reasons and Persons. Oxford: Clarendon Press.

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

Pettigrew, R. (forthcoming a). "On the Accuracy of Group Credences," in T. Szabó Gendler & J. Hawthorne (Eds.) Oxford Studies in Epistemology, vol. 6, Oxford University Press.

Pettigrew, R. (forthcoming b). "The Population Ethics of Belief: In Search of an Epistemic Theory X," Nous.