Epistemology in *Group Agency*: 6 Objections in Search of the Truth*

Fabrizio Cariani

Northwestern University

Group Agency's main foray into epistemology starts with a few simple observations, which we can roughly summarize as follows (the following are not quotations):

(1) Groups can (and often have to) form propositional attitudes. In particular, groups form binary truth-directed attitudes. Call these 'judgments', or 'opinions' or 'acceptance states' (but refrain from calling them 'beliefs', because they have a volitional component that beliefs may lack).

(2) For any agent capable of forming such attitudes, one can evaluate that agent's devices for forming those judgments.

(3) Epistemology provides important dimensions of evaluation for judgment-forming devices.

(4) For groups, the judgment-forming mechanism involves centrally the group's organization, in particular, the sorts of *aggregation rules* the group decides (or is forced) to adopt. Here, an aggregation rule is a function that inputs the individual judgments of the group members and outputs a group judgment.

With these raw materials, List and Pettit (henceforth L&P) carry out a two-part program:[1] first, they sketch a manifesto for how aggregation rules should be assessed epistemically; second, they deduce more specific theses about how groups could arrange themselves to maximize the epistemic quality of their output. I am quite favorable to the global manifesto, but I find myself in disagreement with several of the specific theses and recommendations. This paper details my disagreements and their basis. Before getting to my criticisms, l will reconstruct the key elements of L&P's analytic set-up.[2]

* Special thanks to Mike Titelbaum for detailed comments on a late draft of the paper. Thanks to Ezra Cook, Franz Dietrich, Branden Fitelson, Sandy Goldberg, Jennifer Lackey and to the audience at the 2011 Midwest Epistemology Workshop for conversations that have helped me shape my arguments.

[1] Here, I mostly focus on their chapter 4.

[2] This necessarily involves some repetition of their exposition.

1 The Condorcet Jury Theorem.

We might call L&P's evaluation framework *Bayesian epistemology with a reliabilist twist*. The central measures they invoke are probabilistic analogues of Nozick's tracking conditions.[3] Consider a judgment-forming device δ; let $A_\delta(\varphi)$ mean that φ is accepted by δ. In the following, I omit the subscript, since its intended reference (if it has any) can generally be reconstructed on the basis of the context. Our first pair of Bayesian measures are:

Positive Tracking Reliability	$PTR[\varphi]$	$Pr[A(\varphi) \mid \varphi]$
Negative Tracking Reliability	$NTR[\varphi]$	$Pr[\sim A(\varphi) \mid \sim\varphi]$

These measures have a natural statistical interpretation. PTR is the probability of true positives. NTR is the probability of avoidance of false positives. For judgmental tasks, PTR is also called the *competence* of a judge (we might call NTR the *negative competence*). It is a bit unclear whether L&P intend to settle what interpretations these probabilities should receive (but most of their claims stand or fall independently of the interpretation).

In addition to tracking reliability, L&P introduce *indicating reliability*—a measure that flows in the opposite direction.

Positive Indicating Reliability	$PIR[\varphi]$	$Pr[\varphi \mid A(\varphi)]$
Negative Indicating Reliability	$NIR[\varphi]$	$Pr[\sim\varphi \mid \sim A(\varphi)]$

L&P make three points concerning indicating reliability. First, it is naturally interpreted as capturing the degree of belief that it is rational for an external observer to assign to φ upon observing that that some relevant (individual or collective) agent accepts φ. Second, Bayes Theorem establishes a connection between tracking and indicating reliability. Third, it is often easier to estimate tracking reliability than it is to estimate indicating reliability, so Bayes theorem is usually invoked to recover the latter from the former (together with the prior probabilities).

With these measures, we can score aggregation rules on a few dimensions. Abstractly, aggregation rules are functions that map individual judgments to collective judgments. The key epistemic question is: how can a group composed of reliable judges get the most epistemic benefit out of the reliability of its members? In other words: if the group members have good tracking reliability, what is the best aggregation rule in terms of 'channeling' their reliability into a collective judgment? In a suitably simplified environment, there is a case to be made for aggregating via some form of majority rule.

The simple, and at once stunning, result that sustains this case is the Condorcet Jury Theorem (Condorcet, 1785). The simplification I alluded to consists of four assumptions:

(i) Only one proposition is at stake—say, whether φ is true or false.

[3] Nozick (1981). See Roush (2006) for a modern (Bayesian) version of the tracking theory.

(ii) Individuals form their opinions independently.[4]

(iii) All group members are *competent* towards φ (i.e., their PTR's and NTR's are all above .5).

(iv) All group members are *equally competent* towards φ. Any two group members have identical PTRs, r^+ and identical NTRs, r^-.

With these assumptions in hand, the landmark result is:

> **Condorcet Jury Theorem (CJT)**
> Given (i)-(iv), forming opinion by majority has:
>
> (a) PTR (resp. NTR) greater than r^+ (resp. r^-)—and hence exceeding the competence of each group member.
>
> (b) PTR and NTR approaching 1 (maximum) as group size increases.

The epistemic upshot of CJT is roughly as follows:

> When (i)-(iv) are satisfied, a more democratized aggregation rule, such as majority, performs better than less democratized ones, such as dictatorships.

This conditionalized formulation reveals the limits of CJT. Its assumptions are rarely, if ever, all satisfied (Dietrich, 2008). In light of this, an important theoretical project is to understand how much we can weaken the assumptions while still preserving some of the epistemic punch of CJT.

The independence assumption is the subject of an extensive literature.[5] Since any satisfactory treatment of independence would be beyond the scope of this discussion, I will accept it as an idealization and hold it as a fixed point. I will instead focus on L&P's proposals for how to relax the *other* assumptions.

If we abandon the equal competence assumption, one the key claims of CJT fails. Consider a group of competent judges with distinct reliabilities. In such a group, the reliability of the majority judgment is not guaranteed to exceed the reliability of *every* individual.[6] If some one judge has much higher reliability than the others, the group may be better off deferring to the most reliable judge.

[4]Two judges, say Liz and Joe, form their opinion independently iff

$$Pr[A_{Liz}(\varphi) \,\&\, A_{Joe}(\varphi) \mid \varphi] = Pr[A_{Liz}(\varphi) \mid \varphi] \cdot Pr[A_{Joe}(\varphi) \mid \varphi]$$

Informally, this means that, conditional on the state of the world with respect to φ, the probabilities that each judge accepts φ are independent.

[5]See, among many others Berg (1996); Estlund (1994); Hawthorne (ms.); Dietrich and List (2004); Dietrich (2008); Dietrich and Spiekermann (ms.).

[6]Majority is however guaranteed to exceed the *average* of the individual competence levels. See Grofman, Owen and Feld (1983).

Given this observation, L&P make a majoritarian proposal (based on results reported in Grofman, Owen and Feld (1983), Shapley and Grofman (1984)): a *weighted* form of majority can still perform better than any individual in the group. The weights are specified as a function of the PTR.[7] This sort of weighted majority can be understood as a generalization of the majority rule in the case of (possibly) variable competence: if all the judges have the same competence, they will receive the same weight, and the rule will simply coincide with majority.

2 Complex Judgment Aggregation

Much of the novelty of the project that converges into *Group Agency* stems from relaxing assumption (i). A complex Judgment Aggregation problem arises when a group is deliberating about multiple logically connected judgments. L&P work with this example:

> [...] the group could be a university committee deciding on whether a junior academic should be given tenure with three relevant propositions involved: first, the candidate is excellent at teaching (T), second, the candidate is excellent at research (R), and third, the candidate should be given tenure (Y) where excellence at both teaching and research is necessary and sufficient for tenure (p.92).

Such a group needs individuals to submit binary opinions on three propositions T, R and Y, linked by the rule $(T \ \& \ R) \equiv Y$. The set of relevant propositions is normally called the *agenda*.

As we learned from L&P's prior work, complex Judgment Aggregation problems are like kryptonite for the majority rule. Even if each individual submits consistent opinions, we may have an inconsistent majority at the group level:

	T	R	$T \ \& \ R$
Judge 1	Y	Y	Y
Judge 2	Y	N	N
Judge 3	N	Y	N

Applying the majority rule to each proposition individually would recommend accepting: $\{T, R, \sim Y\}$ which contradicts the rule $(T \ \& \ R) \equiv Y$.

As List (2005a) remarks, the Majority rule satisfies the 'Knowledge challenge': it does pretty well at producing acceptance states with high tracking reliability. However, it fails the 'Rationality challenge': it fails to produce logically consistent outputs, even when the individual judges are consistent. This result raises the question:

> Among the aggregation rules that do meet the Rationality challenge, which ones are epistemically valuable?

[7] For a proposition φ, the weights are proportional to: $Log[PTR[\varphi]/(1 - PTR[\varphi])]$.

Before trying to answer this question, let's take a look some examples of aggregation rules that meet the Rationality challenge. At the extremes, we have dictatorships (in which the group's reliability is identical to the reliability of the dictator) and the (proposition-wise) Unanimity rule, according to which the group accepts φ if and only if it is accepted by every member of the group. Between these extremes, there are rules that are closer to majority.

- *Premise-Based Majority (aka PB):* Designate some propositions as "premises" (in the tenure example the two propositions concerning teaching and research); take majority on the premises and propagate by entailment on the conclusion. If the premises were chosen appropriately, the collective verdict is guaranteed to be consistent.

- *Conclusion-Based Majority:* The group could also just do the opposite: adopt the majoritarian judgment on the conclusion and just leave open the issue of which premises to accept, if there is disagreement.

- *Majority with a referee:* The group aggregates by majority if it is consistent. When majority is inconsistent, it defers to one particular judge (maybe one selected at random).

Many other rules are, of course, possible.[8] The important point for now is that L&P reflect a sentiment in the literature that, in the context of complex Judgment Aggregation, it is epistemically advantageous to aggregate by Premise-Based Majority (where possible).

This position is based on an extension of CJT to the multi-proposition case. Here, we measure the reliability with different aggregation rules get *all of the relevant propositions* right. Let me illustrate with the tenure example. Think of all the possible, logically consistent, combinations of truth-values involving T, R, and Y (call them *valuations*). Given that we're holding fixed $(T \,\&\, R) \equiv Y$, there are only four such combinations:

	T	R	Y
C_1	$\top$	$\top$	$\top$
C_2	$\top$	$\bot$	$\bot$
C_3	$\bot$	$\top$	$\bot$
C_4	$\bot$	$\bot$	$\bot$

To estimate the reliability of an aggregation rule R, we ask how likely is it that R will output the true valuation. This move effectively reduces the multi-proposition analysis to a single-proposition analysis: the single propositions we analyze are true in exactly one valuation. To assess the reliability on valuations, we must figure out the values of $PTR[C_1]$,

[8] In particular, my short catalogue does not include the important family of *distance-based* rules. For some examples, see Konieczny and Pino-Pérez (2002); Pigozzi (2008); Miller and Osherson (2009). For epistemic analyses that include discussion of distance-based rules, see Hartmann and Sprenger (ms.); Hartmann, Pigozzi and Sprenger (2010).

$PTR[C_2]$, $PTR[C_3]$, $PTR[C_4]$. If we want a single quantity, the natural Bayesian answer is to calculate a weighted average of these PTR's, where the weights are the probabilities of each valuation. Letting m be the (finite) number of valuations, we can call this quantity *global competence* and define it as:

$$\textbf{Global Competence} = \sum_{k=1}^{m} PTR[C_k] \cdot Pr[C_k]$$

We can define a parallel quantity for NTR, but in the following I focus on this positive version of Global Competence. Also, since nothing hangs on this but it will make life easier, let us assume that the probabilities of the valuations are all equal.

Working with complex Judgment Aggregation problems requires additional assumptions (Bovens and Rabinowicz, 2006):

v Each judge has identical competence on any two premises.

vi The premises are probabilistically independent.

vii The probability that a given judge is correct on a premise is independent of the probability that the same judge is correct on the other premise.

(v) is just a simplifying assumption. By contrast, (vi) and (vii) are both extremely substantive and, in my view, rather counterintuitive in the tenure example: excellence at teaching does not seem uncorrelated with excellence at research (contrary to assumption vi). Perhaps, excellent researchers are also excellent teachers; perhaps, less optimistically, excellent researchers are *worse* teachers. Even if both effects were present, we have no evidence that they cancel out exactly (not to mention that we could concoct cases in which there clearly are correlations only going one way). As for assumption (vii), having a correct belief on the one proposition does not seem uncorrelated with having a correct belief on the other: being competent and careful in assessing the evidence could be a common cause of both.[9]

Both assumptions severely impact the analysis. I will concede them, at least for now. With these assumptions in hand we can claim:

Premise-Based JT (PBJT)
Under assumptions (ii)-(vii) and on an agenda with two premises and one conjunctive conclusion, forming opinion by Premise-based majority has:

(a) global competence (and negative global competence) greater than the global competence of individual members.

[9]Bovens and Rabinowicz (2006) deploy the assumptions, but they are clear about their counterintuitive status and cautious about deriving conclusions from them.

(b) global competence (and negative global competence) increasing (and eventually approaching 1) with increase in group size.[10]

PBJT establishes that the Premise-Based approach performs quite well from an epistemic point of view. Clearly, PBJT does not establish that groups should adopt PB as their aggregation rule, even if its preconditions are satisfied. It is possible that some other rules may share the same properties.

3 Objections

In this section, I highlight some key problems in L&P's epistemic analysis. L&P cash out the significance of their chapter 4 in very cautious terms. In the chapter's conclusions, we read: "there may not be a 'one size fits all' organizational design that is best for all group agents and all epistemic tasks." (p.102). I agree that the best organizational design may vary with the nature of the group as well as the task. But that leaves room for disagreement as to *how* it depends. I do not suppose L&P to be unaware of—or unable to respond to—the problems I discuss below; my point is simply that they are unsuccessfully resolved within *Group Agency*.

I. Possibility Results?

As is clear, one of the central problems in interpreting Condorcet-style results is exactly how to handle the heavy dependency on assumptions. If I interpret L&P correctly, they take a rather defensive line here: "[...] our primary aim has been to establish a possibility result, showing that a group agent can indeed live up to the epistemic desideratum." (p.102)

I find this too weak. Talk of possibility results doesn't seem to have sufficient normative force. Consider an analogy: Harman (1986) famously pointed out that even when p entails q and I believe p, it doesn't follow that I ought to believe q. This was part of an argument against the view that one can read off normative conclusions from a logical entailment relation. It wouldn't work to reply to Harman's argument by arguing that *sometimes* there are benefits to believing the consequences of what one believes. Similarly, pointing out that information pooling is *sometimes* good is not enough to convince someone who is skeptical of the idea. I do not mean to argue that L&P should demonstrate that information pooling is *always* epistemically valuable. What they have to do is provide a goal that is more robust than just a possibility result.

Moreover, if the possibility result were the only goal of the chapter, I would happily declare myself convinced of it by a relatively small subset of L&P's arguments. I note, for

[10] In an important sense, part (b) is more important in this version than in the original CJT: the global competence of a rule can be $< .5$ even though r is clearly above .5 (for $r^2 > .5$ to hold, r needs to be $> .71$). So it is only for relatively large groups that we can have a guarantee that the Premise-Based procedure can latch reliably on the true valuation.

example, that the possibility result would be established independently of L&P's discussion in §4.3 of what happens when we relax the assumptions of CJT. For that discussion to be relevant, they must be aiming for something more ambitious.

One plausible interpretation of the project is that, in addition to the possibility result, L&P also claim:

(i) that the cases of which the assumptions hold true are in some sense especially significant.

(ii) as we increase the class of circumstances that we consider, the results are sufficiently robust to give us optimism about the epistemic value of pooling information in groups.

In the following, I will take something like this to be their central theoretical goal. I share the optimism on (ii), but I am more skeptical of (i). I discuss one problem below in Objection V.

II. Bias in Favor of Complete Rules.

It is not clear how the reliability analysis should ground comparisons among rules. I grant that in some elementary cases it is not too difficult to derive epistemic verdicts of the sort that L&P produce. For example, they reject the Unanimity rule on easily reconstructed epistemic grounds. Start with a modest principle:

> **Low Convergence:** a rule R does not have good epistemic performance if (though not necessarily only if) whenever judges are competent but fallible, R's global competence converges to some value below .5 as group size increases.

Next, observe that group competence under the Unanimity rule converges to 0 as group size increases.[11] Low Convergence sustains the judgment that aggregating by Unanimity is not good for truth-tracking purposes.[12] However, it is not strong enough to do all the necessary work. For example, it does not rule out dictatorships when individuals are competent but significantly fallible (e.g., competence between .5 and .7). Moreover, L&P's emphasis on Premise-Based rules implies that we have an epistemic reason to prefer such rules to other rules that are also unproblematic vis-á-vis **Low Convergence**—for example, Majority with a referee. Some stronger principle must be grounding these judgments.

[11] If the judges are competent but fallible, then as the number of independent judges increases, it is nearly certain that someone will make a mistake (and so the group's pattern of acceptance will fall short of unanimity).

[12] This is not to say that there is no possible justification for Unanimity. Recall that we are still working under the assumption of independent judges: some of the cases in which a Unanimity rule seems appropriate are cases in which there are certain dependencies among the judges.

L&P propose something along these lines:

> **Tracking Necessity:** "an agent is a good truth-tracker on p only if it has a high positive tracking reliability and also a high negative one."

However, there are a few problems with Tracking Necessity. The intensity of these problems depends on what 'high' means. Two things need to be resolved:

(i) how high is 'high'?

(ii) should we check a rule's global competence in the limit or for a group of a realistic size?

I would think it is more natural to evaluate group organization based on the group's actual composition, rather than how the group's organizational design would perform in the limit.[13] This leaves open (i): we could say that the threshold for 'high' is .5 but then Tracking Necessity doesn't give us a number of verdicts we would like (e.g. it doesn't imply that dictatorships in which the dictator has reliability .6 are bad). Alternatively, we could say that the threshold is some high value (say .8): it becomes hard even for PBP in otherwise optimal conditions to meet the threshold.

Additionally, it is unclear how Tracking Necessity can sustain comparisons among rules. One such example is the rule I earlier called *Majority with a referee*. In the agenda of the tenure example, provided that individual competence is high (>.71), *Majority with a referee* converges to 1 as group size increases. Is it better than Premise-Based majority? Is it worse? Tracking Necessity is not specific enough to adjudicate.

One alternative idea would be to appeal to a notion of epistemic dominance. Let us say that a rule *R dominates* a rule *T* just in case, for all individual competence levels in (.5, 1] and all odd group sizes,[14] the group's global competence according to *R* is at least as large as (and in some cases strictly larger than) the group's global competence according to *T*.

> **Rule Dominance:** *R* does not have good epistemic performance if (though not necessarily only if) there is another rule that dominates *R*.

Majority with a referee dominates any dictatorship in this sense, so it might appear that Rule Dominance implies that dictatorships do not have good epistemic performance. The problem is that Rule Dominance is too strong: it can be used to disqualify any incomplete rule.[15]

[13] This is not to say that the convergence results are irrelevant.

[14] We must restrict the criterion to odd group sizes, because if the group size is even, majority ties are possible.

[15] *R* is incomplete iff for some proposition p and some distribution of individual opinions, *R* does not accept p and does not accept $\sim p$.

To see this point, suppose R is incomplete: for some patterns of individual opinion, R doesn't return a full valuation. That's to say, it leaves some propositions undecided. Let π be such a pattern. Consider now a rule R^* that behaves exactly like R, except that on π it selects one valuation at random among those consistent with R's other verdicts. R^* will dominate R but it is not, intuitively, an epistemic improvement on R^*.

The problem is made worse by the fact that any arbitrary completion of an incomplete rule R must dominate R. The reason for this is that Rule Dominance is only sensitive to Global Competence; Global Competence in turns depends only on PTR; and PTR does not give any penalties for mistaken opinions.

A more nuanced and intuitive picture would be to score rules by some weighted mix of positive and negative tracking reliability. One possible implementation of this idea exploits resources that L&P invoke at one point, but do not otherwise draw upon:

> In jury decisions, false positives – that is, convicting the innocent – are usually consider worse than false negatives – that is, acquitting the guilty. [...] In medical decisions, by contrast, we usually prefer diagnostic tests that minimize the occurrence of false negatives at the expense of more false positives. (p.90)

These points seem to suggest a notion of value for an acceptance state (possibly a purely epistemic such notion, or possibly a mix of epistemic and pragmatic values).[16]

A simple application of this notion would be to score aggregation rules by a mixture of positive and negative tracking reliability—assigning variable weight to PTR and NTR according to the relative importance of accepting truths and avoiding errors. An alternative, in a similar spirit, would be to assess aggregation rules by their expected values.[17] Developing this sort of proposal in detail is evidently beyond the scope of this note, but it is a start in the direction of evaluating rules in a way that is sensitive to the relative values of accepting truths and avoiding error.

Summing up, there seems to be a problem with how PTR and NTR are used in L&P's argument. It is generally not enough to have a clearly defined measure: we also need a clear way of applying the measure in comparing rules. Low Convergence is too weak, as is Tracking Necessity; Rule Dominance is stronger, but it is biased towards complete rules.[18] They may be good to evaluate complete rules, but we should not assume they can fruitfully serve other comparisons.

[16] Such a notion is likely to vary from domain to domain, and must certainly vary from proposition to proposition within a domain.

[17] If the notion of value is purely epistemic, this would be an instance of epistemic decision theory. These kind of theories are discussed in Maher (1993); Percival (2002). There are some canonical objections against them (Stalnaker, 2002; Goldman, 1986, p.101-102). I have no space to develop them here, but it seems that these objections, if persuasive, extend to L&P's original tracking analysis.

[18] I should add that Tracking Necessity is also biased towards complete rules, although to a lesser degree.

III. Should Aggregation Rules Be Complete?

Perhaps there is no problem in appealing to measures that are biased in this way. This is so if there is an independently motivated argument for complete rules. L&P do offer one consideration in favor of privileging complete rules:

> [...] if a group is to perform robustly as an agent, it must generally avoid attitudinal incompleteness; it must be able to make up its mind on the main issues it confronts. Recall, for example, the multi-member court deciding a breach-of-contract case or the expert panel giving advice on global warming. Incompleteness may not be an option in such groups, as they may be expected to take a clear stance on every proposition-negation pair brought to them for adjudication (p.53).

I assume that none of these cases turn in any special ways on the urgency to make a decision. Even then, I do not find the argument conclusive: from the fact that incompleteness may not be an option, it does not follow that the group's judgmental task needs to be resolved by the aggregation rule itself. After all, a group may attempt to resolve deadlocks by further deliberation or by consulting some additional external judges.

Moreover, the argument would also seem to undercut PB when there is an even number of judges. If a group is perfectly split on some particular issue, they have to resolve their deadlock by somehow going beyond their aggregation rules. The need to resolve deadlocks is just as urgent for groups with an even number of members as it is for groups with an odd number. Notice that many paradigmatic examples of group agents—the US Senate, some Philosophy departments, or most nuclear families—may lack the opportunity to vote with an odd number of judges.

Finally, even for groups with odd sizes, there are agendas that create majority inconsistencies in which the perfect symmetry between the judges' opinions makes it implausible to try to solve the inconsistency by "finding the right aggregation rule". Consider what we may call the *Resource* agenda. Suppose that the university gets money to fund two exactly new projects. It has three possible candidates: a new Philosophy lounge (P), a new Math lounge (M) and a new History lounge. (H). The connection rule is $\sim(P \,\&\, M \,\&\, H)$. The university defers to a three-member committee which judges as follows:

	P	M	H
Judge 1	Y	Y	N
Judge 2	Y	N	Y
Judge 3	N	Y	Y

Once again, the majority here is directly inconsistent with the connection rule. If each project truly has the same weight, it seems odd to insist that the resolution of the dead-

lock be delivered by the aggregation rule alone.[19] The group may need to deliberate further, either about the reasons that support each judgment, or about the general criteria surrounding the decision.[20] Since these other possibilities are available, the fact that sometimes groups *have* to form a complete judgment does not entail that the *aggregation rule* is responsible for producing such complete outputs.

IV. The Significance of Indicating Reliability

There is one more argument in favor of Tracking Necessity—an argument that does not explicitly assume an independent preference for complete rules. L&P argue that, through Bayes Theorem, we can see that having a high *PIR* requires having both high *PTR* and *NTR*. Perhaps, that counts as a strike in favor of Tracking Necessity. This argument would be cogent if there was an independent reason to adopt rules that have high *PIR*. However, the only such reason I can reconstruct depends on a slight misinterpretation of the role of *PIR*. This is how L&P describe the role of Indicating Reliability:

> An agent's indicating reliability has a natural interpretation in terms of rational belief updating by an outside observer. An agent's positive indicating reliability on 'p' is the credence that you, the outside observer, are entitled to assign to 'p' on learning that the agent has judged that p. Likewise, the agent's negative indicating reliability on 'p' is the credence that you, the observer, are entitled to assign to 'not p' on learning that the agent has not judged that p. (p. 83)

I agree with the suggested interpretation of indicating reliability for *individual* agents. However, it does not apply very well to group agents. I illustrate the point with the majority rule, but it applies generally. Suppose a school board must determine whether a school has sufficient fire safety protection (call this proposition φ). Suppose that the board consists of independent judges. Finally, suppose that the majority of board members accepts φ. It does not follow that an external observer should fix their credence in φ exactly at the value of $PIR[\varphi]$ for the majority rule. The actual distribution of votes (regardless of the aggregation rule used) provides more specific information about how to set one's credence.

This point is best visualized by considering two scenarios: in scenario 1, the board votes three against two in favor of φ; in scenario 2, they vote unanimously for φ. In both scenarios, the majority accepts φ. It does not follow that an external observer should set their credence in φ at the positive indicating reliability of the majority rule. In fact, the external observer should not set their credence in φ at the same value: it should be higher in scenario 2. But then the indicating reliability of the majority rule can't be the guide

[19] The distance-based rule described in Pigozzi (2008) leaves the output in cases like this incomplete. It strikes me that this is the right verdict.

[20] See List's discussion of agreement and meta-agreement in List (2002).

for group agents. In any case, credences should be conditionalized on *all* the available information.

If individual votes are available, the observer should conditionalize in scenario 1 on the three to two vote, and in scenario 2 on the unanimous vote.[21] The *PIR* of the aggregation rule should not affect one's credence states, because it is not (or at least not always) the total available evidence. If individual votes are not known, and the group is concerned with setting their audience's credences appropriately, they are best off simply making the breakdown of individual votes public.[22] The upshot is that the *PIR* of an aggregation rule does not seem to play the conceptual role that L&P want it to play. This, in turn, undermines the reason to think that this quantity should be especially high and hence undermines the argument for Tracking Necessity.

V. Reliability and Complex Propositions.

The framework for generalizing the Condorcet-style analysis to complex aggregation problems assumes that the probability that an agent (individual or collective) will get all of the salient propositions right is simply determined by the reliabilities on each premise. This seems to me to be relatively uncommon.[23]

Suppose that Liz, faced with the judgment in the tenure case, accepts every proposition in the set: $\{T, R, T\&R\}$. Suppose that r is Liz's competence level on each of T and R. The analysis that L&P rely upon assumes that Liz's Global Competence in this case must be r^2. *The reasoning:* Liz's competence on each premise is r; since these competence levels on the premises are independent, her probability of being right on both premises is r^2; since Liz is right on both premises if and only if she is right on the conclusion, Liz's competence on the whole valuation must also be r^2. Strikingly, the final reliability calculation is unaffected by whether Liz has independent reliability on the conclusion; her degree of reliability on the conclusion is simply derived from her reliability on the premises.

Consider a disjunctive agenda. A committee needs to decide whether Mary is qualified to be hired for a research job (Q); the committee members agree that she is qualified if and only if she has a PhD in Electrical Engineering (E) or in Computer Science (C). The connection rule here is $Q \leftrightarrow (E \vee C)$. The analysis requires that a subject's competence on Q be derived from the judge's competence on C and E. Suppose that the information in Mary's references suffices to establish with near perfect reliability that she has one of those two degrees. Suppose, finally, that we have less solid evidence as to which degree she has (so a committee member's competence in judging $\sim C$ is some lower value r).

[21] In fact, these points flow from the very analysis of List (2004)—whose results are also discussed in *Group Agency* (p.91).

[22] When I say 'best off' here I mean it in a purely epistemic sense. Of course there could be non-epistemic reasons to avoid making the breakdown of votes public.

[23] Thanks to Franz Dietrich for bringing to my attention an example like the one I discuss in this subsection. The point is also acknowledged by in *Group Agency*, p. 92.

In such a case, the judgment that Q is extremely reliable, but it is *not* a function of my competence on the premises. By contrast, in this case, the analysis instead requires that Global Competence be r^2. There is no general reason to assume that this relationship must hold.[24] Global Competence need not be a direct function of the competence on the premises.

VI. Weighted Majorities.

The solution to the problem of variable competence is also unsatisfying. The suggestion, recall, was that when judges have variable competence, we should assign them weights that are determined by their PTR.

This method of assigning weights is, in my view, problematic. A group can organize itself as the rule demands only if the degrees of competence are precisely *known* to the group members. To see this, consider two striking features of the rule. First: Incompetent judges ($r < .5$) receive negative weight. If a judge with $PTR = .4$ accepts φ, her judgment will be counted as supporting $\sim\varphi$. Unlike 'standard' majority, the rule does just as well if the competence level is .8 as it does if it is .2.

Second: it is possible for the tiniest differences in a judge's competence to make the group swing from accepting φ to accepting $\sim\varphi$, even if we hold fixed the individual opinions. To see how strong the swings can be, I have recorded in the table below some sample (normalized) weights for different assignments of competence to the judges.

competence	**weight**	**behaves like**
$\langle .60, .70, .80\rangle$	$\langle .153, .321, .525\rangle$	dictatorship of most reliable
$\langle .65, .72, .80\rangle$	$\langle .209, .320, .469\rangle$	majority
$\langle .71, .72, .52\rangle$	$\langle .466, .492, .041\rangle$	majority
$\langle .71, .73, .52\rangle$	$\langle .454, .504, .041\rangle$	dictatorship of most reliable

Suppose now that we can only establish reliability with say .05 margin of error. In several very realistic cases, we won't be able to tell whether a judge should receive positive or negative weight. In other cases, we won't be able to tell whether a judge should operate as a dictator on the issue at hand or whether the rule should behave like majority rule.

This extreme sensitivity makes it hard to see this kind of rule as the natural response to the question: how should a group organize itself when individual competence is variable?

[24] Under one interpretation, this relationship follows from assumption (vii). Suppose I accept $C \vee E$ with perfect competence. Suppose also that I accept $C \vee E$ iff I either accept C or I accept E. Then my competence on the premises would appear to be related. If this is how assumption (vii) is being deployed, its content should be specified accordingly. Moreover, once it is thus specified, it looks to be even less intuitively plausible.

4 Conclusions.

To conclude, I want to emphasize that these points are not meant as hostile to the epistemological project of *Group Agency* (the epistemological project being, of course, a tiny fragment of the much more ambitious project of the book). The take-home point of my discussion is that a comprehensive picture of the epistemology of group agents needs a slight reorientation from L&P's discussion. On the one hand, we need to focus on a loftier goal than merely establishing a possibility result. It is not easy to specify what such goals might look like, but without such characterization, it is difficult to assess the significance of the epistemic analysis. On the other hand, once we do specify such goals, we should bear in mind the possibility that incomplete rules may serve some epistemic goals more effectively than complete ones and settle on appropriate epistemic desiderata accordingly.

BIBLIOGRAPHY

Berg, S. (1996) "Condorcet's Jury Theorem and the Reliability of Majority Voting", *Group Decision and Negotiation*, 5(3), pp. 229-238.

Bovens, L., and **Rabinowicz, W**. (2003) "Democratic Answers to Complex Questions", *Synthese*, 150, pp.131-153.

Condorcet, N. C. (1785) *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*, Paris.

Dietrich, F., (2008)"The Premises of Condorcet's Jury Theorem Are Not Simultaneously Justified", *Episteme*, 5(1), pp. 56-73.

Dietrich F. and **List, C.** (2004) "A Model of Jury Decision Where All Jurors Have the Same Evidence", *Synthese*, 142, pp. 175-202.

Dietrich, F. and **Mongin, P**. (2010) "The Premise-Based Approach to Judgment Aggregation", *Journal of Economic Theory*, 145, pp. 562-582.

Dietrich, F. and **Spiekermann, K.** (ms.) "Independent Opinions?", manuscript, CNRS (Paris) and LSE.

Estlund, D., (1994) "Opinion Leaders, Independence and the Condorcet Jury Theorem", *Theory and Decision*, 36, 131-162.

Goldman, A. (1986) *Epistemology and Cognition*, Cambridge, MA, Harvard University Press.

Grofman, B., **Owen G.**, and **Feld, S. L.** (1983) "Thirteen Theorems in Search of the Truth", *Theory and Decision*, 15, 261-278.

Harman, G., (1986) *Change in View*, MIT Press.

Hartmann S. and **Sprenger, J.**, (ms.) "Judgment Aggregation and the Problem of Truth-Tracking", forthcoming in *Synthese*.

Hartmann S. , **Pigozzi, G.**, and **Sprenger, J.** (2010) "Reliable Methods of Judgment Aggregation" in *Journal of Logic and Computation*, 20, 603-617.

Hawthorne, J. (ms.) "Voting in Search of the Public Good", University of Oklahoma.

Konieczny, S., **Pino-Pérez, R.** (2002) "Merging information under constraints: a logical framework", *Journal of Logic and Computation*, 12, 773-80.

List, C. (2002) "Two Concepts of Agreement", in *The Good Society*, 11(1), pp. 72-79.

List, C. (2004) "On The Significance of the Absolute Margin", in *British Journal for The Philosophy of Science*, 55, pp. 521-544.

List, C. (2005) "Group Knowledge and Group Rationality: a Judgment Aggregation Perspective", *Episteme*, 2(1), pp. 25-38.

List, C. (2005b) "The Probability of Inconsistencies in Complex Collective Decisions", *Social Choice and Welfare*, 24, 3-32.

List, C. and **Pettit, P.** (2002) "Aggregating Sets of Judgments: An Impossibility Result", *Economics and Philosophy*, 18, pp. 89-110.

List, C. and **Pettit, P.** (2011) *Group Agency*, Oxford, O.U.P..

Maher, P. (1993) *Betting on Theories*, Cambridge, C.U.P..

Miller, M. K. and **Osherson, D.**, (2009) "Methods for Distance-Based Aggregation", *Social Choice and Welfare*, 32, 575-601.

Nozick, R. (1981) *Philosophical Explanations*, Oxford, O.U.P..

Percival, P. (2002) "Epistemic consequentialism", *Proceedings of the Aristotelian Society: supplementary volume*, 76, 121-151.

Pettit, P. (2001), "Deliberative Democracy and the Discursive Dilemma", in *Philosophical Issues*, 11, 2001.

Pigozzi, G. (2008), "Belief Merging and the Discursive Dilemma", *Synthese*, 152(2), pp. 285-98.

Shapley, L. and **Grofman, B.** (1984) "Optimizing Group Judgmental Accuracy in the Presence of Interdependencies", *Public Choice*, 43, 329-343.

Stalnaker, R. (2002) "Epistemic consequentialism", *Proceedings of the Aristotelian Society: supplementary volume*, 76, 152-168.

Roush, S. (2006) *Tracking Truth*, Oxford, O.U.P.