When to defer to supermajority testimony – and when not

Christian List[1]

Original version 20 October 2006, this version 12 November 2013

Pettit (2006) argues that deferring to majority testimony is not generally rational: it may lead to inconsistent beliefs. He suggests that "another ... approach will do better": deferring to supermajority testimony. But this approach may also lead to inconsistencies. In this paper, I describe conditions under which deference to supermajority testimony ensures consistency, and conditions under which it does not. I also introduce the concept of "consistency of degree *k*", which is weaker than full consistency by ruling out only "blatant" inconsistencies in an agent's beliefs while permitting less blatant ones, and show that, while super-majoritarian deference often fails to ensure full consistency, it is a route to consistency in this weaker sense.

Keywords: Testimony, deference, judgment aggregation, majority judgment, supermajority judgment, full consistency, *k*-consistency, discursive dilemma

1. The problem

Philip Pettit (2006) has argued that although it is sometimes rational to defer to majority testimony on perceptual matters – say, whether a car went through the traffic lights on the red – this is not generally the case with matters more deeply embedded in one's web of belief – say, whether abortion is wrong. A key problem is that deference to majority testimony may lead to inconsistent beliefs. For example, suppose one agent believes that *p* and *q* are both true, a second believes that *p* is true and *q* is false, and a third believes that *p* is false and *q* is true. Then *p*, *q*, and *not*-(*p*&*q*) are each believed by a majority, and thus deference to these majorities would lead to inconsistent beliefs. This is a version of the much-discussed "discursive dilemma" or "paradox of majoritarian judgment aggregation" (e.g., Pettit 2001, List and Pettit 2002, drawing on Kornhauser and Sager 1986).[2]

---

[1] This is a revised version of a paper that was first circulated in October 2006. I am grateful to Franz Dietrich, Philip Pettit, and Wlodek Rabinowicz for discussions and feedback, and to the anonymous referees of this paper and the editor of the present volume for very helpful comments. Address: C. List, Departments of Government and Philosophy, London School of Economics, London WC2A 2AE, U.K.

[2] Kornhauser and Sager's "doctrinal paradox" consists in the fact that, in decisions on a conclusion whose truth-value depends on multiple premises (e.g., the conclusion might be that a defendant is liable for breach

Pettit (2006) suggests that "[t]here is another … approach that will do better … This is not to allow just any majoritarian challenge to reverse a belief but to allow only a certain sort of supermajoritarian challenge to do so" (p. 184). As an illustration, he observes that, assuming consistent individual beliefs, there can never be supermajorities of 70% believing each of *p*, *q*, and *not*-(*p*&*q*) to be true. If there were such supermajorities, the inconsistency would have to show up in the beliefs of at least one individual agent.

It is easy to see, however, that a 70% supermajority requirement is insufficient to prevent an inconsistency between a larger number of propositions. In a group of four agents, for example, there can easily be 75% supermajorities for each of *p*, *q*, *r*, and *not*-(*p*&*q*&*r*), even when each agent holds individually consistent beliefs, such as when the first agent accepts all but the first of these four propositions, the second accepts all but the second, and so on, as shown in Table 1.

Table 1: A supermajoritarian inconsistency

	p	*q*	*r*	*not*-(*p*&*q*&*r*)
Agent 1	False	True	True	True
Agent 2	True	False	True	True
Agent 3	True	True	False	True
Agent 4	True	True	True	False
Supermajority of 75%	True	True	True	True

When does deference to supermajority testimony guarantee consistency, and when not? In this short paper, I sketch an answer to this question, drawing on formal results from the theory of judgment aggregation (particularly Dietrich and List 2007, generalizing List 2001, ch. 9; for related results, see Nehring and Puppe 2007).[3] Thus I follow Pettit (2006)

---

of contract, and the premises might be that there was a valid contract in place and that the defendant did a particular action), majority voting on the premises may lead to a different verdict than majority voting on the conclusion. The "discursive dilemma" (e.g., Pettit 2001, List and Pettit 2002) consists in the fact that simultaneous proposition-by-proposition majority voting on multiple interconnected propositions (which need not be partitioned into premises and conclusions) may lead to inconsistent majority judgments.

[3] One difference between the results in my earlier related works (List 2001 and Dietrich and List 2007) and the results in Nehring and Puppe (2007) is that the latter require completeness of judgments, while the

in focusing on the consistency aspect of rationality (other aspects of rationality are beyond the scope of this paper).[4] I state necessary and sufficient conditions for achieving consistency through supermajoritarian deference and also for achieving something less than full consistency: namely what I call "consistency of degree k", in short "k-consistency". This is the requirement that inconsistencies in an agent's beliefs, if there are any, should not be too blatant, where k is an integer number capturing the degree of "blatancy" of the inconsistencies ruled out, in a sense to be made precise. My argument generalizes but also qualifies the observation that deference to supermajority testimony can sometimes be rational, at least on the consistency dimension.[5]

---

former permit incomplete judgments (as I do in this paper). For review articles on judgment aggregation, see List and Puppe (2009) and List (2012).

[4] The somewhat broader title of this paper echoes the title of Pettit's paper (2006). At the end of this paper, I offer some brief remarks about some aspects of rationality other than consistency.

[5] For a more general treatment of the related problem of "judgment transformation" (how agents can/should revise their judgments in light of the judgments of others) and a baseline impossibility theorem, see List (2011). It is worth addressing one potential objection to the present project of analyzing the "logic" of supermajority deference: why should we not simply let our response to any sort of testimony – whether majoritarian, supermajoritarian, or other – be guided by Bayesian conditionalization? In particular, the testimony in question may be interpreted as a piece of information, and Bayesian conditionalization may tell us how to update our beliefs in light of it. This Bayesian approach, however, has two problems from the present perspective. First, it tells us how to update credences (degrees of belief), not beliefs *simpliciter*, which, like Pettit, I am focusing on here. Second, Bayesian conditionalization becomes possible in the present context only once we make fairly precise assumptions about the testimony-generating process, including assumptions about the reliability of the reported beliefs and the nature of their mutual dependence or independence. (For further discussion, see also the last section of this paper.) By contrast, the present analysis of whether deference to supermajority testimony can yield consistent beliefs applies to beliefs *simpliciter* and requires no assumptions about the testimony-generating process. The cost is that this approach informs us only about one aspect of rationality, namely the consistency aspect, while being silent on other aspects that may be required for rationality *simpliciter*.

2. Minimally inconsistent sets and supermajority testimony

What are the simplest inconsistencies that can arise in an agent's belief set?[6] Call a set of propositions *minimally inconsistent* if it is inconsistent but all its proper subsets – obtained by removing at least one proposition from the set – are consistent.[7] For example, the sets {*p*, *q*, *not*-(*p*&*q*)} and {*p*, *q*, *r*, *not*-(*p*&*q*&*r*)} are each minimally inconsistent: each of them becomes consistent as soon as we remove any one of its members. By contrast, the set {*p*, *p*&*q*, *not*-*p*}, although inconsistent, is not minimally inconsistent: even if one of *p* or *p*&*q* is removed from it, it remains inconsistent. Any inconsistent set of propositions has at least one, and possibly many, minimally inconsistent subsets. It follows that any agent with inconsistent beliefs has at least one minimally inconsistent set of propositions among his or her beliefs. Conversely, any agent whose beliefs include no minimally inconsistent set of propositions is consistent throughout.

Under what conditions can deference to supermajority testimony lead an agent to believe a minimally inconsistent set of propositions?

> **Fact 1:** It is possible for a minimally inconsistent set of k propositions to be each supported by a supermajority among agents with individually consistent beliefs if and only if the supermajority size is less than or equal to ${}^{k\text{-}1}/_{k}$.

---

[6] For the purposes of this paper, I define a *belief set* as a set of propositions accepted by an agent. Thus the present focus is on beliefs *simpliciter* (i.e., a proposition is either believed or not), not on beliefs that come in degrees (as captured by credence or subjective probability functions). Crucially, belief sets need *not* be complete. (A belief set is *complete* if it contains a member of each proposition-negation pair from some relevant reference set of propositions under consideration.) Propositions are represented by sentences in a suitable logic, such as standard propositional or predicate logic. Generally, any logic satisfying some minimal conditions (including compactness), as defined in Dietrich (2007), is suitable. Apart from standard propositional and predicate logics, many modal, conditional, and deontic logics are examples of logics to which the present analysis applies.

[7] The significance of the notion of minimal inconsistency for problems of attitude aggregation was identified by Nehring and Puppe (e.g., 2007), who made extensive use of the notion of a "critical family" of binary properties, which, translated into the language of propositions, is equivalent to the notion of a minimally inconsistent set of propositions.

To prove this fact, consider any minimally inconsistent set of k propositions. Call them $p_1, p_2, \ldots, p_k$. I first show that supermajorities of size ${}^{k-1}/_k$ (and by implication of smaller sizes) among agents with individually consistent beliefs can support each of these propositions. Take any set of agents divisible into k subsets of equal size. Suppose the agents in the first subset believe all of the k propositions except p_1, the agents in the second subset believe all except p_2, and so on. As every proper subset among $p_1, p_2, \ldots, p_k$ is consistent – in particular, every subset obtained by dropping precisely one of these propositions – any such agent holds consistent beliefs. But now each of $p_1, p_2, \ldots, p_k$ – that is, each proposition in a minimally inconsistent set of size k – is supported by a supermajority of size ${}^{k-1}/_k$.

Conversely, I show that supermajorities of size greater than ${}^{k-1}/_k$ among agents with individually consistent beliefs can never support all of $p_1, p_2, \ldots, p_k$. Assume, for a contradiction, that there are k such supermajorities. For any two of these supermajorities, even if maximally distinct, the overlap must exceed ${}^{k-1}/_k - (1 - {}^{k-1}/_k) = {}^{k-2}/_k$. For any three, the overlap must exceed ${}^{k-2}/_k - (1 - {}^{k-1}/_k) = {}^{k-3}/_k$. Continuing, for all k supermajorities, the overlap must exceed ${}^{k-k}/_k = 0$. So the supermajorities must have a non-empty overlap, implying that at least one agent belongs to their intersection. But this would mean that this agent holds inconsistent beliefs, contradicting the assumption that all agents in question have individually consistent beliefs. This completes the proof.[8]

3. Ensuring consistency

What, in light of Fact 1, could a rational policy of deference to supermajority testimony look like? Or at least, what could such a policy look like from the perspective of preventing inconsistency in our beliefs?

Suppose the aim is to arrive at fully consistent beliefs. Consider the entire set of propositions on which beliefs are to be formed or revised. (In the theory of judgment aggregation, this is called the *agenda*.) This set could, for example, contain all those propositions that occur somewhere in an agent's web of belief. Let k be the size of a largest minimally inconsistent set constructible from these propositions and their

---

[8] A version of this argument was given in List (2001, ch. 9).

negations.[9] To illustrate, if the only propositions on which the agent forms or revises his or her beliefs are *p*, *if p then q*, and *q*, then the largest minimally inconsistent set constructible from these propositions and their negations would be {*p*, *if p then q*, *not-q*}, and thus *k* would be 3. In the earlier example, where the relevant propositions were *p*, *q*, *r*, and *not-*(*p*&*q*&*r*), *k* is 4. If the set of propositions is larger and more complex, *k* can of course be significantly larger.

Fact 1 immediately implies that the policy of adopting all and only those beliefs held by a supermajority of size greater than $^{k-1}/_{k}$ can never lead to an inconsistency. If it did, the resulting inconsistent belief set would have to include a minimally inconsistent set of propositions; but that set would contain at most *k* propositions (as *k* is the size of the largest minimally inconsistent set constructible from the given propositions and their negations), and Fact 1 implies that no such set can be supported by supermajorities of size greater than $^{k-1}/_{k}$ among agents with individually consistent beliefs. Thus the following holds (Dietrich and List 2007, generalizing List 2001):[10]

> **Fact 2:** Let *k* be the size of a largest minimally inconsistent set of propositions constructible from the propositions on which beliefs are to be formed or revised and their negations. The set of propositions that are each supported by a supermajority of size greater than $^{k-1}/_{k}$ among agents with individually consistent beliefs is consistent.

However, for any supermajority size below unanimity, the set of propositions supported by supermajorities of that size is not guaranteed to be *deductively closed*: the propositions receiving the required supermajority support may entail other propositions that fail to

[9] Formally, the agenda is assumed to be closed under (single) negation and thus to contain proposition-negation pairs. The significance of the largest minimally inconsistent subset of the agenda for the possibility of consistent attitude aggregation was also identified by Nehring and Puppe (e.g., 2007), albeit in a property-based rather than propositional-logic-based framework.

[10] Related but not fully equivalent results (relying on a combinatorial notion called the "intersection property") can be found in Nehring and Puppe's work on binary attitude aggregation (e.g., 2007). As noted above, their relevant results require – unlike here – that belief sets be complete, i.e., for every proposition-negation pair under consideration, the agent has to accept either the proposition or its negation. In the present case, following List (2001, ch. 9) and Dietrich and List (2007), incompleteness is permitted: the agent may have no belief on some proposition-negation pairs.

receive such support (Dietrich and List 2007; see also Nehring 2005). This means that deferring to supermajority testimony on propositions on which there is the required supermajority agreement while suspending belief on all other propositions (within the set on which beliefs are to be formed or revised) may not be a rational policy: it may make the agent vulnerable to "lottery-like" paradoxes: the agent may believe several propositions but fail to believe some of their implications, even among the relevant propositions. (The classical lottery paradox consists in the possibility that an agent may believe of every single one among a million lottery tickets that this particular ticket will not win, while also believing that one of the million tickets will win.) In the present context, this problem can generally be avoided only if the supermajority threshold is raised to unanimity, a condition that ensures a deductively closed set of supported propositions. Empirically, however, unanimous agreement is rare, and so a policy of unanimitarian deference would seldom lead us to acquire new beliefs.

Independently of the issue of deductive closure, another consideration may also push us in the direction of a unanimity threshold. As the set of propositions on which beliefs are to be formed or revised increases in size and complexity, the value of *k* – the size of the largest minimally inconsistent set constructible from these propositions and their negations – typically increases as well, and thus the supermajority threshold required to ensure consistency approaches unanimity.

4. Avoiding blatant inconsistencies

Achieving full consistency in one's beliefs may not always be feasible. Indeed, it is perhaps unrealistic to expect the beliefs of a normal human agent to be consistent. On the other hand, we do expect those beliefs to be free at least from the most blatant inconsistencies. When is an inconsistency blatant? An agent who believes a single proposition that is self-contradictory, such as *p*&(*not-p*), is clearly blatantly inconsistent. An agent who simultaneously believes a proposition and its negation, such as *p* and also *not-p*, is also fairly blatantly inconsistent, even if each of *p* and *not-p* is not contradictory by itself. An agent who believes three propositions which are in contradiction, such as *p*, *if p then q*, and *not-q*, is still rather blatantly inconsistent, but not as much as one who believes a self-contradictory proposition or a proposition-negation pair. An agent with inconsistent beliefs across five propositions, such as four logically independent conjuncts and the negation of their conjunction, is still inconsistent, but intuitively less so than any one of the earlier agents.

Now suppose that, although my large set of beliefs is inconsistent in its entirety, it turns out that every combination of 1588 or fewer propositions among my beliefs is consistent, and the smallest set over which I hold inconsistent beliefs contains 1589 propositions. Should my beliefs still be described as blatantly inconsistent? Intuitively, the inconsistency here is much less blatant than in any of the earlier cases.

My proposal is to measure the blatancy of an agent's inconsistency by the size of the smallest minimally inconsistent set of propositions believed by the agent. The smaller this size, the more blatant the agent's inconsistency. To be sure, this is a rather simple and crude measure, but I illustrate its usefulness in a moment. In the examples just given, the values of the measure are 1, 2, 3, 5, and 1589, respectively, capturing the intuitive ranking of how blatant the inconsistencies in question are.

Just as the blatancy of an agent's inconsistency can be measured by the size of the smallest minimally inconsistent set of propositions among the agent's beliefs, so the degree of consistency of the agent can be measured in a closely related way. Call an agent whose belief set is free from any minimally inconsistent subset of k or fewer propositions *consistent of degree k*, or in short *k-consistent*. For example, an agent who believes no self-contradictory proposition is 1-consistent. An agent who, in addition, believes no proposition-negation pair (and no inconsistent set of similar complexity) is 2-consistent. One who further does not believe any inconsistent set of the form {*p*, *if p then q*, *not-q*} is 3-consistent. And so on. In the contrived example of my less-than-fully-consistent beliefs, I would be 1588-consistent. Full consistency, finally, is the special case of k-consistency for an infinite value of k.

Perhaps the best a human agent can ever hope to achieve is k-consistency for a reasonably large value of k. What could a policy of deference to supermajority testimony look like if the aim were to achieve k-consistency for some finite value of k? The following corollary of the proof of Fact 1 answers this question.

Fact 3: For any value of k, the set of propositions that are each supported by a supermajority of size greater than $^{k-1}/_{k}$ among agents with individually consistent (or merely k-consistent) beliefs is k-consistent.

To prove this fact, fix some value of k and assume, for a contradiction, that some set of propositions that are each supported by a supermajority of size greater than $^{k-1}/_{k}$ among agents with individually consistent (or k-consistent) beliefs violates k-consistency. This set will then have at least one minimally inconsistent subset of k or fewer propositions. Consider the k or fewer supermajorities of size greater than $^{k-1}/_{k}$ supporting those propositions. The proof of Fact 1 shows that these supermajorities must have a non-empty intersection, implying that at least one agent belongs to all of them. By implication, this agent must support a minimally inconsistent set of k or fewer propositions, which contradicts his or her individual consistency (or k-consistency). This completes the proof.

Of course, if the underlying set of propositions on which beliefs are to be formed or revised has no minimally inconsistent subsets of size greater than k, then k-consistency implies full consistency. In this case, Fact 3 reduces to Fact 2. Otherwise, Fact 3 is more general.

Fact 3 suggests that, while full consistency may often be hard to achieve through deference to supermajority testimony short of unanimity, supermajoritarian deference may nonetheless be a good route to k-consistency for a suitable value of k. And this remains true even if the agents constituting the supermajorities in question are themselves merely k-consistent. Thus, for any value of k, deference to supermajorities of size greater than $^{k-1}/_{k}$ preserves k-consistency.

In summary, the larger the supermajority threshold we require for the acquisition of a belief, the less blatant the inconsistencies we are liable to run into.

5. Coherence and correspondence

For a sufficiently high threshold, deference to supermajority testimony may yield consistent beliefs; and for lower thresholds, it may yield beliefs that are not too blatantly inconsistent. In both cases, other beliefs, on which there is no sufficient supermajority agreement, may need to be revised accordingly.

Does this make supermajoritarian deference rational? My focus has been on "coherence" considerations: supermajority testimony is less prone to inconsistency than majority testimony. But "coherence" is only one aspect of rationality and may not be sufficient for rationality *simpliciter*. A different aspect, on which my analysis has been silent, is "evidential well-supportedness", which, in turn, is motivated by "correspondence" considerations. Does supermajority testimony in support of a proposition provide good evidence for the truth of that proposition?

There is one situation in which the answer to this question is positive, as in the simple-majority case discussed by Pettit (2006). This is the situation in which agents meet the (very restrictive) conditions of Condorcet's jury theorem. That is, they each have an *independent* and *better-than-random* chance of making a correct judgment on each proposition in question (e.g., Grofman, Owen and Feld 1983). The probability of a correct majority judgment on such a proposition will then converge to certainty with increasing group size. Further, using Bayesian reasoning, it can be shown that the conditional probability of the proposition being true, given that it is supported by a majority (and, *a fortiori*, the conditional probability given supermajority support), will approach certainty as well (see, e.g., List 2004). Under these Condorcetian conditions, deference to supermajority testimony may be epistemically rational *simpliciter*, over and above yielding consistent beliefs.

In general, however, agents need not meet these demanding conditions on all propositions, let alone on matters deeply embedded in their webs of belief (see also Bovens and Rabinowicz 2006 and List 2006, Section VI). For example, as a simple consequence of the laws of probability, an agent cannot generally be as reliable at detecting the truth of a conjunction as he or she is at detecting the truth of each conjunct. The agent may have a probability of 0.7 of detecting the truth of p and also a probability of 0.7 of detecting the truth of q, but – if his or her judgments on p and q are independent from one another – only a probability of 0.49 of the detecting the truth of $p\&q$. Thus, even if the agent meets Condorcet's conditions (particularly better-than-random reliability) on each of *p* and *q*, he or she may not meet them on their conjunction. Agents who are highly reliable on “simple” matters (such as *p* and *q* taken separately) can still be less reliable on “composite” or “derivative” matters (such as *p&q*). Since the belief set of any agent (expert, witness, and so on) consists of “simple” as well as “derivative” propositions, we cannot assume that the Condorcetian conditions will be satisfied across the board. As a result, there is no guarantee that the majority, or supermajority, will be reliable on all propositions. Deference to majority or supermajority testimony may not be epistemically rational, even if such testimony is internally consistent.

An analysis of when supermajority testimony – over and above being consistent – is a good indicator of the truth of the relevant propositions is beyond the scope of this short paper (for some relevant results, see Feddersen and Pesendorfer 1998 and List 2004). But it seems wise to exercise caution in deferring to such testimony. Before you take on a set of beliefs because there are supermajorities supporting them, make sure these beliefs are not only internally consistent – which they may well be – but also likely to be correct.

References

Bovens, L. and W. Rabinowicz. 2006. Democratic Answers to Complex Questions – An Epistemic Perspective. *Synthese* 150: 131-53.

Dietrich, F. 2007. A generalised model of judgment aggregation. *Social Choice and Welfare* 28(4): 529-565.

Dietrich, F. and C. List. 2007. Judgment Aggregation by Quota Rules: Majority Voting Generalized. *Journal of Theoretical Politics* 19(4): 391-424.

Fedderson, T. and W. Pesendorfer. 1998. Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting. *American Political Science Review* 92: 23-35.

Grofman, B. G., G. Owen and S. L. Feld. 1983. Thirteen theorems in search of the truth. *Theory and Decision* 15: 261-78.

Kornhauser, L. A. and L. G. Sager. 1986. Unpacking the Court. Yale Law Journal 96(1): 82-117.

List, C. 2001. Mission Impossible? The Problem of Democratic Aggregation in the Face of Arrow's Theorem. DPhil-thesis, University of Oxford.

List, C. 2004. On the Significance of the Absolute Margin. *British Journal for the Philosophy of Science* 55: 521-44.

List, C. 2006. The Discursive Dilemma and Public Reason. *Ethics* 116(2): 362-402.

List, C. 2011. Group communication and the transformation of judgments: an impossibility result. *Journal of Political Philosophy* 19(1): 1-27.

List, C. 2012. The theory of judgment aggregation: An introductory review. *Synthese* 187(1): 179-207.

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

List, C., and C. Puppe (2009) Judgment aggregation: a survey. In P. Anand, C. Puppe and P. Pattanaik (eds.), *Oxford Handbook of Rational and Social Choice*. Oxford (Oxford University Press).

Nehring, K. 2005. The impossibility of a Paretian rational. Working paper, University of California, Davis.

Nehring, K. and C. Puppe. 2007. The structure of strategy-proof social choice. Part I: General characterization and possibility results on median spaces. *Journal of Economic Theory* 135(1): 269-305.

Pettit, P. 2001. Deliberative Democracy and the Discursive Dilemma. *Philosophical Issues* (supplement to *Nous*) 11 (2001): 268-99.

Pettit, P. 2006. When to defer to majority testimony – and when not. *Analysis* 66: 179-87.