Optimal Judgment Aggregation

Jesús Zamora Bonilla†‡

The constitution of a collective judgment is analyzed from a contractarian point of view. The optimal collective judgment is defined as the one that maximizes the sum of the utility each member gets from the collective adoption of that judgment. It is argued that judgment aggregation is a different process from the aggregation of information and public deliberation. This entails that the adoption of a collective judgment should not make any rational member of the group change her individual opinion, and so the collective judgment can not have any kind of epistemic superiority over the individual ones.

1. Introduction. Judgment Aggregation: A Constitutional View. After centuries of debate, there is no agreement about whether 'knowledge' must be essentially conceived as a cognitive state of individual minds, or must be attributed to some collective entity, that is, whether it's *me*, or *we*, who 'really' knows. A promising analytical approach to this problem has emerged in recent years, which concentrates in a more simple question: what formal connections exist between 'knowledge' as a social entity, and 'knowledge' as a property of those individuals conforming the social? Authors within this *judgment aggregation approach* have mainly employed social choice theory (see, e.g., Sen 1970) as a mathematical tool to analyse how individual judgments determine, according to well specified rules of aggregation, the claims endorsed by the collective, and also to what extent the rationality that individuals may display in their own epistemic states is transferred to the group's opinions.[1] The most consequential result that

1. A related problem, i.e., whether some aggregation rules are capable of producing rational collective judgments out of not-so-rational individual opinions, has not been attacked within this approach until now, but would deserve discussion.

JESÚS ZAMORA BONILLA

TABLE 1. Editorial Decision for a Mathematics Journal.

| Agents | Statements | | |
	(A) The premises of a mathematical proof are true	(B) The steps in the proof are valid	(A&B) The proof is sound (and so the result is a theorem)
1	V	F	F
2	F	V	F
3	V	V	V
Majority decision	V	V	F

has emerged from this approach is the 'doctrinal paradox', or 'discursive dilemma', first described by Lewis Kornhauser and Lawrence Sager (1986), and generalized in a number of 'impossibility theorems' by other authors, mainly Philip Pettit, Christian List, and Franz Dietrich.[2] According to these results, when the members of a group disagree about certain statements, and they form a 'collective opinion' by means of some democratic aggregation procedure, it is possible for the group to reach conclusions that are mutually contradictory, even if each member has an internally coherent set of opinions.

Table 1 illustrates this possibility: the editorial board of a mathematical journal is formed by three members; they have to decide whether to accept a paper which contains the presumed proof of an important theorem, but they are discussing whether the proof is valid or not. This happens if and only if every premise of the theorem is true (*A*), and every deductive step is valid (*B*). Each member is willing to accept the paper if and only if *A&B* is true, but they have to answer the author with some *reasons* justifying the board's decision. The problem they face is that, though a majority of the board accepts claim *A*, and a (different) majority accepts *B*, the majority still rejects *A&B*, and so, the message they are going to send to the author would be something like: "we don't consider valid your proof, *because* we think that it is based on true premises, and all the logical steps are valid."

Independently on one's position in the debate about the social nature of knowledge, this result constitutes a serious problem, both from a theoretical and from a pragmatical point of view, for there is no doubt that many instances of what we take as 'knowledge' in our complex modern societies—based on public deliberation and on the division of intellectual labour—is the result of 'aggregating' thousands of epistemic inputs provided by separate but interrelated individuals. A particularly important case is that of *scientific knowledge* itself, which basically consists in some kind of *negotiated consensus* among the specialists, but a consensus that

2. List and Pettit (2002); Dietrich (2006).

always conceals a number of more or less significant disagreements. It is surprising that the discursive dilemma has escaped the attention of relativistic philosophers, for it might apparently provide new arguments to show the lack of objectivity of accepted scientific claims; however, those tempted to follow this road should also be aware of some results about the probability of 'good' epistemic outputs arising from the aggregation of individual opinions (e.g., List 2005; Goldman 2004). My aim is precisely to show that, in spite of the mentioned 'impossibility theorems', certain mechanisms can guaranteeing the *optimality* of the procedure according to which individual claims are aggregated (mechanisms in which some of the assumptions made in those theorems are obviously relaxed).

I will not employ in this paper social choice theory, but *constitutional political economy*, which studies the collective choice of norms by rational agents.[3] This approach is based on the assumption that social interactions presuppose that the members of a group can jointly establish specific *rules* to govern that interaction; if this is the case, we can also assume that the chosen rules will be *efficient* (in the sense that no other conceivable set of rules would have been better for some members, and at least as good for the rest), and furthermore, that if these norms are to be in force during a long time, and have to be acceptable for people with highly conflicting interests, then they will tend to be *impartial* (in the sense that they would correspond to a choice made 'under a veil of ignorance'). This last hypothesis can be made operational by assuming that the chosen rules maximize the *average expected utility* of the group's members.

The first question from a contractarian view is: why do we worry at all about having something like a 'collective opinion', as something different from the mere *enumeration* of our individual judgments? In most cases this will be due not to some intrinsic interest of the group members, but to the demands of *external* people. For example, complaining customers do not care about the particular opinions of the company's workers or counsellors; citizens demand that a single and coherent law is passed by the Parliament; engineers want scientists to tell them the laws governing some physical system, and so on. Thus, many groups feel an external pressure to provide *unified* reasoned claims. Interestingly, Philip Pettit and his collaborators have been engaged in the literature on judgment aggregation surely in part because the 'discursive dilemma' constitutes a threat to the *deliberative republicanism* advocated by him (cf. Pettit 2001a, 2001b). According to this view, a necessary condition for democracy is that political or administrative agencies are 'rationally contestable,' that is, it must be possible for other agents to engage in a reasoned deliberation

---

3. Brennan and Buchanan (1985). Interestingly enough, List and Pettit (2006) refer to the aggregation mechanisms as 'constitutions'.

with the former. Hence, it is not only that we often want that collective agencies 'speak with a single voice', but also that the claims endorsed by this 'voice' are *logically articulated*, so that rational discussion is possible.

But there is still a very important aspect of the process of judgment aggregation that needs to be taken into account, and that is usually neglected in the literature: a reasonable assumption to make is that the aggregation process takes place *only after every individual has taken into account the judgments of the others*; that is, we can assume that a previous process of *public deliberation* has taken place, during which each agent presents her reasons in favour or against each debated statement, and perhaps revise her own judgment accordingly. (Christian List has suggested that this can be modelled as if each individual carried out a process of judgment aggregation within her own mind). *Public aggregation*, instead, starts when deliberation has finished, that is, when no one of the reasons publicly discussed makes anybody else change her opinion, or, stated differently, when an equilibrium is attained in the deliberation process. This apparently innocent assumption, together with the hypotheses that the agents are *rational*, has a dramatic consequence for the philosophical discussion on judgment aggregation: *there is no reason to suppose that the collectively adopted judgment is 'epistemically superior' in any sense to the individual judgements*.

This means that a judgment aggregation problem is not equivalent to an analogous problem of aggregation of *information*. In the latter case, we can take the opinion of every individual as a kind of *statistical estimator* of the truth of the relevant propositions, and, by knowing each individual's reliability, it would be possible to make an inference to the theory that is *most likely* true. But we are assuming that, were there some compelling, public argument showing that the collective judgment is *more* likely true than the opinion of an individual, then *this individual* would have a reason to change her mind, but we have supposed that *every* rational change of individual opinions has already been made. Hence, *the collective judgment should not force any member of the group to change her individual opinion in any way*. From an epistemic point of view, this means that we must not think of the collective judgement as a 'better' opinion than the individual ones. The problem of judgment aggregation, then, is not that of 'finding the truth amongst a bundle of contradictory opinions', but rather, that of *how to live and act in a group in which there are irreducible cognitive disagreements*. It is important not to forget that this is primarily a problem *for the members of the group*, and not for the philosopher observing them from the outside. This is the main justification of the *contractarian* approach defended in this paper: judgment aggregation mechanisms need not be justified by means of philosophical or mathematical arguments (though some of these can obviously be relevant), but mainly by means of the

practical advantages or disadvantages that having one mechanism or another will have for the people whose judgments are going to be aggregated.

2. Two Examples. I shall consider a couple of idealized scenarios, which are just taken as illustrations. In the first place, I will assume that agents have purely epistemic preferences, that is, they only care about the 'distance' between their own individual opinions and the collective claims, but are not worried at all about the practical consequences this distance may entail (or these practical consequences' value is proportional to the epistemic distances). In the second place, I will make the opposite assumption: individuals don't care at all about the epistemic difference between the collective opinion and their own, but they fear that, the most informative this collective opinion is, the higher is the risk of the group taking a decision contrary to their individual interests (if they happen to disagree with the group's opinion). At least in the examples chosen below, where there is an optimal judgment aggregation mechanism it will also happen that the members of the group will be constraining their decision to the choice of a *consistent and deductively closed set of claims*. This assumption contradicts one of the conditions on which the impossibility theorems are based: systematicity, that is, that the same aggregation rule is uniformly applied to all the propositions one by one.[4] But, leaving aside the discussion about the possible reasons to prefer rules that obey systematicity, our assumption allows to solve in a straightforward way the problem of the possible inconsistency of the collective judgments: since the choice is now directly made on *sets* of propositions, and not claim by claim, the members of the group just abstain from including among the available options those sets that are internally inconsistent. Furthermore, if the options only include deductively closed sets of propositions, then all the possible deductive relations between propositions are automatically taken into account in the collective choice, and this debilitates the paradoxical appearance of the discursive dilemma (which is due to the fact that a different collective decision may emerge if the choice is made at the level of the 'premises' or at the level of the 'conclusions'); I will assume, instead, that the group is not choosing isolated claims, but a *theory* (in the logico-mathematical sense of the word), that necessarily incorporates all the relevant deductive connections between the proposition it contains.

2.1. Judgment Aggregation by Purely Epistemically Oriented Agents. Imagine there is a group composed by an odd number (n) of individuals,[5]

4. These rules are what List and Pettit (2006, 12) call 'set-wise supervenient'.

5. That n is odd is assumed to avoid ties.

each one having a certain opinion about k independent atomic proposi-
tions, $p^1, p^2, \ldots, p^k$. Since we assume that every individual has a definite
judgment about every proposition, we say that each agent believes a *com-
plete and consistent theory*. In this propositional framework, complete
theories can be axiomatised by a proposition of the form: $p = \pm p^1 \& \pm$
$p^2 \& \cdots \& \pm p^k$, where each symbol '$\pm$' is to be replaced by a negation
or by nothing; since every proposition is assumed to be logically inde-
pendent of the rest, these complete theories are consistent (inconsistencies
will only take place in someone accepts p^n and $\neg p^n$). Each complete and
consistent theory is then equivalent to some row of a traditional truth
table. The theory accepted by individual i will be called $p_i = \pm p_i^1 \& \pm$
$p_i^2 \& \cdots \& \pm p_i^k$. Now suppose that the group has to take a collective
decision about *what complete and consistent theory represents in the best
way the judgments of the group's members*. In order to answer this question,
we need some information about the *epistemic* preferences of the indi-
viduals. The particular assumption I am going to make in this section is
that every member of the group only cares about the 'distance' between
the theory which is collectively adopted and the theory she personally
believes. This distance can be measured in a very simple way:[6]
$d(p, q) = (1/k)$(number of mismatches between p and q), and hence I will
assume that the utility i receives if theory q is collectively accepted is given
by the formula $u_i(q) = 1 - d(q, p_i)$. (If the individual is uncertain about
which p is true, an expected utility function can be employed. (I owe this
observation to Christian List.) Unfortunately, I have no space here to
develop the consequences of this idea. From these assumptions, several
interesting theorems can be derived:

(1)   For every distribution of individual opinions, there is one theory
      which maximizes the sum of the individual utilities.

This is straightforward: given the opinions of the individuals, each com-
plete theory will have associated with it a certain degree of total utility,
and for one of these theories this sum will have a maximum value.[7] A
much more relevant question is whether the members of the group have
some way of finding which one the optimum theory is. The following two
theorems show that this is certainly the case: *premise-based majority voting*

6. The notion of 'logical distance' has been exploited in the literature on 'verisimilitude'
(cf. Niiniluoto 1987).

7. For every atomic proposition p, only one of p or $\neg p$ minimizes the sum of the
distances to the individual beliefs about p; so $B \neq A$ and B will have at least one
atomic proposition for which the sum of distances to individual judgments is not
optimal, and hence B cannot be optimal. (The proof depends on there being an even
number of individuals.)

(PMV), which consists in each member casting a separate vote on each atomic proposition p^n, and selecting as the collective judgment p^n or $\neg p^n$, depending on which option receives more votes, has the desired properties.

(2)   The outcome of PMV maximizes the sum of the individual utilities.

Proof: For each atomic proposition p^j, majority voting guarantees that the outcome has fewer mismatches with the individual opinions about p^j than its negation.

(3)   PMV is nonmanipulable.

Proof: For each atomic proposition, no individual can attain a higher utility level by voting the negation of the proposition she accepts, than by voting this proposition.

So, PMV can be seen as the optimal judgment aggregation rule for agents that have the type of preferences assumed in this section.[8] Of course, by accepting the outcome of this voting mechanism, the group will be committed to accept many claims that some members would individually reject, and it is also possible that *some* propositions the group is forced to accept (because they logically follow from the adopted complete theory) are rejected by a majority. Furthermore, it would not be strange that the outcome of PMV were a theory that *everybody* would individually reject! In our scenario, however, this would only be an apparently dramatic conclusion, because the voted theory is just taken as something that *represents* in the best possible way the variety of opinions of the group's members. It is really the result of a *compromise*, and, as in most cases of bargaining, the final outcome simply does not coincide with the optimum choice of any of the parties, though it minimises the aggregated losses.[9]

2.2. Judgment Aggregation by Cynical Agents. I will consider next a situation which, in a sense, is a mirror image of the previous one. In the first example, I assumed that the members of the group only care about how far the collective opinion lies from their own individual judgment about the truth. The *practical* consequences that the group will draw from having formed one opinion or another have not been taken into account,

8. PMV was defended by Pettit (2001a, Chapter 5) because it generates a consistent collective opinion. What my argument adds is that, in this idealized scenario, the rule is also optimal from an epistemic point of view, and it forces individuals to sincerely reveal their true opinions.

9. Two further complications that deserve study refer to the conventional nature of the choice of atomic propositions, and to the possibility of individuals having a reservation utility from not reaching a consensus. The first problem relates to the problem known as 'language variance' in the literature on verisimilitude.

or have been simply assumed to have an effect on each individual's utility function that is strictly proportional to the distance between the collective and the individual judgment. But this will certainly not happen in many real situations (perhaps scientific research is the best example of an institution relatively close to the idealized one depicted in Section 2.1). Now I will assume, instead, that the individuals are utterly *cynical*, in the sense that, no matter how 'close' your personal opinion is from the collectively agreed one, if the latter is inconsistent with the former, that is, if your own preferred theory happens to 'loose' in the voting, then the collective judgment will be *interpreted* by the winners in the most beneficial way for them, and the least beneficial way for you, that is, they will use the collective judgment to justify those practical decisions that satisfy in the best possible way *their own* interests, at the cost of yours. Logic imposes some limits to this cynical use of reason, but the limits are often wide enough as to permit a considerable degree of *exploitation* of those that disagree on the public opinion. My strategy in this section will be to assume that the members of the group, knowing this, may want to establish some constitutional mechanism that minimizes the chances of being exploited, or, more exactly, that maximizes the difference between the benefits they derive when they win and the costs they suffer when they lose. Now the agents can choose, not only amongst complete theories, but amongst all consistent and deductively closed set of sentences.[10] In order to calculate the sum of individual utilities derived from the choice of a particular theory, we have to make it explicit some assumptions about the individuals' preferences ('$u_i(A)$' indicates the utility agent i receives if theory A is the collective choice):

(a) If $p_i \vdash A \vdash B$, then $u_i(B) \leq u_i(A)$.

(b) If $A \vdash B \vdash \neg p_i$, then $u_i(A) \leq u_i(B)$.

(c) $\forall A \forall i \forall j$, if $p_i \vdash A$, and $p_j \vdash \neg A$, then $u_i(A) = -u_j(A)$.

(d) $\forall A \forall i \forall j$, if $i, j \vdash A$, then $u_i(A) = u_j(A)$.   (4)

(e) $\forall i \ u_i(Taut) = 0$.

(4a)–(4b) assert that, amongst two theories that i accepts, she will prefer as the collective opinion the one with more content, and, amongst two

10. Adding these options in the case of Section 2.1 does not vary its result, because all noncomplete theories give a total utility lower than that associated to the optimum complete theory, if a noncomplete theory's is measured as the average of the complete theories of which it is the disjunction.
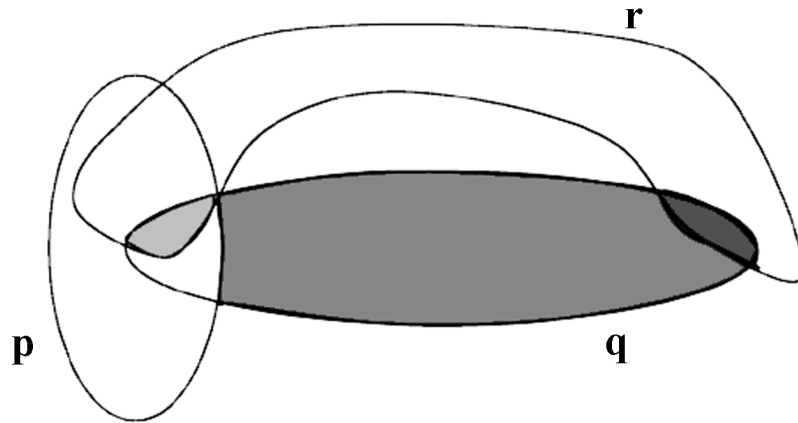
Figure 1. The weaker theory ¬*p*&*q* is 'closer' to the individual opinion than the stronger theory ¬*p*&*q*&*r*.

theories she does not accept, she will prefer the less contentful; these assumptions reflect the 'cynical' attitude agents have towards the collectively adopted claims: if a member of the group agrees with the collective opinion, then she will want this opinion to be as strong as possible, but if she disagrees, she will want it to be as weak as possible. On the other hand, (4c–e) are assumed by analytical convenience ('*Taut*' stands by the tautology); in particular, (4d) seems reasonable when discussing a choice made 'under a veil of ignorance'.[11] A possible objection to (4b) has been suggested by Franz Dietrich: if an individual believes *p*&*q*&*r*, this assumption entails that she will prefer that ¬*p*&*q* is collectively adopted, rather than ¬*p*&*q*&*r*, though in the latter case a new proposition accepted by her has been added. My answer is simply that those individuals that happen to have this type of preferences (or those situations that generate this type of payoffs) are better represented by the scenario of Section 2.1. On the contrary, when people are afraid of linguistic manipulation of reasons, the new assumptions seem more reasonable. Take into account that the 'distances' between the different theories may depend on the set of concepts with which the language operates (recall note 9), and, in absence of a clearly predetermined way of 'measuring' the similarity between several propositions, those 'distances' become extremely subjective. Figure 1 represents this possibility: the individual opinion is *p*&*q*&*r* (the lightest gray area); ¬*p*&*q* is the medium gray area, while ¬*p*&*q*&*r* is the

11. The hypotheses are not logically independent; for example, (4b) can be derived from (4a) and (4c), and (4c) also entails (4d).

dark gray area. In this example it is clear that the weaker theory ¬*p&q*
is 'closer' to the individual opinion than the stronger theory ¬*p&q&r*,
though this includes an additional claim the agent accepts. I am not
assuming that examples like this one are the norm, but, as long as the
possibility exists of using the collectively accepted claims to take decisions
that serve to exploit the 'dissidents', the new hypotheses about individual
utility functions become more justifiable.

From the point of view of the group's members, the most important
fact is that, in each collective choice situation, there will be some theory
for which the sum of individual utilities attains a maximum, and agents
would like to have an aggregation procedure which systematically leads
the group to accept that theory. I will show that there is a mechanism
that, even if it fails to select the optimum theory in each particular col-
lective choice, it generates an optimal pattern of choices on the average.
By *theory based majority voting* (TMV) I will refer to a process in which
the members of the group can form coalitions that propose a theory, A,
which is then voted. If a majority of the members of the full group vote
in favour of A, it becomes the collective opinion. If no theory attains a
majority, then the group suspends judgment (this can be represented by
the choice of *Taut*), which results in everybody having a utility equal to
0. I introduce a further difference between *simple* majority voting and
qualified majority voting; in the latter case, some predetermined percent-
age w (≥ 0.5) of the group must vote in favour of the proposed theory if
it is to be socially accepted (in the case of simple majority voting, $w =$
0.5). A theory A is *w-defeatable* if and only if there is another theory B
such that the set of members for which $u_i(A) < u_i(B)$ constitutes a w-
majority. The following proposition states some very basic properties of
this voting procedure (S^w represents the outcome of applying the w-TMV
aggregation procedure; $p_i$fo represents the complete theory accepted by i,
as before):

(a) $\forall\, w \exists i, j, \ldots, l, S^w = p_i \vee p_j \vee \ldots \vee p_l$ \hfill (5)

(b) S^w is non-w-defeatable.

Proof: (5a) follows from (4a), for, the individuals voting in favour of S^w
will always prefer it to any other theory entailed by it. (5b) follows from
the fact that, if S^w were w-defeatable, then another coalition would propose
some theory which defeats S^w (this is equivalent to saying that S^w is the
theory which maximizes u_i for those individuals belonging to the winning
w-majority).

Let S^* be the theory which would maximize the sum of individual
utilities if collectively chosen. Then the most important result is the fol-
lowing theorem:

There is some qualified majority level, w', such that $S^* = S^{w'}$.   (6)

Proof: (4c–d) entail that S^*, having a positive sum of individual utilities, will have a majority of members in favour of it (i.e., for which $u_i(S^*) > u_i(\neg S^*)$). Let w' be the proportion of members in favour of S^*. If this theory is w'-defeatable, there will exist another theory, S', such that at least the same number of members prefer S' to S^*, but this, together with (4c–d), entails that S' has a bigger aggregated utility than S^*, contrarily to the definition of S^*.

The next relevant question is whether S^* can be reached by simple majority voting (i.e., whether $w' = 0.5$). It is easy to see that, in general, this will not be the case.[12]

So, there will be, for each collective choice situation, a particular value of w' guaranteeing that w'-TMV selects the optimal theory. Nevertheless, the extent of the optimal qualified majority will probably change from case to case. What the members of the group would like to choose as a *constitutional* rule, under this 'cynical' scenario, will be some value of w that maximizes the average value of the outcomes of w-TMV. The more predisposed they are towards exploiting the other members of the group, the higher the value of w they will choose. The outcome of judgment aggregation in a situation like the one depicted in this section is a collective claim consisting simply in the disjunction of the beliefs of a high proportion of the members of the group. Perhaps this collective claim does not look like a powerful victory of deliberative reason, but we think that it is an extremely important point having shown that even in circumstances utterly inhospitable to reason and dialog, like the ones assumed in this section, agents can find a way of carrying out epistemic negotiations in an efficient way. One really ironic consequence is that the collective judgment made by cynical agents is usually a proposition that a big majority of the members believe to be *true*, whereas in the case of purely epistemic agents we have seen that the collective judgment can be recognise as *false* by all the members of the group (though typically, it is more informative). Hence, cynicism and self-interest are not necessarily an obstacle in the constitution of an epistemically sound consensus.

---

12. (4c–d) entail that the social utility associated to simple majority is equal to the utility of just *one* of the individuals voting for the winner theory (since there are $2n + 1$ individuals, $n + 1$ of them will get $u_i(S)$, and the remaining n will get $-u_i(S)$). Let S^1 be the winning theory if w is set equal to $(n + 2)/(2n + 1)$; in this case the total utility attained by the group is $3u_i(S^1)$ (from a similar argument), and so, simple majority voting will be collectively better than w-majority voting only if $u_i(S)/u_i(S^1) > 3$. Hence, if individual utility decreases 'slowly' from the level it attains with the outcome of simple majority voting to the level attained under unanimity (i.e., when w equals 1), S^* will necessarily correspond to a majority level higher than 0.5.

REFERENCES

Brennan, Geoffrey, and James Buchanan (1985), *The Reason of Rules. Constitutional Political Economy.* Cambridge: Cambridge University Press.

Dietrich, Franz (2006), "Judgment Aggregation: (Im)possibility Theorems", *Journal of Economic Theory* 126: 286–298.

Goldman, Alvin (2004), "Group Knowledge versus Group Rationality: Two Approaches to Social Epistemology", *Episteme* 1: 11–22.

Kornhauser, Lewis A., and Lawrence G. Sager (1986), "Unpacking the Court", *Yale Law Journal* 96: 82–117.

List, Christian (2005), "The Probability of Inconsistencies in Complex Collective Decisions", *Social Choice and Welfare* 24: 3–32.

List, Christian, and Philip Pettit (2002), "Aggregating Sets of Judgments: An Impossibility Result", *Economics and Philosophy* 18: 89–110.

——— (2006), "Group Agency and Supervenience", unpublished.

Niiniluoto, Ilkka (1997), *Truthlikeness.* Dordrecht: Reidel.

Pettit, Philip (2001a), *A Theory of Freedom: From the Psychology to the Politics of Agency.* Cambridge: Polity.

——— (2001b), "Deliberative Democracy and the Discursive Dilemma", *Philosophical Issues* 11: 268–299.

Sen, Amartya (1970), *Collective Choice and Social Welfare.* San Francisco: Holden Day.