CAMBRIDGE
UNIVERSITY PRESS

Eliminating Group Agency

Lars J. K. Moen

Department of Philosophy, University of Vienna, Universitätsstraße 7, 1010 Vienna, Austria
Email: lars.moen@univie.ac.at

Abstract

Aggregating individuals' consistent attitudes might produce inconsistent collective attitudes. Some groups therefore need the capacity to form attitudes that are irreducible to those of their members. Such groups, group-agent realists argue, are agents in control of their own attitude formation. In this paper, however, I show how group-agent realism overlooks the important fact that groups consist of strategically interacting agents. Only by eliminating group agency from our social explanations can we see how individuals vote strategically to gain control of their groups and produce collective attitudes we cannot make sense of if we treat groups as agents.

1. Introduction

Explaining how certain groups function, group-agent realists argue, requires us to interpret them as agents in their own right (Tollefsen 2002*a*, 2002*b*, 2015; List and Pettit 2011).[1] Group-agent realism is thus a rejection of the eliminativist view of group agency as at most a metaphorical shorthand and unhelpful for understanding groups' place in the social world. Like eliminativists, the group-agent realists I focus on accept the individualist view that there are no psychologically mysterious forces to be accounted for in explanations of social phenomena. Groups are nothing more than individuals organised in accordance with a certain structure (List and Pettit 2011: 4–6). They nonetheless argue that we need to treat certain groups as agents in their own right to fully understanding how they function.

In this paper, I explore group-agent realism and consider whether it offers an approach to explaining groups' behaviour that is superior to that of its eliminativist rival. As I go into in section 2, group-agent realism is based on an interpretational account of agency, on which a system, or object, is an agent as

---

[1] List and Pettit develop their group-agent realism in several works, but their most comprehensive account is List and Pettit (2011).

long as we can best interpret its behaviour by viewing it as an agent capable of rationally forming and acting on its own attitudes. Groups are agents, then, insofar as their behavioural patterns can reliably be interpreted by treating them as agents.

The paper's main purpose is to demonstrate why we cannot expect group members to behave in the way required for groups to function in this manner. I show how group-agent realism and its corresponding rejection of eliminativism rest on an implausible account of individual agency. The eliminativist approach, on the other hand, allows for a more plausible view of how individuals behave within the group structure. Only on this view can we appreciate how groups themselves consist of strategically interacting agents. I show how group members behave strategically to make their groups function in ways that will make no sense to us if we treat the groups themselves as agents. By fully appreciating the implications of groups being constituted by intentional agents, therefore, we shall see why group-agent realism fails.

In section 3, I start exploring developments in social choice theory over the last two decades that have strengthened the case for group-agent realism. By aggregating individual group members' consistent attitudes towards a set of logically interconnected propositions, we might get inconsistent collective attitudes. For example, there might be majority support for p and q separately but not for '$p \land q$' conjunctively. In cases of such inconsistency, groups need the capacity to select only some of their members' attitudes and adopt whatever attitudes follow whether the members endorse them or not. To adopt p and q separately, a group must also adopt '$p \land q$' conjunctively even when there is majority opposition to this proposition.

This is a cognitive process, Christian List and Philip Pettit argue, in which the group reflects on the inputs from its members and forms its own attitudes, which might be irreducible to those of its members. I show in section 4 how List and Pettit understand this process to force an autonomous group mind into existence, as the content of group attitudes is in part determined by which decision procedure the group applies, and not just by individuals' attitudes. Crucially, two groups whose members hold identical attitudes can therefore form different attitudes (List and Pettit 2011: 66). We therefore cannot make sense of how groups function by treating them as mere collections of individuals, group-agent realists argue. We must instead see them as agents in their own right. Group agency is consequently non-redundant in social explanations (List and Pettit 2011: 6–10).

I begin developing my objection to group-agent realism in section 5, where I first present the necessary conditions for the collective-level inconsistency on which group-agent realism relies. Crucially, one such condition is that group members vote sincerely. I show, however, why we can expect group members to vote strategically to make sure their group forms attitudes consistent with their preferences. By misrepresenting their attitudes towards some propositions on the agenda, the group members can make their group adopt a set of attitudes consistent with their own attitudes towards the propositions most important to them. By such strategic voting, individuals can make the process from which List and Pettit believe group agency emerges irrelevant. Group attitudes are then determined by individuals' preferences, not by their group's correction mechanism.

The importance of group members' sincerity for group-agent realism is further illuminated by the observation that it can lead groups to form false judgements. By strategically misrepresenting their beliefs, individuals can make their group form judgements they know to be false. As I show in section 6, this causes problems for the view of groups as rational agents whose beliefs are generally true (Tollefsen 2002*a*, 2002*b*; List and Pettit 2011: 24, 36–7). This view is unsuited for making sense of false collective judgement formation resulting from strategic voting. To understand this phenomenon, we must descend to the individual level.

Sections 5 and 6 thus show how strategic voting is a problem for group-agent realism. In sections 7 and 8, I explore possible solutions to this problem by considering how truthful voting can be made incentive-compatible. In section 7, I show how familiar social choice results imply that no collective decision procedure can ensure this outcome. I therefore move on to consider the two ways in which List and Pettit (2011: Ch. 5) suggest group members can be made to prefer sincerity. As List and Pettit themselves seem to admit, however, neither of these solutions can be expected to work if implemented.

I therefore conclude that group-agent realism depends on assumptions about individual behaviour that we cannot expect to hold when we seek to explain real social phenomena. We are therefore better off eliminating group agency and focusing on individual agency.

2. Group-agent realism

Group-agent realism is the view that groups can form attitudes that are irreducible to those of their human members, and thus be agents with their own minds. Group agency is therefore distinct from the idea of 'shared agency', associated particularly with Michael Bratman (1999: pt. 2; 2014), where a shared intention is understood as a state of affairs where individuals hold the same intention to perform an action together. As Bratman (1999: 111) notes, such an intention is not attributed to a 'superagent'. The mental states occur only in individuals' minds. Group-agent realists, on the other hand, do attribute mental states to a superagent, or group agent.[2] Certain groups, such as political parties, multi-member courts, legislatures, expert committees, commercial corporations, and tenure committees, they argue, have their own attitudes, and should therefore be understood as agents.

Group-agent realism rests on a functionalist theory of mind that understands a system to possess a mind on the basis of how it functions rather than its physical composition (List and Pettit 2011: 28; Tollefsen 2015: 68–9). Two systems with entirely different physical components can therefore both possess a mind insofar as their parts perform the same function. If a physical event, *a*, in one system performs the same function as a physical event, *b*, in a different system, then *a* and *b* are functionally analogous. If some human action, *a*, within the group structure performs the same function in this structure as some physical event, *b*, in the human brain does with respect to the human mind, then *a* and *b* are

---

[2]For further elaboration on the distinction between group agency and shared, or joint, agency, see Pettit and Schweikard (2006) and Tollefsen (2015: Ch. 2).

functionally analogous. We may then say that *b* is a function of a human mind and *a* is a function of a group mind. This is how a group and a human can both be understood to have a mind despite the differences in their physical make-ups.

The group-agent realists I focus on lean on a particular functionalist theory of mind and agency by taking Daniel Dennett's (1971, 1987, 1991) intentional strategy towards groups to explain their behaviour (List and Pettit 2011: Intro., Ch. 1; Tollefsen 2002a, 2002b, 2015: Ch. 5). On this approach, a system is an agent insofar as we can best interpret its behaviour from the 'intentional stance', from which we treat the system as an agent and predict its behaviour based on the beliefs and desires it rationally ought to have. We assume that the system is rational and capable of forming its own mental states in response to its environment and of acting accordingly. We take the system to act in accordance with a pattern of generally true and consistent beliefs. If this intentional strategy proves to reliably detect such a predictive pattern, the system is an agent. According to group-agent realism, we fail to detect such patterns at the collective level, and therefore miss out on a powerful way of explaining and predicting groups' behaviour, if we see groups as collections of individuals rather than agents in their own right.

Taking the eliminativist approach to group agency is to take the 'physical stance' towards groups. From the physical stance, Dennett explains, we expect behaviour in accordance with the system's physical constitution. Taking the physical stance towards groups, then, means interpreting them by looking at their physical components – that is, their human members. As List and Pettit (2011: 6) note, their approach towards groups 'parallels the move from taking a 'physical stance' towards a given system to taking an 'intentional stance'' (see also List 2019: 155). Seeing groups from the physical stance as mere collections of individuals, they argue, will prevent us from understanding their behaviour.

List and Pettit (2011: 36–7) identify three kinds of standard of rationality a system must satisfy to be interpretable from the intentional stance, and thus qualify as an agent. First, 'attitude-to-action' standards require that a group member can act on the group's behalf. Although important, I shall not focus on these standards here. I devote my attention to the other two kinds of standard. The 'attitude-to-attitude' standards require that the group forms consistent attitudes. These standards are at the centre of the discussion in the next three sections, where the focus is on groups' capacity to form consistent attitudes that are irreducible to their members' attitudes. Finally, the 'attitude-to-fact' standards require the group to reliably form true beliefs based on accessible evidence. I turn to these standards in section 6, where I reveal a problem for groups' ability to form true beliefs.

3. Attitude aggregation

An important step in the argument for group-agent realism is showing why satisfying the attitude-to-attitude standards of rationality requires groups to form attitudes that are irreducible to those of their members. For a group to form consistent attitudes – that is, none of its attitudes logically precludes another – it

might have to adopt attitudes only a minority, or even none, of its members hold. This is why group-agent realists see the need to attribute agency to groups themselves in order to explain their behaviour; such explanation is impossible if we only look at individuals' attitudes.

To understand how groups can form attitudes that are irreducible to those of their members, let us first take collective beliefs or desires to be formed by aggregating group members' binary judgements or preferences, respectively, towards some proposition, p. That is, individuals express a belief that p or not p, or a desire for p or not p. The set of propositions the group members make collective attitudes towards is called the agenda. The agenda is simple if the propositions are not logically interconnected, so that an attitude towards any proposition is not implied by attitudes towards other propositions. The agenda is non-simple when the propositions are logically interconnected. Logical connectives, such as ¬ (not), ∧ (and), ∨ (or), → (implies), or ↔ (if and only if), tie the propositions on the non-simple agenda together. For example, '$p \rightarrow q$' connects p and q, and so does '$p \wedge q$'. An agenda with the two propositions p and q is therefore simple, but we make it non-simple by adding a connecting proposition, such as '$p \rightarrow q$'.

The group can obviously form consistent attitudes reliably reducible to its members' attitudes when the agenda is simple. But a crucial step in the argument for non-redundant group agency is to show why this is not the case if the agenda is non-simple (List and Pettit 2002). On a non-simple agenda, if a system judges that p is the case and that '$p \rightarrow q$', we expect it to judge that also q is the case. However, aggregating individuals' complete and consistent judgements of connected propositions might deliver inconsistent collective judgements.[3]

More precisely, such results may occur in groups where three or more members take part in determining the collective judgements. Let us assume that three group members submit complete and consistent judgements on a non-simple agenda consisting of the three propositions p, '$p \rightarrow q$', and q. But as we see in Table 1, a majority of group members judge that p and '$p \rightarrow q$' are both true but that q is false.

List and Pettit (2002) prove the general result that no non-dictatorial aggregation function that forms collective attitudes towards the propositions on a non-simple agenda can satisfy the following four conditions (quoted from List and Pettit 2011: 49):

Table 1. A group of three members governed by the majority makes inconsistent judgements of three interconnected propositions

	p	$p \rightarrow q$	q
Member A	True	True	True
Member B	True	False	False
Member C	False	True	False
Majority	**True**	**True**	**False**

---

[3]An individual's attitudes are complete when the individual expresses an attitude towards each of the propositions on the agenda.

Universal domain. The aggregation function admits as input any possible profile of individual attitudes towards the propositions on the agenda, assuming that individual attitudes are consistent and complete.[4]

Collective rationality. The aggregation function produces as output consistent and complete group attitudes towards the propositions on the agenda.

Anonymity. All individuals' attitudes are given equal weight in determining the group attitudes. Formally, the aggregation function is invariant under permutations of any given profile of individual attitudes.

Systematicity. The group attitude on each proposition depends only on the individuals' attitudes towards it, not on their attitudes towards other propositions, and the pattern of dependence between individual and collective attitudes is the same for all propositions.

In summary, a non-dictatorial aggregation function cannot both make complete and consistent attitudes and be equally responsive to each group member's attitudes whatever they might be. Pettit (2001: 272) therefore refers to this problem as a dilemma. And he calls it a 'discursive dilemma' because it involves individuals coming together to form collective attitudes. A solution to the discursive dilemma must weaken one of the four conditions.

One solution is to relax collective rationality by allowing for either incomplete or inconsistent attitude-formation. While relaxing the consistency requirement will lead to obvious problems for the group's ability to operate properly, we might consider weakening completeness by introducing a super-majoritarian voting rule. If the group members vote on k propositions, a super-majority greater than $(k-1)/k$ is guaranteed to deliver consistent judgements (List and Pettit 2002: 106). So, if there are three propositions, a super-majority of more than two-thirds will ensure consistency. With four propositions, the required supermajority is more than three-quarters, and so on. But by weakening collective rationality, this aggregation function will make the group unable to form attitudes whenever there is no required super-majority either for or against a proposition. Since this means the group will frequently fail to form an attitude, List and Pettit dismiss this solution.

They also consider weakening universal domain by requiring that group members' attitudes be single-peaked (List and Pettit 2011: 52). If the individuals can be ordered from left to right along a spectrum so that, for any proposition on the agenda, those accepting the proposition are exclusively to the left or to the right of those rejecting it, majority attitudes will correspond with the median individual, whose attitudes, as we have assumed, are complete and consistent (List 2003). This obviously means the group's attitudes will be reducible to those of its members. List and Pettit note, however, that even after deliberation and even in relatively conformist groups, like expert panels, we have no good reason to expect the group members' attitudes to fall into such unidimensional alignment. This is therefore no reliable solution.

---

[4]'Profile' refers to the n-tuple judgement sets across the m individual group members.

A more plausible solution would be to weaken anonymity. By making some individuals' attitudes weightier than those of others, the aggregation function can deliver complete and consistent attitudes. However, List and Pettit (2011: 53–4) reject this solution as objectionable for democratic reasons, and especially point to undesirable epistemic implications, as the group loses access to all the information that is distributed across the group members.[5]

The only remaining solution, then, is to weaken systematicity (List and Pettit 2011: 54–8). That means the aggregation function can assign attitudes to the group that are rejected by most, or even all, of its members insofar as this is necessary to ensure consistent collective attitudes.[6] For example, if the aggregation result is p, q, '$\neg p \wedge q$', a decision rule that weakens systematicity can go against the majority and assign '$p \wedge q$' or $\neg p$ or $\neg q$ to the group.

List and Pettit consider this the most attractive solution to the discursive dilemma. And since it allows the group to form attitudes that are irreducible to those of its members, they argue that we cannot explain the group's behaviour by looking at the individual level. We must instead detect an intentional pattern at the level of the group, thus treating the group itself as an agent. This solution to the discursive dilemma is therefore a crucial step in the defence of group-agent realism.

4. Autonomy

Decision procedures that violate systematicity differ in the ways they correct inconsistencies in collective attitudes – that is, in how they ensure the group satisfies the attitude-to-attitude standards. When considering the different procedures List and Pettit discuss, it will be useful to distinguish between weak and strong interpretations of these standards. On a weak interpretation, all they require is that the group forms consistent attitudes. On a strong interpretation, the standards demand that the group itself be in control of its own attitude formation, autonomously from its members.

Group-agent realism depends on groups satisfying strong attitude-to-attitude standards, as group agency will only then be non-redundant in social explanations. A weak interpretation of these standards is insufficient, as it is compatible with the group forming complete and consistent attitudes that are reducible to those of the group members. List and Pettit (2011: 59) have a strong interpretation in mind when they find a 'surprising degree of autonomy' in group agents. And they clearly take group agents to be in control of their

---

[5]I show in section 6 how an aggregation function that satisfies anonymity can lead groups to form obviously false beliefs. This observation could be understood to challenge List and Pettit's reason for not weakening anonymity.

[6]Pettit (2003: 183–4) gives an example where a group adopts a judgement all its members reject. Three group members vote on three propositions, p, q, and r. They agree that each of these propositions is necessary and that they are jointly sufficient for a fourth proposition, s ('$(p \wedge q \wedge r) \leftrightarrow s$'). In Table 2, we see that a majority of the group members support p, q, and r, but not s. One solution to this inconsistency problem – the premise-based procedure, which I explain in section 4 – results in the group forming a judgement of s that all its members reject.

Table 2. A group forms a judgement all its members reject

	p	q	r	$(p \wedge q \wedge r) \leftrightarrow s$	s
Member A	Yes	Yes	No	Yes	No
Member B	Yes	No	Yes	Yes	No
Member C	No	Yes	Yes	Yes	No
Group	**Yes**	**Yes**	**Yes**	**Yes**	~~**No**~~ **Yes**

attitude formation by understanding them to want to hold consistent beliefs and to pursue consistency in order to satisfy this desire (List and Pettit 2011: 30–1).[7]

What List (2004) calls 'functionally inexplicit priority rules' characterize decision functions that do not give groups the autonomy they need to meet strong attitude-to-attitude standards. Such priority rules do not mechanically give priority to particular propositions but are instead flexible as to which propositions to prioritize in cases of inconsistency. Pettit actually favours a functionally inexplicit priority rule by seeing deliberation among the group members as a good way of dealing with inconsistent aggregation results (Pettit 2007: 512; 2009: 81–8; 2012: 193–4; 2018: 20–1). In the deliberation, the group members decide collectively how to solve the inconsistency. They receive information about how others have voted and then deliberate and figure out how to vote in a second round so as to ensure complete and consistent collective attitudes. In this procedure, List and Pettit (2011: 61–4) see the group members not as separate individuals but as parts of a group mind reflecting on its own attitudes and trying to make them consistent. Tollefsen (2002a: 401), similarly, understands individuals deliberating in an organisational setting to adopt not their own personal point of view but that of their organisation.

The flexibility of such a procedure might make it attractive in the absence of good reasons for prioritising some propositions over others, but it does not justify a kind of group agency that is non-redundant in social explanations. The outcomes of deliberation reflect individuals' strategic bargaining, which is fully accounted for in eliminativist decision theory: the group adopts the attitudes individual group members form on the basis of their preferences. We need not ascribe agency to a group to understand that its members act strategically in response to how others behave or how they expect others to behave. The deliberative procedure is therefore the basis for no more than a redundant kind of group agency, as collective attitudes will be reducible to those of the group members as they express them within the group structure.

The autonomous group-level reasoning necessary for non-redundant group agency seems more plausible on the basis of aggregation functions with a functionally explicit priority rule. Here a mechanism is built into the group's constitution so that it adopts attitudes on particular propositions regardless of the group members' attitudes towards them. List and Pettit (2011: 56–8) discuss

---

[7]List (2018) and Pettit (2018) have recently argued that group agents are conscious, since they must be aware of their own attitudes to ensure their consistency.

various decision procedures with functionally explicit priority rules (see also Tollefsen 2002*b*: 36–9).

First, a 'premise-based procedure' assigns the majority attitudes towards the premises to the group and then the attitude towards the conclusion that follows whether it conforms to the majority attitude or not (List and Pettit 2011: 56). In the above example (Table 1), we can see the first two propositions, *p* and '*p* → *q*', as premises and the third, *q*, as the conclusion. Table 3 illustrates how a premise-based procedure corrects an inconsistency in majority judgements. By adopting the majority judgements that *p* and that '*p* → *q*', the group will automatically accept *q*, even though the majority rejects *q*. In List and Pettit's (2011: 58) view, premise-based procedures 'make the pursuit of agency explicit: they lead a group to perform in the manner of a reason-driven agent by deriving its attitudes on conclusions (or 'posterior' propositions) from its attitudes on relevant premises (or 'prior' ones)'.

A 'conclusion-based procedure' gives priority to the conclusion, so that the group's judgements of *p* and '*p* → *q*' must be adjusted if they conflict with the majority judgement of *q*. Unlike the premise-based procedure, this procedure will not always produce a complete set of attitudes, as there might be more than one way of correcting the group's judgements of premises (List and Pettit 2011: 126–7). In the example above, the majority rejects *q* but accepts *p* and '*p* → *q*'. The group can therefore make its judgements consistent by rejecting either or both of the premises (Table 4). A conclusion-based procedure must therefore be supplemented by a rule determining how the group is to judge the premises.

List and Pettit (2011: 57) also consider the 'distance-based aggregation function'. Collective judgements are here made on the basis of a 'distance metric' between different combinations of attitudes towards the propositions on the agenda. The

Table 3. The group applies the premise-based procedure and accepts *q* in spite of the majority rejecting *q*

	p	*p* → *q*	*q*
Member A	True	True	True
Member B	True	False	False
Member C	False	True	False
Group	**True**	**True**	~~**False**~~ **True**

Table 4. The conclusion-based procedure. Note that this procedure is indecisive on which of the first two propositions to reject, as it can reject either of them to achieve consistency at the group level

	p	*p* → *q*	*q*
Member A	True	True	True
Member B	True	False	False
Member C	False	True	False
Group	**True/False?**	**True/False?**	**False**

complete and consistent set of attitudes with minimal total distance from the individual judgements is assigned to the group. Like any of the other solutions List and Pettit propose, distance-based aggregation will attribute the majority judgements to the group as long as they are consistent. In the above example, however, this procedure will, like the premise-based procedure, accept the majority's judgements of the first two propositions but reject its judgement of the last one, since the 'nearest' consistent attitude set is the one that deviates from only one individual attitude. This is group member B or C's attitude towards q. In some cases, more than one consistent collective attitude set will be nearest, and the distance-based procedure therefore needs a rule for how to break such ties.

Crucially, these decision procedures with a functionally explicit priority rule can produce different collective attitudes on the basis of the same individual attitudes. As we have just seen, with the same inputs from the group members, the premise-based and conclusion-based procedures produce different outcomes, and the distance-based procedure can deliver an outcome that differs from either of them. The same micro-foundations, in other words, can produce different macro-phenomena. Tollefsen (2002b: 41) therefore concludes that we must 'accept the adequacy of macro-level explanations that appeal to the intentional states of groups'.

Group-agent realists understand these procedures to give groups control over their own attitude formation, thus satisfying strong attitude-to-attitude standards. These procedures, List and Pettit (2011: 76) argue, give the group 'an important sort of autonomy'. To understand how groups function, we therefore cannot treat them as mere collections of individuals. We must appreciate that groups can form their own attitudes and make up their own minds. We need to recognize how their self-correction mechanisms turn them into reasoning agents in intentional pursuit of consistent attitudes.

In the rest of this paper, however, I show why procedures with a functionally explicit priority rule fail to give groups the self-control of autonomous agents. Individuals can control their groups' attitudes also under such procedures.

5. Strategic voting

Strategic behaviour can give group members control of their group also under a decision procedure with a functionally explicit priority rule. Awareness of this phenomenon requires that we do not treat the group as an agent; we must instead descend to the individual level of the group members. We must, in other words, take the physical stance towards the group and look at its physical components.

This is to take the intentional stance towards these individuals, from which we see that they act in response to the environment they find themselves in. Within their group, individuals can be expected to behave differently than outside of this structure. Among their fellow group members, they are in a strategic situation and they know it. Particularly, in the situation of collective decision-making, they know that expressing their attitudes without concern for how others express their attitudes might be a poor strategy for making the group

reach the decisions they prefer. How they vote, and, consequently, how the group forms its attitudes, will therefore depend on this knowledge.[8]

To appreciate that group members are strategically interacting agents, we must first recognize that an individual's attitudes and the attitudes she expresses are conceptually distinct and might not coincide. In cases of strategic voting, the two do not coincide. We can expect group members to vote strategically if it will contribute to bringing about their preferred outcome. Strategic voting on a non-simple agenda occurs when individuals care more about which attitudes their group forms towards some propositions than towards others. They give more importance to their 'propositions of concern', to use List and Pettit's (2011: 112) terminology. An individual votes strategically by misrepresenting her attitude towards one or more propositions on the agenda to bring about her desired outcome.

We shall see how such strategic voting can give the group members control of the group's attitude formation, thus eroding the idea of autonomous, non-redundant group agency. I make this evident by first showing how strategic voters can ensure the outcomes they prefer regardless of which correction mechanism the group applies to deal with the inconsistency that would have arisen had they voted sincerely. Group attitudes are therefore determined by individuals' preferences, not by any group-level mechanism. While List and Pettit (2011: Ch. 5) are aware of the possibility of strategic voting, they do not recognize its threat to their group-agent realism (see also List 2004: 505–10; Dietrich and List 2007). But if we fully appreciate the significance of strategic voting, we shall see why group-agent realism collapses.

We have seen how the discursive dilemma is said to give rise to autonomous group agency. Let us now identify three necessary conditions for a discursive dilemma for any non-simple agenda. The first two concern the set of possible profiles of group members' sets of attitudes that result in a discursive dilemma. First, a majority of group members cannot have consistently positive or negative attitudes towards all the propositions on the agenda. In the above example, members B and C ensure that their group meets this condition by supporting one proposition while rejecting the other two. If, on the other hand, a majority of group members have consistently positive or consistently negative attitudes, there cannot be majority support for an inconsistent set of attitudes. This is because we get unidimensional alignment, where each individual can be positioned on the left or the right side of a spectrum for every proposition, so that the majority attitudes towards the propositions on the agenda corresponds with the complete and consistent attitudes of the median individual (List 2003).

Second, within the subgroup of members with a combination of positive and negative attitudes, there can be no majority of members with the same set of

---

[8]Such strategic behaviour does not conflict with Bacharach's (1999) understanding of 'team reasoning', where a group member works out what an imaginary coordinator would require her or him to do as part of a team doing what is best for all its members. Nor is there a conflict with Tuomela's (2013: esp. Ch. 7) very similar account of 'we-mode reasoning', which is the consideration of how to contribute to an outcome that best satisfies one's group's preferences. Group members might vote strategically based on what they consider best for their group as much as on what they believe is best for themselves. In fact, strategic voting in the cases I consider contributes to outcomes favoured by a majority of group members.

attitudes.Group members can vote strategically based on one If, however, a majority of the members of this subgroup hold the same set of attitudes, we again get unidimensionality and these members will be decisive.

These two conditions concerning the group members' attitudes are not sufficient for a discursive dilemma, since we also need a motivational condition to hold. Specifically, the group members with a combination of positive and negative attitudes must express their attitudes truthfully. They must be epistemically motivated, meaning they want their group to adopt the judgements they believe to be true. The more epistemically motivated voters are, the more likely they are to vote truthfully, even when strategic voting will produce their desired outcome. We can therefore see that group-agent realism depends on epistemically motivated group members. It is this sincerity condition I challenge by showing how group members have an incentive to misrepresent their attitudes.

Group members can vote strategically based on one of two kinds of preference orientation (Dietrich and List 2007: 289–91). They have reason-oriented preferences if their propositions of concern are those identified as reasons, or premises, for the proposition identified as the outcome, or conclusion. They have outcome-oriented preferences if their proposition of concern is the outcome, or conclusion. Whenever the first two conditions for the discursive dilemma hold, the voters with a combination of positive and negative attitudes have an incentive to vote strategically, and thus undermine the third condition, insofar as they have reason-oriented or outcome-oriented preferences. C in the above example believes that $\neg p$, '$p \rightarrow q$', and $\neg q$. If C has outcome-oriented preferences, she votes strategically by voting against '$p \rightarrow q$', since contributing to majority opposition to '$p \rightarrow q$' makes her group form attitudes consistent with rejecting q. By voting consistently against each proposition, C cannot contribute to any other outcome than $\neg q$. Strategic voting is thus meant to ensure that the collective attitude formation be determined by the individual's preference orientation rather than by the group's decision rule.

To further clarify the conditions for strategic voting and to show how it puts individuals in control of the collective attitude formation, I begin with an example. A committee consisting of three members will decide by majority voting whether or not to give a junior academic, J, tenure.[9] The committee must give reasons for its decision and therefore bases it on collective judgements of whether J is (*a*) a good researcher and (*b*) a good teacher. The three committee members agree that each of these two conditions is necessary and that they are jointly sufficient for the decision to (*c*) award J tenure ($c \leftrightarrow (a \wedge b)$). To make the case relevant to the question of group agency, let us assume that the committee members hold complete and consistent beliefs about the three propositions and that expressing their judgements according to these beliefs will generate inconsistent majority judgements. Table 5 shows how the majority judgements will be inconsistent if the committee members vote truthfully.

---

[9]Dietrich and List (2007) also use the tenure committee example, but I extend it here to illustrate the significance of strategic voting.

I assume the voters know how many others vote for or against each proposition.[10] They also know that if the committee members vote sincerely, the resulting inconsistency will be resolved by one of the correction rules introduced in the previous section. Sincerity is a dominant strategy for members with consistently positive or negative attitudes towards all propositions on the agenda (except the connective proposition, '$c \leftrightarrow (a \wedge b)$', which all members must accept for the possibility of the discursive dilemma), regardless of preference orientation and decision rule. This is the case for A in the tenure committee example. Members with a combination of both positive and negative attitudes, however, such as B and C, might vote strategically if they attach greater importance to some propositions than to others, as we shall see below.

To demonstrate how strategic voting makes collective decision-making dependent on individuals' preferences instead of the group's correction mechanism, I shall compare how the committee members vote if they have reason-oriented preferences with how they vote if they have outcome-oriented preferences. I do so under each of the three procedures with functionally explicit priority rules I introduced in the previous section. We shall see that by voting strategically based on preference orientation, the members will ensure the same consistent set of collective attitudes regardless of decision rule. What determines the outcome is therefore not the decision rule but the members' preference orientation.

I first suppose that the committee employs the premise-based procedure. If committee members vote truthfully, the committee will accept *a* and *b* and go against the majority by accepting *c*. With reason-oriented preferences, group members have no incentive to vote untruthfully when the group employs the premise-based procedure (Dietrich and List 2007: 290). With this procedure, the committee adopts its members' collective judgements of the premises and will reject the majority judgement of the conclusion if it conflicts with the collective judgements of the premises. The members therefore have no incentive to vote strategically insofar as they have reason-oriented preferences.

If they have outcome-oriented preferences, on the other hand, the premise-based procedure does incentivize strategic voting. In Table 5, we see that committee member B does not believe that *c* is the case, but she believes that *a* is the case.

Table 5. Three committee members' *beliefs* about propositions relevant to whether or not a junior academic ought to be given tenure

	a	*b*	$c \leftrightarrow (a \wedge b)$	*c*
Member A	Yes	Yes	Yes	Yes
Member B	Yes	No	Yes	No
Member C	No	Yes	Yes	No
Majority	**Yes**	**Yes**	**Yes**	**No**

---

[10]Attributing to each voter knowledge of other voters' attitudes is common in analyses of strategic voting. Cox (1997: 78–9), for example, uses such knowledge as a mechanism underlying Duverger's law, which says plurality systems are always two-party systems, since people will not vote for a candidate who will end up third in a plurality-rule election.

B consequently accepts only one necessary condition for *c*, and rejects the other one, *b*. If B has outcome-oriented preferences and he knows that the committee's judgement of *c* will be whatever follows from the majority judgements of *a* and *b*, then B will misrepresent his positive belief of *a* to make sure the committee rejects *c*. Likewise, if C, who also rejects *c*, has outcome-oriented preferences, she has an incentive to misrepresent her belief about *b*. B and C's strategic voting thus ensures that the committee consistently adopts their negative judgement of *c* on the basis of their knowledge that the group will apply the premise-based procedure in the case of an inconsistency (Table 6). The general point here is that outcome-oriented group members will vote on the premises in whatever way necessary to make the group's judgements coherent with their desired judgement of the conclusion (Ferejohn 2007: 133–4).

Table 6. Committee members with outcome-oriented preferences express their judgements when the committee will apply the premise-based procedure in response to an inconsistency

	a	*b*	$c \leftrightarrow (a \wedge b)$	*c*
Member A	Yes	Yes	Yes	Yes
Member B	~~Yes~~ No	No	Yes	No
Member C	No	~~Yes~~ No	Yes	No
Majority	**~~Yes~~ No**	**~~Yes~~ No**	**Yes**	**No**

Let us now instead assume that the committee employs the conclusion-based procedure, which means the committee members vote directly on *c*. In this case, committee members with outcome-oriented preferences will vote truthfully on *c*. The committee's conclusion is therefore the same under the premise-based and conclusion-procedures when the members have outcome-oriented preferences. Generally, if group members have outcome-oriented preferences, the premise-based and conclusion-based procedures will generate identical collective judgements on the conclusion (Dietrich and List 2007: 290–1).[11]

If the committee members have reason-oriented preferences, on the other hand, the conclusion-based procedure will incentivize untruthful voting. B and C will misrepresent their beliefs of *c* to ensure the committee adopts their beliefs about the premises, *a* and *b*. Their strategic voting will then lead the committee to accept all the propositions on the agenda (Table 7). This was also the result we got with the premise-based procedure on the condition that the committee members have reason-oriented preferences. So, we see that the premise-based and conclusion-based procedures will produce the same set of collective judgements also if the committee members have outcome-oriented preferences.

The distance-based procedure also conforms to this pattern. If the committee members have reason-oriented preferences, B and C have an incentive to misrepresent their beliefs of *c* in order to ensure minimal distance to a collective

---

[11]As Dietrich and List (2007: 292) conclude, this observation questions the view common in the deliberative democracy literature that the premise-based procedure is superior to the conclusion-based solution for epistemic reasons.

Table 7. Committee members with reason-oriented preferences express their judgements when the committee will apply the conclusion-based procedure in response to an inconsistency

	a	b	$c \leftrightarrow (a \wedge b)$	c
Member A	Yes	Yes	Yes	Yes
Member B	Yes	~~No~~ Yes	Yes	~~No~~ Yes
Member C	~~No~~ Yes	Yes	Yes	~~No~~ Yes
Majority	**Yes**	**Yes**	**Yes**	**~~No~~ Yes**

Table 8. The outcomes of each procedure under different assumptions about group members' preference orientations

	Reason-oriented preferences	Outcome-oriented preferences
Premise-based procedure	a, b, c	$\neg(a \wedge b), \neg c$
Conclusion-based procedure	a, b, c	$\neg(a \wedge b), \neg c$
Distance-based procedure	a, b, c	$\neg(a \wedge b), \neg c$

judgement set compatible with their beliefs of a and b. And if they have outcome-oriented preferences, they have an incentive to misrepresent their beliefs of a and b to ensure minimal distance to a collective judgement set compatible with their beliefs of c. In this example, the result of such strategic voting will be no distance at all to B and C's desired outcome, as there will be no inconsistency for the group to deal with. We have just seen that we get the same result with the premise-based and conclusion-based procedures. By voting strategically, then, the committee members remain in control of the committee's judgements regardless of decision procedure (Table 8).

In each scenario above, the committee members have the same preference orientation; they have either all reason-oriented or all outcome-oriented preferences. However, their strategic voting will ensure consistency also if some have reason-oriented preferences and others have outcome-oriented preferences. This is so because at least two of the three members will then have consistently positive or consistently negative attitudes towards the propositions on the agenda. When B and C strategically vote consistently for or against the propositions on the agenda, they will ensure consistently positive or negative majority judgements either by combining with A or with each other. The same would hold if A had consistently negative attitudes.

Generally, in any group of three or more members, where the first two conditions for the discursive dilemma hold, so that a majority of the members do not hold consistently positive or consistently negative attitudes, and no majority in this subgroup hold the same attitudes, group members will vote strategically to ensure consistent collective attitudes insofar as they have either reason-oriented or outcome-oriented preferences. The case for autonomous group agency therefore depends on these group members having neither reason-oriented nor outcome-oriented preferences, and therefore no incentive to vote strategically. In all other cases, collective decision making over a non-simple agenda will depend not on the decision rule employed but on the members' preference orientation. Group-agent

realists must therefore show why we should expect sincere voting. In sections 7 and 8, I consider different attempts at making sincerity incentive-compatible.

6. Breaking patterns

While the strategic voting discussed in the previous section weakens the autonomy and non-redundancy of group agency, it does not challenge groups' ability to satisfy weak attitude-to-attitude standards. After all, strategic voting does not jeopardize the consistency of collective attitudes. However, as I have also argued elsewhere, strategic voting can pose a further problem that can undermine groups' agency, not just their autonomy (Moen 2019).

From the intentional stance, we look for patterns not just of consistency but also of factuality. We expect the system in focus to form true beliefs based on the evidence it gathers from scanning its environment. We take the system to adopt the belief that p only if p is the best, or most predictive, interpretation of the matter given the available evidence (Dennett 1987: 29). An unreliable truth-tracker, however, is unpredictable from the intentional stance, and therefore not an agent. This condition for agency is expressed in List and Pettit's (2011: 36–7) attitude-to-fact standards of rationality (see also Tollefsen 2002*a*: 399–400). We have seen how group members might have an incentive to express judgements they believe to be false. We shall now see how this could make their group uninterpretable from the intentional stance.

Group members, List and Pettit (2011: 36) write, are their group's 'eyes and ears', as they provide it with information about the world. When they vote in favour of p, the group treats that as evidence for p, just as perceptual input is evidence for an individual. But this view is distorted by the observation that group members might deliberately misinform their group about the world. When group members vote strategically, they express an attitude they believe to be false. In the cases I have considered so far, any proposition is always believed by at least one individual. These cases therefore suggest the problem of false belief formation is no greater at the group level than at the individual level. But if the group members lack complete information of others' attitudes, there can be outcomes where the group forms an attitude with which all the group members disagree.[12] In such situations, strategic voting problematizes group-agent realists' attribution of predictive patterns of generally true beliefs to groups.

Let us consider an example where strategic voting generates a false negative. In this example, a political party is making a policy decision and will employ the premise-based procedure in the case of an inconsistency. On the basis of their outcome-oriented preferences, the party members deny an obviously true proposition in order to ensure that their party collectively has reasons in support of their desired policy decision. Each of the three party members votes on the following propositions:

---

[12]For an account of strategic voting under conditions of uncertainty of others' attitudes, see Clough (2007).

- *p*: Greenhouse gas emissions cause global warming.
- *q*: Reducing greenhouse gas emissions will not harm the economy.
- *r*: Reducing greenhouse gas emissions will not intolerably infringe on individual liberty.
- $(p \wedge q \wedge r) \rightarrow s$: If *p*, *q*, and *r*, then the party should support policies to reduce greenhouse gas emissions.
- *s*: The party should support policies to reduce greenhouse gas emissions.

All the party members believe *p*, but since B and C reject *s*, they have an incentive to vote against *p*, as well as *q* and *r*, to ensure the party makes their desired decision. The result is that the party forms a belief of *p* that all its members believe to be false (Table 9).

A failure to assess relevant information to form true beliefs indicates that the group has malfunctioned, at least from the perspective of the intentional stance. But any human agent occasionally also forms false beliefs. If infallibility were a requirement for agency, there would be no human agents. We can, however, attribute false beliefs to a system from the intentional stance, but then, as Dennett (1987: 49) writes, 'special stories must be told to explain how the error resulted from the presence of features in the environment that are deceptive relative to the perceptual capacities of the system'. We should look for something in the system's environment that can explain the false belief. For example, gathering relevant information might have been costly, and the system was therefore understandably trying to operate on the limited information it had (Dennett 1987: 96–7). No such explanations are possible, however, when groups form false judgements due to group members strategically misrepresenting their beliefs. The issue may be obvious and the relevant evidence easily accessible, but the group nonetheless fails to form the correct belief.

When a system functions in a way uninterpretable from the intentional stance, we must descend to the physical stance. Only from the physical stance are we aware of any faults that prevent the system from functioning as a rational agent. As Dennett (1971: 89) notes, 'the physical stance is generally reserved for instances of breakdown, where the condition preventing normal operation is generalized and easily locatable'. To make sense of a group's failure to form true beliefs due to its members' strategic voting, we cannot view it as an irreducible agent. Only by looking at the group members and their attitudes can we make sense of the group's failure to form true beliefs.

Table 9. A political party rejects a proposition all its members believe is true, as a result of members voting strategically on the basis of their outcome-oriented preferences

	p	*q*	*r*	$(p \wedge q \wedge r) \rightarrow s$	*s*
Member A	Yes	Yes	Yes	Yes	Yes
Member B	~~Yes~~ No	No	~~Yes~~ No	Yes	No
Member C	~~Yes~~ No	~~Yes~~ No	No	Yes	No
Majority	~~**Yes**~~ **No**	~~**Yes**~~ **No**	~~**Yes**~~ **No**	**Yes**	**No**

From the physical stance, we are aware of the different physical make-ups of groups and individuals. Crucially, groups, unlike individuals, consist of strategically interacting agents, and this is what makes them vulnerable to false belief formation in a way individuals are not. They do not just form false beliefs because their human members are imperfect truth-trackers, but also because their human constituents may intentionally feed false judgements into the aggregation function. To explain and predict the group's behaviour, then, we should descend to the level of the group members and consider their reasons for misrepresenting their beliefs. Only then can we make sense of the group's malfunctioning: its members have strategically misrepresented their beliefs. And by descending to the physical stance to understand the group's behaviour, we disqualify it as an agent.

Of course, strategic voting will not always make the group form false beliefs. However, the cases where strategic voting does lead the group to false belief formation illuminate a potential problem applying only to groups, and not to individuals. The group members' strategic behaviour weakens their group's capacity to form true belief, thus making it less interpretable from the intentional stance. It becomes harder to detect an intentional pattern that can reliably predict the group's behaviour, and therefore less plausible to treat it as an agent.

7. Strategy-proofness

The last two sections have revealed how group members can undermine group agency by misrepresenting their attitudes to make their group adopt their attitude towards their proposition of concern. The challenge for group-agent realists, then, is to explain how group members can be motivated to vote truthfully, so that groups can be interpreted as autonomous agents. They must show how sincerity can be made incentive-compatible.

One way of disincentivizing strategic voting would be to introduce a strategy-proof aggregation function. Strategy-proofness is a game-theoretic concept we can understand by first seeing the aggregation function as a game form with the individual voters as the players (Gibbard 1973). In the case of attitude aggregation over a set of propositions, the players' possible actions are all the different attitude sets they can submit to the aggregation function (Dietrich and List 2007). By assigning payoffs for each player to the different possible outcomes, we have modelled the situation as a game. Which strategies the players pursue determines each player's payoffs.

Strategy-proofness depends on truthfulness being a weakly dominant strategy – that is, no player gets a higher payoff by voting untruthfully regardless of what the other players do. More precisely, an attitude aggregation function is strategy-proof for any individual, for any profile, and for any preference relation, when each individual weakly prefers the outcome following sincere voting to any possible outcome resulting from misrepresenting her or his attitudes (Dietrich and List 2007: 287). Notice that whether an aggregation function is strategy-proof depends on motivational assumptions. All else being equal, strategy-proofness is easier to achieve the stronger the individuals' preference for truthfulness is.

Dietrich and List (2007) distinguish strategy-proofness from manipulability, which is not a game-theoretic concept and does not depend on individuals' preferences. An aggregation function is manipulable if and only if there is a logical possibility of a profile such that a player can alter the outcome by voting untruthfully. An aggregation function can therefore be manipulable but nonetheless strategy-proof if voting truthfully is each player's dominant strategy. Non-manipulability is therefore a more demanding condition than is strategy-proofness.

The Gibbard–Satterthwaite theorem proves the general impossibility of a non-manipulable and non-dictatorial aggregation function with at least three possible outcomes (Gibbard 1973; Satterthwaite 1975). Dietrich and List (2007: 282) extend this result to attitude aggregation to prove that on a non-simple agenda, an aggregation function can satisfy universal domain, collective rationality, responsiveness,[13] and non-manipulability if and only if it is a dictatorship of some individual. We saw in section 3 why List and Pettit will not weaken collective rationality or universal domain. And a dictatorial procedure, in addition to its other unattractive features, is obviously no solution for group-agent realism, since the group's attitudes will then be reducible to those of a single individual. Dietrich and List (2007: 283) also refrain from weakening responsiveness, as that would imply a degenerate aggregation rule. Consequently, the aggregation function must be manipulable.

However, group-agent realism does not depend on a non-manipulable aggregation function. A strategy-proof one will do, since it makes the group members' preferences for the group to adopt their attitude towards each proposition on the agenda stronger than any other preference they might have. It will thus solve the whole problem of strategic voting. The key question is therefore under what conditions group members will be motivated to vote truthfully. Dietrich and List (2007: 280–3) show that an aggregation rule is strategy-proof whenever the following two conditions are satisfied simultaneously (quoted from List and Pettit 2011: 112–13):

> **Independence on the propositions of concern.** The group attitude on each proposition of concern – but not on other propositions – should depend only on the individuals' attitudes towards it (and not on individual attitudes towards other propositions).

> **Monotonicity on the propositions of concern.** [A] positive group attitude on any proposition in that set never changes into a negative one if some individuals rejecting it change their attitudes towards accepting it.

Simple majority voting satisfies both conditions, as it requires that a collective judgement of a proposition depends only on the majority judgement of that proposition. The problem with majority voting, as we have seen, is that it violates collective rationality. On a non-simple agenda, majority attitudes might

---

[13]Responsiveness: For any admissible profile, if one or more voters change their votes in one direction, while all other votes remain unchanged, the collective decision cannot change in the opposite direction.

be inconsistent. List and Pettit therefore reject simple majoritarianism and opt for an aggregation function that satisfies collective rationality but violates systematicity. But to weaken systematicity is to weaken the attitude-aggregation version of Arrow's (1963) independence of irrelevant alternatives condition: the collective attitude towards each proposition on the agenda is to be determined strictly on the basis of the group members' attitudes towards that proposition and no other proposition.[14] To follow List and Pettit in relaxing systematicity is to allow for an aggregation function that rejects this independence condition by adopting collective attitudes towards some propositions on the basis of their connection to other propositions.

The conditions relevant for strategy-proofness, however, are independence and monotonicity *on the propositions of concern*. By specifying which propositions can be of the group members' concern, the aggregation function can satisfy both of these conditions as well as universal domain, collective rationality, and anonymity. The solution, in other words, is to make sure the group members hold preferences towards the propositions on the agenda so that they have no incentive to vote strategically. They will consequently vote on each proposition as if it were logically separate from any other proposition, which means voting on each proposition as if there were only two possible outcomes. By thus not seeing the propositions on the agenda as connected, they will have no reason for voting strategically. Ensuring strategy-proofness, then, means motivating sincerity.

8. Motivating sincerity

List and Pettit (2011: 113–14, 124–8) discuss two ways of making truthfulness incentive-compatible. The first is to assume that the group members' preferences are fixed and then apply an aggregation procedure that satisfies both independence and monotonicity of the propositions of concern. This works with both premise-based and conclusion-based procedures, they explain, '[i]n the lucky scenario in which the propositions of concern are mutually independent and fit to serve either as premises or as conclusions' (List and Pettit 2011: 113). If the group members all have outcome-oriented preferences, and the proposition of concern therefore is the conclusion, then the conclusion-based procedure will meet the two conditions and be strategy-proof. The same holds for the premise-based procedure if the group members' preferences are reason-oriented, so that the premises are the propositions of concern.

As we saw in section 5, however, truthful voting under these conditions does not strengthen the case for group-agent realism, since the individuals remain in control of the collective attitude formation. That is, the outcome is the same regardless of which collective decision function is employed. So, while List and Pettit's first way of

---

[14]Independence of irrelevant alternatives is one of the components that make up the systematicity condition. The other component is a version of May's (1952: 681–2) 'neutrality' condition, according to which the pattern of dependence between individual and collective attitudes will be the same for all propositions on the agenda. Changing the order of the propositions will then not change the result.

ensuring truthfulness will work insofar as the condition of fixed preferences holds, it cannot respond to the problems strategic voting poses for group-agent realism. Furthermore, Dietrich and List (2007: 281–3) show that when individuals' preferences are not fixed, so that they do not all share the same proposition, or propositions, of concern, the only aggregation functions satisfying monotonicity and independence on the propositions of concern are degenerate ones that violate either the condition of non-dictatorship or unanimity preservation.

The other solution List and Pettit (2011: 113) propose is to 'try to change the individuals' preferences by persuading or convincing them that they should care about a different set of propositions of concern'. In connection to the first solution, List and Pettit note that by making individuals' preferences reason-oriented, we can apply the premise-based procedure to make truthfulness a dominant strategy for each individual. In the tenure committee, for example, the problem of strategic voting in response to the premise-based procedure is due to the committee members having outcome-oriented preferences. But if we can make their preferences reason-oriented, they will have no incentive to vote strategically. B will have no incentive to vote untruthfully on a, since he wants the committee to adopt his belief that a is the case. And C will have no incentive to vote untruthfully on b, since she wants the committee to adopt her view that b is the case. Like the first solution, however, this second solution implies that the individuals' preferences, and not the group's correction mechanism, will determine the group's attitude.

Another version of this second solution of changing individuals' preferences is to simply make the group members prefer honesty – that is, expressing their attitudes towards each proposition truthfully with no special consideration for any proposition of concern. If successful, this approach achieves strategy-proofness and the group members will not attempt to prevent collective-level inconsistencies. Which attitudes the group adopts will then depend on which correction mechanism it applies, and the group will be in control of its own attitude formation in the way group-agent realists suggest. Making sense of how the group functions will then require us to treat it as an agent.

The obvious problem here is how to actually change individuals' preferences in this manner. List and Pettit (2011: 127–8) also admit to having difficulties seeing how to make individuals adopt the right kind of preference. They suggest trying to create a social ethos under which people approve of one another's truthfulness and disapprove of each other's strategic voting. Truthfulness will then be an equilibrium strategy on the assumption that people want each other's approval and that they will get it by voting sincerely.[15]

This solution can only work under three conditions. First, individuals' sincere attitudes must be common knowledge. In section 6, I assume voters have knowledge of the aggregate support for or opposition to each proposition on the agenda. This common knowledge condition, however, is different and more demanding, as the group members must know not just how many others support or reject a proposition, but also the attitudes of each particular voter.

---

[15]This idea is based on Brennan and Pettit (2004).

They also know that others have this information and that others know that they have this information. Second, the individuals want to gain each other's approval and to avoid each other's disapproval. While this condition seems realistic, at least in groups whose members frequently interact, the third is less prevalent. This is the condition that individuals actually receive approval for expressing their attitudes truthfully and disapproval for expressing attitudes untruthfully. In my examples above, this condition does not seem to hold, as a majority of the group members vote strategically. It is also unlikely when the strategic voting leads to an outcome most members desire, as in the political party example in section 6.

Without a plausible mechanism showing why individuals will be motivated to vote sincerely, individuals cannot be expected to behave as the group agency model predicts. Group-agent realists have not provided such a mechanism, however, and therefore lack grounds for arguing that group agents are non-redundant in social explanations.

9. Conclusion

Group-agent realists reject eliminativism for failing to capture how groups can reliably form attitudes that are irreducible to those of their individual members. To appreciate this phenomenon, they argue, we must treat groups as agents in their own right. They base this argument on the observation that aggregating individuals' complete and consistent attitudes towards a set of interconnected propositions can produce inconsistent collective attitudes. To reliably form consistent attitudes, the group must therefore have the capacity to form attitudes that are not held by a majority of their members. Group-agent realists understand this as the capacity of a reasoning group mind that reflects on the inputs from the group members and accepts them only if they are consistent. In this process, groups might form different sets of attitudes based on the same profile of the group members' attitudes. To make sense of this phenomenon, group-agent realists argue, we cannot look at the individual level, as eliminativism suggests. We must instead take the step up to the collective level and treat the group as an agent in its own right.

In this paper, I have highlighted a fatal problem in the micro-foundations of group-agent realism. By fully appreciating that groups consist of strategically interacting agents, we can see why we cannot expect groups to function in the way group-agent realism predicts. Collective-level inconsistency depends on a majority of group members voting sincerely when reason-oriented or outcome-oriented preferences will motivate them to vote strategically. Group-agent realism depends on a mechanism showing why we can expect truthful voting from such individuals, since strategic voting will make the decision-making dependent on group members' preferences rather than on the group's decision procedure. However, group-agent realists have provided no such mechanism. And if they cannot show why group members should not be expected to vote strategically, we are better off eliminating group agency, as we can then appreciate the significant impact of individuals' preferences on how groups function.

References

Arrow K.J. 1963. *Social Choice and Individual Values*. Second edn. New York, NY: John Wiley & Sons.

Bacharach M. 1999. Interactive team reasoning: a contribution to the theory of co-operation. *Research in Economics* **53**, 117–147.

Bratman M.E. 1999. *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge: Cambridge University Press.

Bratman M.E. 2014. *Shared Agency: A Planning Theory of Acting Together*. New York, NY: Oxford University Press.

Brennan G. and P. Pettit. 2004. *The Economy of Esteem: An Essay on Civil and Political Society*. New York, NY: Oxford University Press.

Clough E. 2007. Strategic voting under conditions of uncertainty: a re-evaluation of Duverger's Law. *British Journal of Political Science* **37**, 313–332.

Cox G.W. 1997. *Making Votes Count: Strategic Voting in the World's Electoral Systems*. Cambridge: Cambridge University Press.

Dennett D.C. 1971. Intentional systems. *Journal of Philosophy* **68**, 87–106.

Dennett D.C. 1987. *The Intentional Stance*. Cambridge, MA: MIT Press.

Dennett D.C. 1991. Real patterns. *Journal of Philosophy* **88**, 27–51.

Dietrich F. and C. List 2007. Strategy-proof judgment aggregation. *Economics and Philosophy* **23**, 269–300.

Ferejohn J. 2007. Conversability and deliberation. In *Common Minds: Themes from the Philosophy of Philip Pettit*, eds. R. Goodin, G. Brennan, F. Jackson and M. Smith, 121–142. Oxford: Oxford University Press.

Gibbard A. 1973. Manipulation of voting schemes: a general result. *Econometrica* **41**, 587–601.

List C. 2003. A possibility theorem on aggregation over multiple interconnected propositions. *Mathematical Social Sciences* **45**, 1–13.

List C. 2004. A model of path-dependence in decisions over multiple propositions. *American Political Science Review* **98**, 495–513.

List C. 2018. What is it like to be a group agent? *Noûs* **52**, 295–319.

List C. 2019. *Why Free Will Is Real*. Cambridge, MA: Harvard University Press.

List C. and P. Pettit 2002. Aggregating sets of judgments. *Economics and Philosophy* **18**, 89–110.

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

May K.O. 1952. A set of independent, necessary and sufficient conditions for simple majority decision. *Econometrica* **20**, 680–4.

Moen L.J.K. 2019. Redundant group agency. *Philosophy of the Social Sciences* **49**, 364–384.

Pettit P. 2001. Deliberative democracy and the discursive dilemma. *Philosophical Issues* **11**, 268–299.

Pettit P. 2003. Groups with minds of their own. In *Social Epistemology: Essential Readings*, ed. A.I. Goldman and D. Whitcomb, 167–93. Oxford: Oxford University Press.

Pettit P. 2007. Rationality, reasoning and group agency. *Dialectica* **61**, 495–519.

Pettit P. 2009. The reality of group agents. In *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice*, ed. C. Mantzavinos, 67–91. Cambridge: Cambridge University Press.

Pettit P. 2012. *On the People's Terms: A Republican Theory and Model of Democracy*. Cambridge: Cambridge University Press.

Pettit P. 2018. Consciousness incorporated. *Journal of Social Philosophy* **49**, 12–37.

Pettit P. and D. Schweikard 2006. Joint actions and group agency. *Philosophy of the Social Sciences* **36**, 18–39.

Satterthwaite M.A. 1975. Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory* **10**, 187–217.

Tollefsen D. 2002*a*. Organizations as true believers. *Journal of Social Philosophy* **33**, 395–410.

Tollefsen D. 2002*b*. Collective intentionality and the social sciences. *Philosophy of the Social Sciences* **32**, 25–50.

Tollefsen D.P. 2015. *Groups as Agents*. Cambridge: Polity Press.

Tuomela R. 2013. *Social Ontology: Collective Intentionality and Group Agents*. New York, NY: Oxford University Press.

Lars J. K. Moen is a postdoctoral fellow at the University of Vienna. His research focuses on group agency and strategic voting and political concepts of freedom. URL: http://www.lars-moen.com.