

## Rationality, Reasoning and Group Agency

Philip PETTIT<sup>†</sup>

### ABSTRACT

The rationality of individual agents is secured for the most part by their make-up or design. Some agents, however – in particular, human beings – rely on the intentional exercise of thinking or reasoning in order to promote their rationality further; this is the activity that is classically exemplified in Rodin's sculpture of *Le Penseur*. Do group agents have to rely on reasoning in order to maintain a rational profile? Recent results in the theory of judgment aggregation show that under a range of plausible conditions they do. In a slogan: group agents are made, not born.

### *Introduction*

Agents have to display a modicum of rationality in the formation and enactment of their attitudes, else they will not pass as agents at all; their performance will be too random or erratic to count as action. This is true of group agents as well as of individual agents and so it raises the question that this paper addresses. How do group agents achieve the rationality that their status as agents requires? Can we expect rationality to emerge spontaneously when groups are organized properly? Or does the maintenance of a rational group configuration require a continuing surveillance on the part of members? Does it require this as a matter of feasible practice, if not as a matter of logical necessity?

The rationality of individual agents is secured for the most part by their make-up or design. Some agents, however – in particular, human beings – rely on the intentional exercise of thinking or reasoning in order to ensure the maintenance of rationality, and to further its improvement. This is the activity that is classically exemplified in Rodin's sculpture of *Le Penseur*; we all recognize what the bent-over, almost clichéd figure is doing: he is lost in thought. This use of reasoning sharpens the question addressed in the paper. Do group agents have to rely on reasoning in order to maintain a rational configuration of attitudes? Or can such a configuration be maintained without an analogue of thinking?

The first section of this paper sets up the distinction between rationality and reasoning, as it applies with individual subjects, and then the second and third sections ask after its application to the case of group agents. A principal requirement of reasoning, according to the first section, is access on the part of reasoners

---

<sup>†</sup> Department of Philosophy, Princeton University, 308 Marx Hall, Princeton, NJ 08544, USA; Email: ppettit@princeton.edu

to the content of their existing commitments; they must have feedback on the things that they already hold and seek, believe and plan. The second section argues, by contrast with the individual case, that under plausible conditions the absence of feedback to members of a group on their existing group commitments is going to make it hard, perhaps even effectively impossible, for them to achieve group rationality. And the third section then shows that the presence of such system-level feedback, and specifically its use in group reasoning, would suffice to make group rationality accessible.

The argument exploits a lesson of recent results on the aggregation of judgments and provides a novel perspective on the emergence of agency among groups of individuals (Pettit 2003c; List and Pettit 2005; 2006). A brief conclusion construes the upshot in terms of a distinction that is familiar from other areas of discussion, between self-organizing and self-governing systems. This serves to highlight the core message: group agency among human beings does not emerge without effort; group agents are made, not born.

## 1. *Rationality and reasoning*

### *Simple agents*

To begin with, some basics.<sup>1</sup> If a creature is to count as an intentional agent, then it must have desires or goals for which it is disposed to act and it must form beliefs about its environment to guide its action, identifying suitable opportunities and strategies. Such desires and beliefs can be characterized as attitudes towards propositions, with the desire consisting in the targeting of a proposition, the belief in the acceptance of a proposition, and with the distinction between targeting and acceptance being given by a difference in direction of fit. An agent will act to make the world fit a targeted proposition – a would-be goal – and will adjust to make its mind fit a proposition it accepts: a would-be fact (Anscombe 1957; Searle 1983; Smith 1987). It will act for the realization of its desires, seeking to bring the world in line with them; and it will act in this way according to its beliefs, where its beliefs are brought into line by the world. This will be so, at any rate, within what we think of as feasible limits and favourable conditions.

If a system is to fulfil world-changing desires according to world-tracking beliefs, then it must satisfy three sorts of standards. Attitude-to-evidence standards will require, among other things, that the system's beliefs be responsive to evidence. Attitude-to-attitude standards will require that, even as they adjust under evidential inputs, its beliefs and desires do not assume such an incoherent form that they support inconsistent options. And attitude-to-action standards will

<sup>1</sup> These are rehearsed more fully in List and Pettit forthcoming.

require that the system tend to act, and to form intentions to act, on the lines that its beliefs and desires support.

These are standards of rationality, as I understand the term. Rational standards are nothing more or less than desiderata of agency: standards such that agents will generally do better as agents by robustly satisfying them, at least within relevant limits and conditions.<sup>2</sup> Let an agent form beliefs that run counter to the evidence and it will tend to adopt strategies that do not satisfy its desires. Let it form attitudes that support inconsistent lines of action and it will fail even more dramatically in the pursuit of desire-satisfaction. And let it not be disposed to act or intend to act as its beliefs and desires require, and it will fail more theatrically still. Further intuitive desiderata of rationality may be less important but, other things being equal, their satisfaction will also enhance agency. One example is the attitude-to-evidence desideratum that an agent form those general beliefs for which a run of evidence provides inductive support. And another is the attitude-to-attitude desideratum that the agent form beliefs and desires with respect to propositions that are entailed by propositions already believed or desired, or reject the belief or desire held for the entailing propositions.

Quite a simple system can merit the ascription of propositional attitudes, and characterisation as an agent. Consider the little robot that navigates a table top on wheels, scanning various cylinders on the table with bug-like eyes, and moving to set upright any cylinder that falls or lies on its side. Even a system as rudimentary as this can be said to accept propositions to the effect that this or that or another cylinder is upright or on its side and to be disposed, with any cylinder on its side, to target or realize a proposition to the effect that it is upright once again.

Any creature, including one as simple as this robot, will have to display a minimal level of rational competence, if it is to deserve the name of agent. The movement of the robot's eyes will have to pick up relevant evidence about the orientation and location of cylinders on their side. Its cognitive processing will have to ensure that it forms a set of consistent representations as to where the cylinders are. And the representations will have to interact with its overall goal to generate attempts to set those cylinders back in upright position. In other words it will have to display a minimal level of rationality in attitude-to-evidence, attitude-to-attitude and attitude-to-action relations.

Or at least it will have to do this within what we take to be intuitively feasible limits and intuitively favourable conditions. Suppose that the robot tends to knock cylinders at the edge of the table onto the floor. Do we say that besides the desire on which it generally acts, it has a desire to knock such cylinders off the table?

<sup>2</sup> Some of these standards might be taken as relevant, of course, to systems that do not qualify fully as agents. Thus we might invoke the standards of belief-formation that would be appropriate with an agent in assessing a system that can form beliefs but not act upon them.

That will depend on our knowledge of its design specifications. It will be plausible that there is no specification for such a goal, if the actions it uses with cylinders at the edge of the table are just the same as those it uses with other cylinders. In that case we will treat the condition where a cylinder is at the edge of the table as less than favourable and will stick with the original characterisation of its intentional attitudes.

### *Sophisticated agents*

Non-human creatures, certainly non-human animals, get to be much more complex agents than the robot imagined. There are a number of ways in which the robot might be designed to more complex specifications. It might be built to search out and pick up other behavioural strategies, on a trial-and-error basis, if its existing efforts at raising cylinders run into problems – say, if they knock cylinders onto the floor. It might form beliefs about other objects besides the cylinders or about other properties besides the location and orientation of the cylinders. And it might embrace a number of purposes, not just the single goal of setting certain cylinders upright.

But even if such complexities are introduced, robotic agents in this mould will remain simple in one salient respect. They will form propositional attitudes of belief and desire with respect to concrete objects like the cylinders and the salient properties of such objects; they will come to form beliefs in, and desires for, various propositions involving such items and such features. But nothing in the story told means that they will conceptualize and attend to those very propositions, forming attitudes about their properties and relations in turn. Nothing entails that they will practice what we might describe as propositional ascent.

These agents will form ordinary propositional attitudes but not attitudes of a meta-propositional character. An ordinary propositional attitude is an attitude towards a proposition in which only concrete objects and their properties figure, whether that proposition be singular or existential or universal in form. A meta-propositional attitude is an attitude towards a proposition – if you like, a meta-proposition – in which propositions may themselves figure as objects of which properties and relations are predicated. Some meta-propositions will ascribe properties like truth and evidential support and relations like consistency and entailment to propositions, as in the claim that ‘p’ is true or that it is inconsistent with ‘q’. Others will identify propositions as scenarios that are believed or desired by agents or as scenarios that are credible or desirable, as in the assertion that someone believes the proposition ‘p’, or that ‘p’ is credible. Others again will identify them as scenarios that the agent is disposed to realize or as scenarios the agent invites others to rely on his or her realizing, as in the claim that someone is

going to make 'p' true or that 'p' is something someone intends or promises to make true.<sup>3</sup>

The absence of meta-propositional attitudes in the robot and its more flexible counterparts means that they are subject to a salient restriction. A robot might ask itself a question, as we can put it, when it registers a movement in its peripheral visual field and then focuses its eyes on the relevant location out of a desire, derivative from its cylinder-raising desire, to know whether or not a cylinder has fallen; this desire will be satisfied so far as the focusing of the eyes has the effect of letting a belief form on the matter. But because the robot and its counterparts don't have meta-propositional attitudes, they cannot ask themselves similar questions about connections between propositions, say about whether they are consistent or inconsistent, and then do something – pay attention to the inter-propositional relations – out of a desire to have a belief form one way or the other.

This restriction means that the robotic creatures cannot reason. To be rational, as we saw, is to satisfy certain desiderata in the attitude-to-evidence, attitude-to-attitude and attitude-to-action categories. To be able to reason, under the model I shall adopt here, is to be able to conduct an intentional activity that is designed – and perhaps explicitly intended – to raise the chance of satisfying such desiderata (Pettit 1993). Specifically, it is to be able to ask oneself questions about certain propositional properties and relations; to be able thereby to let beliefs form on such matters; and to be disposed to adjust rationally to whatever beliefs one forms. This is a sort of activity that the robot clearly cannot conduct.

Suppose I currently believe that p and that q. Perhaps because of worrying about the conditions in which I formed those beliefs, I may ask myself whether 'p' and 'q' are consistent propositions, setting in train a process of forming a belief in answer to the question and adjusting to the result. And I may do so with benefit. For if I come to form the belief that the propositions are inconsistent, I will have brought to the surface the fact that I believe that p, that q and that 'p' and 'q' are inconsistent. And, if things go right, I will then be prompted to eliminate the inconsistency among those beliefs and achieve a higher degree of rationality; in the ordinary course of events, I will be prompted to give up on my belief that p or that q, depending on which is the more weakly supported.

<sup>3</sup> We often ascribe meta-propositional attitudes, not by identifying a proposition and the property that is predicated of it, but by using propositional operators that make the attitudes seem to be of a regular sort. Instead of saying that people believe that 'p' is true, signaling their capacity to think about the proposition 'p', we say they believe that it is true that p; instead of ascribing a belief that someone believes 'p' we say that they believe that the person believes that p; instead of taking them to see desirability in the realization of a proposition 'q', we say that they believe that it is desirable that q. But I take it that such ascriptions are plausible only with creatures that are capable of meta-propositional thinking. It would be misleading to ascribe beliefs that it is true that p and that it is false that not-p, for example, to a creature that is incapable of recognizing any common element in those beliefs (Evans 1982). And in order to be able to recognize a common element, the creature would have to be able to have beliefs about the proposition 'p'.

The process I will have gone through in such a case constitutes what counts, on the model adopted here, as reasoning.<sup>4</sup> In the example given the reasoning is theoretical in character insofar as the meta-propositional interrogation involved is deployed in the service of checking on the formation of beliefs. An otherwise similar process would constitute practical reasoning if the interrogation were deployed in the service of checking on the formation of intentions. I reason theoretically when the fact of forming the belief that 'p' and 'q' are inconsistent as a result of meta-propositional interrogation leads me to reject at least one of those propositions. I reason practically when the fact of forming the belief that making 'r' true is the only way of making 's' true as a result of meta-propositional interrogation leads me from an intention to realize 's' to an intention to realize 'r'. And so on.<sup>5</sup>

How can subjects like you and me form meta-propositional beliefs and engage in reasoning? The answer surely has to do with the fact that we have language and can use sentences as ways of making propositions into objects of attention. The normal assertoric use of sentences like 'p' or 'q' will be to report the fact that p or that q, expressing a belief in that fact. But the sentences can also be used to exemplify what is said in normal assertoric usage, allowing the speaker to think about those propositions and to let beliefs form as to whether they are consistent or not (Davidson 1984). I say to myself 'p', and I say to myself 'if p, then q', and, registering their relationship, conclude 'so, q'. The robotic agent may be more or less perfectly designed to form the belief that q whenever it forms the belief that p and that if p, q, but it will not register the relationship between those propositions; it will not go through any process that would enable it to use the word 'so' or some cognate.

<sup>4</sup> The model should prove relatively uncontroversial, on two counts. First, it is a model of reasoning, not of good reasoning; and second, it gives an account of reasoning as such, not of the special exercise in which someone makes a judgment as to what good reasons require overall. I reason whenever I set out to form meta-propositional beliefs and let them play out as checks on the process whereby my attitudes form. The exercise will be good reasoning when it takes me where I ought to go or where there is good reason to go, but the theory of reasoning as such need not offer an account of where precisely that is. Thus I do not engage with the issues dividing Niko Kolodny and John Broome, for example; see Kolodny 2005; Broome 2007. Moreover, to go to the second point, when I reason in the sense at issue, I do not need to make any judgment about where there is good reason to go. Why should we say that the exercise constitutes reasoning only in the presence of the rider? And why should we think that we have given an account of reasoning only when we have given an account of what can be put forward as purportedly good reasons? On these matters, see Burge 1998.

<sup>5</sup> The idea is not that the meta-propositional belief will have to be correct and will naturally be authoritative in determining the adjustment for which it argues. The presence of that belief is only meant to put a further check in place, however fallible, by providing another locus at which any irrationality ought to show up. Finding myself with the beliefs that p, that q and that 'p' and 'q' are inconsistent, I might be led to question the belief in the inconsistency, not the belief that p or that q. The possibility is unlikely to materialize to the extent that the inconsistency is a priori demonstrable, or the belief in the inconsistency is the product of more reflective consideration than that which led to the other beliefs; but it is certainly not closed. I am grateful to Rory Pettit for pressing me on this point.

Absent a linguistic or symbolic means of objectifying the propositions, it is hard to imagine how any meta-propositional thinking could emerge (Pettit 1993, ch. 2; Dennett 1994; Clark 1999). Thus it is difficult to see how non-human animals can reason in the active sense described. None ever appears to conduct the activity we naturally ascribe to *Le Penseur*. A dog may perk up its ears on hearing a sound and attend to what is happening, asking itself whether the family is home or dinner is being served. But no dog does anything that we might interpret as asking itself whether certain propositions are really supported by the evidence, whether realizing one will help realize another, whether they are consistent, and so on. Or at least not outside of Gary Larson cartoons.

Reasoning in the sense characterized may be pursued in a content-based or a form-based manner. Standard reasoning involves an activity in which people think about the scenarios that answer semantically to various sentences and reflect on how far they are supported by the evidence to hand, how far the scenarios are consistent with one another, whether one scenario entails another, and so on. But given that many of those sorts of relationships are guaranteed to hold in virtue of the form of the corresponding sentences, reasoning can also be conducted in abstraction from the content, as in seeing that since certain scenarios relate in the form of 'p and q' and 'p', then regardless of how these schematic letters are interpreted, the first must entail the second. Numerical calculation offers an everyday example of form-based reasoning. Using the multiplication rule to work out that eleven times eleven is one hundred and twenty one can be seen as an exemplar of reasoning that abstracts from content.<sup>6</sup>

I said earlier that a simple agent like the robot, or indeed the dog, will have to display a minimal level of rationality in order to earn the name of agent. It must be clear that there are goals that it pursues, that it pursues them according to certain beliefs, and that those beliefs are not randomly or rigidly formed, but display some sensitivity to evidence. Does the same lesson apply to symbol-using creatures like you and me?

Yes and no. Yes, we must display some sensitivity to the demands of rationality. But no, this need not be displayed in the spontaneous, behaviourally vindicated achievement of rational performance. We may display the required sensitivity through showing ourselves to be ratiocinatively if not behaviourally responsive – that is, responsive at the meta-propositional level – to rational requirements. We may not behave as if 'p' and 'if p, q' entail 'q', for example, but that will raise no doubts about our status as agents so far as we can be challenged, forced to recognize the entailment, and led to criticize our own behaviour. Being symbolic creatures, we can have our attitudes identified on the basis of our avowals as well

<sup>6</sup> Such intentional exercises in calculation should be distinguished, of course, from the computations in an uninterpreted language of thought that are sometimes postulated as the means whereby intentionality is sub-personally realized (Fodor 1975).

as our actions and we can vindicate our claim to be agents, not just by acting appropriately, but also by being disposed to recognize the implications of our avowals, to criticize our own performance on that basis and, at least ideally, to defer to those implications, bringing ourselves into line with what they require (Levi 1991; Bilgrami 1998; McGeer and Pettit 2002). We shall notice an analogue of this observation towards the end of the discussion of groups.

*Simple and sophisticated compared*

The rationality of the simple creature is realized sub-personally so far as there is nothing the creature can do in order to improve its rational performance (Dennett 1969). The robot and its counterparts have to be more or less rational in order to count as agents but they cannot exercise their agency with a view to ensuring or enhancing that rationality. They can act on the cylinders and other concrete objects, having beliefs and desires that bear on the properties of those objects. And in order to do this, they must have beliefs and desires that minimally satisfy attitude-to-evidence, attitude-to-attitude and attitude-to-action desiderata. But, not having attitudes that bear on the propositional objects of those beliefs and desires, they cannot do anything to check or channel the process in which the beliefs and desires form, connect, and occasion action. It is by grace of their design, artificial or natural, that they generally satisfy the desiderata of rationality, not by virtue of anything they themselves can do.

We reasoning creatures transcend this limitation. We do not have to rely entirely on the processing for which our nature programs in order to be rational. We can do something about it, as we check out the meta-propositional constraints and connections that are relevant to what we should believe, desire or do. We can monitor and hope to improve our own performance, putting extra checks in place in order to guard against rational failure.

The transcendence of the non-intentional that we thereby achieve is only partial, of course (Carroll 1895). Suppose that paying attention to the relations between the propositions 'p', 'if p, q' and 'q' is to have a rationality-enhancing effect on me, triggering me to move from beliefs in the premises to a belief in the conclusion or to inhibit one of the beliefs in the premises. It can have that effect only if two conditions hold, both of which involve my non-intentional, rational processing. First, I must be able to rely on the attentional activity to generate a correct meta-propositional belief and, second, I must be able to rely on that meta-propositional belief having the required effect. Reasoning does not work in parallel to non-intentional rational process, seeking the same end by different means; it exploits rational process in a novel way, searching out extra inputs to improve as checks on how the process is operating.

Even if it is partial, however, the transcendence that reasoning achieves gives us a degree of personal control over our own rational performance. The control

deserves to be described as personal on two counts. First, it is a form of control in which I intentionally pursue the satisfaction of rational desiderata, rather than merely relying on my non-intentional processing. I act as an intentional system with a view to achieving rationality rather than leaving the task just to sub-systems within me. I am a systemic agent in this domain, not just a site of sub-systemic activity.

The second count on which this form of control deserves to be described as personal rather than sub-personal is that we each exercise it in sensitivity to what we as an intentional system already believe and desire; implicitly or explicitly it requires that we process feedback on where we are already committed, and keep track on those commitments. In principle there might be a creature that asks itself meta-propositional questions without any awareness of what it already believes or desires, and without any awareness of whether the questions asked are relevant to its own attitudes. The creature might even do this with the effect – intended or unintended – of putting extra rational checks on its attitude-formation. But clearly we ordinary reasoners are not like that. If we ask meta-propositional questions about various propositions, at least outside the classroom, then we will do so because those propositions already figure as the objects of our attitudes or are propositions about which we are trying to form attitudes. We will ask meta-propositional questions, and look for the benefits of reasoning, under the guidance of feedback on where we already stand. Thus, if we ask whether ‘p’ and ‘q’ together entail ‘p and q’, that will typically be because of feedback awareness that by our existing lights it is the case that p and it is the case that q.

Not only do I pursue rationality as a systemic agent, then, rather than just leaving it to sub-systems within me. I do it out of a sense of myself as a systemic agent with a record of attitudinal commitment. The first aspect of personal control means that it is I as an intentional agent who seeks to exercise control. And the second means that it is over me, conceptualized as a centre of enduring, available-in-feedback attitudes that I seek to exercise that control.<sup>7</sup>

Our interest in the sections following is in how far the divergence between sub-personally rational and personally ratiocinative agents is replicated with group agents. The most straightforward way of addressing this question will be by looking at the possibility of group agents in which members do not have feedback on the existing commitments of the group and then at group agents where members do enjoy such feedback. The negative thesis of the paper, defended in section 2, is that under a range of plausible conditions groups that lack system-level feedback – and so cannot reason in the ordinary sense – are not going to be able to perform satisfactorily as agents. The positive thesis, defended in section 3,

<sup>7</sup> These considerations do not exhaust all aspects of what is required for personal control: a form of control in which I can properly be said to assume responsibility. The discussion of that topic would take me too far afield. For more, see (Pettit 2001a, ch. 5).

is that groups that have access to such feedback and are able to reason are likely to be capable of a satisfactory performance under those conditions.

## 2. *Group agents without system-level feedback*

A group of individuals will succeed in becoming a single agent or agency to the extent that the members can coordinate with one another and replicate the performance of an individual agent. The question I address in this section is whether the members could form such an agent without any system-level feedback at any stage on where it stands in the space of commitments – on what it already believes and desires. Could they form an agent in a manner that mimics the performance of the simple robot or animal?

### *The issue*

In order to replicate the performance of a single agent, the members of a group will have to subscribe, directly or indirectly, to a common set of goals, plus a method for revising those goals, and to a common body of judgments, plus a method for updating those judgments. And in addition they will have to endorse a method of ensuring that one or more of them – or an appointed deputy – is selected to form and enact any intention, or perform any action, that those group attitudes may require. Or at least they will have to take steps that provide for these results, within feasible limits and under intuitively favourable conditions. Within such constraints, the group as a whole will have to be robustly disposed to achieve the minimal rationality that is required of any unreasoning agent; I put aside for the moment the question of whether the ability to reason weakens this requirement, as it does with individual human beings.

This analysis of group agency highlights the fact that if we take any grouping or organization of people to constitute an agent – in particular, an unreasoning agent – then not only must we hold that as a matter of fact its collective behaviour is representable as the pursuit of plausible goals in accordance with plausible beliefs. We must expect it to be robustly representable in such a manner. As we imagine various changes in its circumstances, for example, we must expect it to adjust so as to continue to act sensibly in pursuit of the goals attributed, or perhaps to alter the goals it pursues. This means that in looking at any group of people, the current evidence may leave it underdetermined whether or not it is a group agent in the sense defined. We may lack data on how it would adjust in counterfactual circumstances and may not be in a position, therefore, to say whether or not it is truly an agent.

How might a group be organized to meet agency requirements without any members having feedback on where it is already committed? Assume, plausibly, that the group agent does not emerge behind the backs of its members, so to speak, whether on the basis of some selectional pressure or the devious plan of some

organizing genius; assume, in other words, that the members are each aware of being part of a group. This assumption makes for a point of contrast with the individual, sub-personally rational agent, but the analogy with such an agent will still remain alive insofar as the following scenario obtains. The members act on the shared intention that they together should establish and enact a plan or constitution under which their efforts are coordinated suitably.<sup>8</sup> And, crucially, that plan or constitution does not require any of them to monitor the attitudes or actions of the group as a whole, gathering information on how things are going at the system level. The group operates without any system-level feedback on the attitudinal configuration generated under the constitution.

Were such an arrangement in place, then each member would play his or her local part and the global consequence would be the appearance of a pattern of rational attitude formation and enactment at the group level. The group's attitudes would form and unform in a rational pattern, and would rationally prompt and direct action. And this would happen without any members having to monitor or regulate how things transpire at the group level. The members would each look after their own local business and the global business of the group would take care of itself.

Under the constitution or arrangement envisaged some individuals would have to play a special role in triggering the group procedure needed to revisit a goal or judgment, in enacting a decision of the group on one or another issue, or on any of a number of fronts. But much of the business of the group would consist in the formation of the attitudes required for agency. How might those attitudes be formed, then, without anyone's having system-level feedback? I shall consider this question with particular reference to judgments. How might the members ensure that on every issue that comes before it, the group will form a suitable judgment? How, in particular, might they ensure this without having any feedback, and so without having any record, of the judgments that the group has already formed?

### *The voting proposal*

The salient method or strategy whereby a group might seek to achieve this result is by recourse to voting. The members of the group are its eyes and ears and voting

<sup>8</sup> I like an account of acting on a shared intention that is broadly in line with Bratman 1999. Set out in Pettit and Schweikard 2006, it stipulates that for individuals to act on an unforced, shared intention to X they must each intend that they together X; intend to do their individual part in a presumptively salient plan; believe that others each intend to do their part too; intend to do their part because of believing this; and enjoy a common form of awareness that those conditions are fulfilled. Those who act on such an intention may do so reluctantly or without relish, so that the account in the text does not suppose an equal commitment on the part of all members. It may be that some members do not even share properly in the intention but acquiesce in the shared intention of others; the acquiescence will mean that they play their allotted parts and will make them indistinguishable from those who endorse the intention.

will enable them each to register their evidence on any issue confronted.<sup>9</sup> Suppose the group has to determine whether  $p$  or  $q$  or  $r$ , then, establishing the answers on which members or their deputies are to act in pursuing the group's goals. The members will determine the group's judgment on such an issue by holding a vote on whether  $p$  or  $q$  or  $r$ , and then aggregating those votes according to some acceptable procedure. They may employ any of a variety of voting procedures, some centralized, others decentralized, for this purpose. Everyone may be invited to vote on every issue, whether in a process of majoritarian or non-majoritarian voting. Or issues may be segregated so that different subunits – at the limit, singletons – are given different questions to resolve, and one or another process of voting is adopted in each. Propositions will be presented to one or more members of the group and they will be treated as matters of belief or desire, depending on whether they command appropriate support.

The recourse to voting will rule out one salient possibility for establishing the attitudes of the group. This is that that members might put degrees of belief or probability together, thereby generating a system of probability for the group as a whole.<sup>10</sup> But this is not a serious loss. Voting can take place over probabilistic issues, such as whether it is more than likely than not that  $p$ , whether it is nearly certain that  $p$ , and the like. And in any case the possibility ruled out is quite fanciful. Even if people each have fine-grained degrees of belief on propositions considered by the group, it is not clear how they could know what they are, and so not clear how they could communicate them to one another in the fashion required. Such degrees of belief may show up in behaviour, particularly in dispositions to accept various gambles, but they need not be available to introspection (Harman 1986).<sup>11</sup>

<sup>9</sup> The members might deliberate with one another before voting, thereby pooling information and comparing their responses. But whether or not deliberation occurs, there is likely to be disagreement on any issue and so a need to determine the non-unanimously supported group position. See Pettit 2003b.

<sup>10</sup> This possibility is significant in the following respect. If the representations that individuals aggregate into group representations come in degrees, and the aggregation reflects those degrees, then there need not be any problem like the discursive dilemma discussed below; this presupposes attitudes of an on-off kind. But the aggregation may be subject to other difficulties. For examples of some difficulties, see Raiffa 1968; Hylland and Zeckhauser 1979, pp 220–37; I am grateful to Arthur Applebaum for drawing my attention to this work.

<sup>11</sup> Frank Ramsey describes a procedure whereby it is in principle possible for an interpreter to construct both a utility and probability function for an individual subject, on the basis of the subject's expression of binary preference as between different items and different gambles over items; for a summary description of the procedure see Pettit 2002, Part 2, Essay 2. Might it be applied with a group? No. Extracting a set of coherent binary preferences from a group will be subject to the same problem as that which arises, as we shall see later, with extracting a coherent set of judgments. The discursive dilemma that is used below to illustrate the problem with judgments can be extended readily to binary preferences; the group may prefer that  $p$ , that  $q$ , that  $r$ , and that not- $p$ & $q$ & $r$ .

The question before us, then, is whether the members of a group might be able to use a voting procedure – a procedure of voting without feedback – so as to determine the judgments of the group after a rational pattern. Could it organise itself on the basis of no-feedback voting? Could it relate in this manner to its own members and satisfy the conditions for rational agency at the group level?

There would certainly be problems under an uncoordinated arrangement that allowed one subunit to decide whether  $p$ , another whether  $q$  and a third whether  $p \& q$ ; these subunits might commit the group to an inconsistent set of judgments. But might a group guard against such inconsistency without any voting members having feedback on the commitments of the group as a whole? Might the subunits in a networked agency be coordinated so as to avoid the problem? Or might the members assemble so as to decide by centralized voting on each of the issues it has to resolve?

We may concentrate, without a serious loss of generality, on the assembly case. If a group cannot operate satisfactorily without feedback in this case, then it is unlikely to be able to operate without feedback in any other mode. The smaller subunits in any networked group will almost certainly have to face the same problem that arises for the assembly, if they are each charged with making judgments on logically connected propositions like ‘ $p$ ’, ‘ $q$ ’ and ‘ $p \& q$ ’. And in any event, there will be the extra problem of how to ensure that the different propositions they support do not form an inconsistent set.

### *The theory of judgment-aggregation*

Recent results on the aggregation of judgments are relevant to the issue we have raised and the possibility of a satisfactory voting solution. Those results reveal that there are severe constraints on how far a group can attain rationality in its judgments over logically connected issues and remain responsive in intuitively important ways to its members: specifically, responsive in ways that voting procedures automatically tend to implement. The results argue that the space for simultaneously ensuring both group rationality and individual responsiveness is very restricted.

There are broadly three respects in which we might expect that a group agent should be responsive to its membership. First, it should be robustly responsive, not just contingently so; the group judgments should be determined by the judgments of members, more or less independently of the form those judgments take. Second, the group should be inclusively responsive, not just responsive to a particular member – a dictator – and not just responsive to named individuals; otherwise it would fail to use its members as its eyes and ears, as epistemic considerations suggest it should do, as well as failing on a democratic count. Third, the group should be issue-by-issue responsive – if you like, proposition-wise responsive

(List and Pettit 2006) – with its judgment on any question being determined by the judgments of its members on that very question.

The recent results on the aggregation of judgments show that under a wide range of specifications of these responsiveness conditions, some quite weak, it is impossible to have a procedure for determining group judgments that both satisfies those conditions and ensures, over issues that are connected in one or another degree, that the judgments will be complete and consistent. One example of such a result is proved in List and Pettit 2002, and others have followed.<sup>12</sup> It demonstrates the relevant sort of impossibility under the following precisifications of the three responsiveness conditions.

- Robust responsiveness: the procedure works for every profile of votes among individuals (universal domain);
- Inclusive responsiveness: the procedure treats individuals as equal and permutable (anonymity); and
- Issue-by-issue responsiveness: the group judgment on each issue is fixed in the same way by member judgments on that very issue (systematicity).

The best way to communicate the lesson of these results is by way of an example. Take the connected set of issues: whether *p*, whether *q*, whether *r*, and whether *p&q&r*. Suppose that a group of three individuals, A, B and C, wishes to form a rational set of judgments over those issues, say because the question of how best to promote some group goals depends on the answers. And imagine now that the group follows a procedure of majority voting. Such a procedure will be robustly responsive in the sense that it will work under any profile of consistent member votes; it will be inclusively responsive in the sense that it gives everyone an equal vote; and it will be issue-by-issue responsive in the sense that it lets the judgment on every issue be determined by the votes of members on that very issue. And because of ensuring such full-blown responsiveness it may generate an inconsistent and therefore irrational set of judgments on the issues considered. The members may vote as follows.

	<i>p</i> ?	<i>q</i> ?	<i>r</i> ?	<i>p&amp;q&amp;r</i> ?
A	No	Yes	Yes	No
B	Yes	No	Yes	No
C	Yes	yes	No	No
A-B-C	Yes	Yes	Yes	No

<sup>12</sup> See in particular Pauly and Van Hees 2006 and Dietrich and List 2006. Notice that the three dimensions of responsiveness are not always reflected in a one-to-one fashion by three exactly corresponding conditions.

In a situation like this the group will face a ‘discursive dilemma’.<sup>13</sup> Either it secures majoritarian responsiveness to the views that are registered in the votes of members, in which case it will have to endorse the inconsistent set of judgments in the last row and fail to be rational. Or it ensures its own coherence, revising one of the judgments in the last row, in which case it will have to offend at least against majoritarian, issue-by-issue responsiveness; it will have to break with the majority view on ‘p’ or on ‘q’ or on ‘r’ or with the unanimously supported verdict on ‘p&q&r’.

#### *Back to the voting proposal*

Back now to the question of whether a group can attain rationality on the basis of a voting procedure, and do so without feedback to its members on where it is already committed as a group. Such a group will confront a growing number of issues over time: now whether p, as it might be; now whether q; and so on. I assume that two conditions will generally hold with such a group. First, the issues it addresses will tend, sooner or later, to form connected sets and to give rise to the sorts of problem addressed in the theory of judgment aggregation (List 2006). And second, it will be important for the group to form complete and consistent judgments over those issues, since it will generally form judgments only on a need-for-action basis; thus, important decisions are liable to be jeopardized by any failures of completeness or consistency.

Under these assumptions, any responsive voting procedure, whether it has a majoritarian or non-majoritarian or mixed character, is liable to lead the group into forming inconsistent sets of judgments. That is the lesson of the theorems on judgment aggregation. Suppose that the group faces a new issue that is logically connected with issues that have been resolved under prior votes. Lacking feedback on the prior resolutions, members won’t know what the group already judges and will have to vote blindly on the issue before them. Hence they are quite likely to vote in such a way that the group ends up with an inconsistent set of judgments. To return to the schematic example given, the members may vote that not-p&q&r, even when prior voting has committed the group to judgments that p, that q and that r.

#### *An unsatisfactory solution*

Is there any way out of this problem under the no-feedback stricture with which we have been working? There is one class of solutions available but I shall argue that they are unsatisfactory in a distinct manner.

<sup>13</sup> See Pettit 2001a; 2003c. The idea of the discursive dilemma is a generalization of the legal idea of a doctrinal paradox. See Kornhauser and Sager 1993. For an overview of the topic, and of other issues, see List 2006.

The most salient of these solutions is the sequential priority rule (List 2004).<sup>14</sup> This would organize issues so that whenever a group faces an issue on which its prior judgments dictate a resolution, then voting is suspended or ignored and the judgment recorded on that issue is the one that fits with existing judgments. There are a number of technologies whereby such an organization of issues might be realized without anyone in the group having to get feedback on the judgments already made by the group. And so the rule may seem to illustrate a procedure whereby the group might ensure its rationality without introducing feedback and so without activating any sort of group reasoning.

Assume that the group registers its views on 'p' and 'q' and 'r' before it confronts the issue of whether  $p \& q \& r$ , so that its existing commitments dictate that it should make a positive judgment. Under the organization postulated, the judgment recorded will be that  $p \& q \& r$ , independently of whether or how the members vote.<sup>15</sup> In following the rule, the group is bound to display a suitable sensitivity to meta-propositional constraints, in particular the constraint of formal consistency, but members need not have any feedback on its existing commitments. They may play their local parts blindly, without anyone keeping track of the group as a whole, yet the global upshot will be the formation of reliably consistent sets of judgments.

The reason why the sequential priority rule enables a group to be consistent, evading the difficulties identified in the impossibility results on judgment aggregation, is that while it forces the group to be robustly and inclusively responsive to its members, on intuitive interpretations of those conditions, it allows failures of issue-by-issue responsiveness. On any question where prior judgments dictate a certain line, the group may adopt a position that goes against the views of the members on that particular issue. The position taken will be driven by the positions that members take on other issues but not by their positions on that issue itself (List and Pettit 2006).

It should be clear that a group might avoid inconsistency by having all of its attitudes formed under the sequential priority rule, or suitable variants. But would it be a rationally satisfactory agent? I argue not.

<sup>14</sup> A variant on this procedure would divide issues into basic, mutually independent premise-issues and derived issues – this will be possible with some sets of issues, though not with all – and treat those judgments as prior, letting them determine the group's judgments on derived issues. See Pettit 2001b; List 2004.

<sup>15</sup> The group would reach the same judgment, if it worked with the (rather implausible) rule that would cast issues involving logically simple propositions as basic and that let other issues be fixed by its commitments on such premises; see the previous footnote. But those rules might come apart in other cases. Suppose that the group registers positive views on 'p' and 'if p, q' before it confronts the issue of whether q. The regular sequential priority rule would deem it to judge that q, even if members are disposed to vote against 'q', but the variant, premise-rule would have the group vote on q and, given commitments for 'p' and against 'q', would deem the group to reject 'if p, q'.

The problem is that while such a group would reliably achieve consistency in the judgments it forms, it would be entirely inflexible in its responses and potentially insensitive to the overall requirements of evidence. When I realize that some propositions that I believe entail a further proposition, the rational response may well be to reject one of the previously accepted propositions rather than to endorse the proposition entailed. Those are the undisputed lessons of any coherence-based methodology and the group that operates under a sequential priority rule, or under any variant, will be unable to abide by them; it will not be robustly sensitive to the requirement of attitude-to-evidence rationality.

The evidential insensitivity of the sequential priority rule is apparent from the path-dependence it would induce.<sup>16</sup> One and the same agent, with access to one and the same body of evidence, may be led to form quite different views, depending on the order in which issues present themselves for adjudication. The group agent that follows the rule will be required to respond to essentially conflicting bodies of testimony – conflicting majority judgments among its members – without any consideration as to which judgment it seems best to reject. It will be forced by the order in which issues are presented not to give any credence to the judgment its members may be disposed to support on the most recent issue before it. And this, regardless of the fact that often it will be best to reject instead a judgment that was endorsed at an earlier stage.

The path-dependence imposed under the sequential priority rule can be illustrated with the now familiar, schematic example. Let the group confront the three atomic issues before it faces the issue of whether  $p \& q \& r$  and it will judge that  $p \& q \& r$ . Let it confront that compound issue earlier, say before the issue of whether  $r$ , and it will judge that  $\text{not-}p \& q \& r$ . When it comes to the issue of whether  $r$ , the rule will force it in consistency with prior commitments to judge that  $\text{not-}r$ . Thus the judgments with which it ends up will vary with the order in which the judgmental issues are presented. And yet, on any intuitive conception of evidential rationality, the order of presentation ought not to be relevant in this manner.

### 3. *Group agents with system-level feedback*

These observations suggest that the search for a no-feedback constitution that would have groups perform on the model of rational but unreasoning agents is

<sup>16</sup> For ways of mitigating the effects of path-dependence see (List 2004). The evidential insensitivity of the sequential priority rule appears in other ways too. Suppose, for example, that the A-B-C group judges in favor of 'p', 'q' and 'r' and is then tested on a conjunction of those propositions with an incontestable, empirical truth: say, ' $p \& q \& r \& e = mc^2$ '. A majority will vote against the compound proposition, since it is disposed to vote against ' $p \& q \& r$ ' alone; and that judgment can stand since it is not consistent with the earlier judgments. But what now will the rule dictate about the group's judgment on ' $e = mc^2$ ', if it is next faced with that issue on its own? The group will have to be deemed, preposterously, to judge that it is not the case that  $e = mc^2$ . I am grateful to Caspar Hare for discussion on this point.

likely to be a futile enterprise. More specifically, the search is likely to be futile in any scenario where the groups face logically connected issues, and where they aspire to achieve a significant degree of responsiveness in the dimensions given. No doubt there are other possible approaches, on a par with the sequential priority rule, which would have a group seek to operate without feedback. But it is hard to see how any could deal satisfactorily with the evidential problem raised for that rule. For how could a group display evidential sensitivity across a number of issues without keeping track of those issues, and of the responses that they each elicited from the group?

### *The straw-vote procedure*

This is not bad news for group-formation as such, however. For it turns out that once we allow members to have feedback on where a group is committed, and once we make arrangements for that feedback to have an effect, then group rationality ceases to be so elusive. The possibility can be illustrated with what we may call a straw-vote procedure. This is a procedure that a group might implement in a centralized assembly where every member votes on every issue; but it can also stand in for analogous procedures with groups of a networked, non-assembling kind.

The idea in the straw-vote procedure is to have members take a straw vote on every new issue; consider whether the judgment supported is consistent with existing judgments; and, if it is not, revise one of its conflicting judgments so as to ensure consistency. The judgment revised is not necessarily the judgment just supported in voting, as it would be under the sequential priority rule. If that is what seems evidentially more appropriate, one of the earlier judgments may be revised instead.

The straw-vote procedure might be detailed in this set of instructions to members of the group:

- (1) With every issue that comes up for judgment take a majority vote on that issue and, as issues get progressively settled in this way, keep a record of the accumulating body of judgments.
- (2) With every new issue that is voted on, check to see if the judgment supported is consistent with the existing commitments of the group.
- (3) If majority voting generates an inconsistency, treat the judgment supported and the set or sets of judgments with which it is inconsistent in the record as candidates for reversal.
- (4) Identify the problematic judgments – say, the judgments that p, that q, that r, and that not-p&q&r – and address the question of how to resolve the inconsistency.

- (5) Take a vote on where it would be best to revise the judgments: whether, in the simple example considered, it would be best to revise the judgment that  $p$ , that  $q$ , that  $r$ , or that  $\text{not-}p\&q\&r$ .
- (6) Take the proposition identified in this way, and hold another vote on how the group should judge that proposition.
- (7) If the group reverses its previous judgment, treat the new verdict on that proposition as the one to be endorsed by the group.
- (8) If the previous judgment is not reversed in that vote, go back to stage 3 and try again.
- (9) If it appears that there is no prospect of success in this process, try to quarantine the inconsistency, and the area of decision it would affect, so that it does not generate problems elsewhere.
- (10) If this quarantining is not possible, perhaps because the area of action affected is important to the group's aims, there is no alternative but to declare defeat on the issues under consideration, even perhaps to disband.

The procedure outlined in these instructions is not a particularly surprising proposal and we can well imagine a group adopting it. The procedure requires a group to treat the appearance of every new, connected issue as introducing a problem of how best to judge, not just over that particular issue, but over that issue together with the previously considered, logically connected issues. The group will recognize the inconsistency of the judgments that were elicited from it in separate votes and will resolve that problem of inconsistency insofar as members can converge on one presumptively best set of judgments overall. The set of judgments it adopts at the end of the exercise will be fixed by the pattern of member judgments over those issues but in a way that violates issue-by-issue responsiveness in the minimal measure required for group rationality.

The straw-vote procedure shows how feedback can make group rationality accessible in circumstances where a no-feedback rule would lead to problems. But the strategy it illustrates not only involves the use of feedback; it also displays the exercise of a sort of group reasoning.

The requirement that the group consider every proposition that is supported by a straw vote for whether it is consistent with propositions already endorsed amounts to a requirement that it practice semantic ascent and look to relations between propositions. And the requirement that it adjust to a judgment that there is an inconsistency amongst the propositions endorsed amounts to a requirement that it respond appropriately to any observed irrationality, removing the inconsistency while also respecting other constraints of rationality like that of evidential sensitivity. If it follows the straw-vote procedure, then the group will count as a reasoning subject in the image of the reasoning subjects that we individuals constitute. It will exercise a sort of control over its own processes of judgment-

formation that resembles the personal control associated with individual reasoning. The members will act together in implementing an intentional exercise of group control. And they will do this in respect of themselves considered as a unified centre of attitude formation and enactment.

There is no saying how exactly the members of a group will adjust so as to rectify perceived inconsistencies, of course, or other forms of irrationality; that is a matter of variable psychology. But the fact that they can be relied on to adjust in such a manner means that the group can go through the exercise of reasoning with confidence that it may prove beneficial. This should not be surprising, since the situation is similar on the personal front. There is no saying how I as an individual may adjust, by grace of my subpersonal nature, to one or another perceived inconsistency. But the fact that I can be relied on to adjust in a suitable manner means that I can go through the exercise of reasoning with a similar confidence that it will bear fruit. As it is on the group front, so it is on the personal.

Although I have used the straw-vote procedure to illustrate how group reasoning may help to resolve issues of group rationality, it should be clear that there are many variations possible on that particular approach. These variations will apply, not just with the assembled group, but also with groups that are networked out of more or less independent subunits. The simplest analogue would be a networked group in which one subunit, with the authorization of the group as a whole, plays the role that is played by the assembly under the straw-vote procedure. This unit would review the commitments that would materialize on the basis of voting in one or more other subunits and revise the emerging commitments in the minimal measure require to secure consistency in the commitments of the group as a whole. It would serve in a role analogous to that served by the highest court in a system of judicial review.<sup>17</sup>

Under this variant system, as under the straw-vote procedure, the group can be said to reason about what judgments to support, rather than relying on rational judgments to bubble up without any need of monitoring. It intentionally conducts an exercise – in this case, via an authorized subunit – that meets the specifications for reasoning. It asks meta-propositional questions out of a desire to identify the answers. And it does this in confidence that the answers will have an appropriate impact on the processes whereby group attitudes form and have an effect.

<sup>17</sup> Stearns 2000 draws attention to a way in which a group might maintain its rational configuration, thanks to the unauthorized intervention of some members. Take the case of a collegial court that has to make judgments on related issues; say, the case where it has to judge in matters of tort on whether there was harm done, whether there was a duty of care, whether there was negligence, and so, whether the defendant was liable; these issues relate under legal doctrine like ‘p’, ‘q’, ‘r’ and ‘p&q&r’. It is possible for a court as a group to vote in a case like this for each of the atomic propositions but against the compound. And in some such cases, there is evidence that at the last minute one or another judge votes inconsistently with his or her commitments on the atomic issues in order to preserve the rational configuration of the court.

*The struggle for group rationality*

But though group rationality may be achieved through recourse to reasoning, it is sometimes achieved only with difficulty. Groups may see what is required in rationality of them and yet, like individual agents, fail or falter on this front. They may display group *akrasia* – collective weakness of will (Pettit 2003a).

Imagine a non-commercial academic journal with an editorial committee of three members that resolves all the issues it faces by majority vote. Suppose that the committee votes in January for promising subscribers that there will be no price rise within five years. Suppose that it votes in mid-year that it will send papers to external reviewers and be bound by their decision as to whether or not to publish any individual piece. And suppose that in December it faces the issue as to whether it should be just as prepared to publish technical papers that involve costly typesetting as it is to publish other papers: whether it should treat them as equal. The earlier votes will argue against its being prepared to do this, since a rise in the number of technical papers submitted and endorsed by reviewers – endorsed, without an eye to overall production costs – might force it to renege on one or other of those commitments. But nonetheless a majority may support the even-handed treatment of technical papers, without any individual being in any way irrational. The members of the committee may vote as follows.

	Price freeze?	External review?	Technical papers equal?
A.	Yes	No	Yes
B.	No	Yes	Yes
C.	Yes	Yes	No

The group now faces a hard choice of broadly the kind we have been discussing. Suppose that the members operate with the straw vote procedure and that they agree that the issue on which the group should revise its view is that of whether to treat technical papers on a par with other papers; they may vote unanimously that it is impossible to revise its position on either of the other issues, perhaps because the editorial position on those questions has already been made public. How, then, may we expect the consequent vote to go?

If members are individually dedicated to the group and are in no way tempted to defect from what it requires of them, then of course they will each vote for offering less than equal treatment to technical papers; they will reverse the previous group position. A group whose members were dedicated in this way would operate like a perfectly virtuous agent, always spontaneously supporting what the balance of available reasons requires of the group. But not all members need be so devoted to the group in which they figure; and when something less than full collective devotion is on offer, then it may prove very difficult for members to get their act together and ensure that the group lives up to the considerations that it endorses.

Take the majority, A and B, who originally supported an open policy on technical papers. That majority may remain individually and stubbornly inclined to support the equal treatment of technical papers. We can imagine them turning their eyes from the group as a whole, and sticking to their votes when the issue is raised again. We can imagine them refusing to hear the call of the group and acting like encapsulated centres of voting who are responsive only to their own modular prompts. As we imagine this, we envisage the group failing to reverse its judgment on an issue where every member of the group thinks it is desirable to reverse judgment. The recalcitrant majority in this sort of case might be moved by a more or less selfish inclination or identification, being technically minded themselves, or they might be moved by a sense of fairness towards those who would be disadvantaged; personal virtue is as likely as personal vice to source recalcitrance towards the collectivity.

Could it really be rational, however, for the recalcitrant members to stick to a deviant pattern of voting, whether out of individual bias or virtue? I don't see why not. They would satisfy their private motives, partial or impartial, by doing so. And they might individually expect to get away with such voting, being outvoted by the others; they might each expect to be able to free-ride. Or they might hope that even if a majority remains recalcitrant, this will not cause problems: there will not be a deluge in the number of technical papers submitted and accepted, and the committee can get away with holding by all of the three commitments involved.

The possibility of people remaining encapsulated in their personal identities in this way, and the danger that that holds out for the survival of the group, shows that it is essential in general that the members should break out of their capsules. If the group is to evolve as a centre of agency, with a capacity to be held responsive to the demands of consistency, then it must be able to discipline members into supporting only certain patterns of judgment. And if a group is to have that capacity, then its members must be willing to put their own views aside and identify with the group as a whole, whether spontaneously or under the impact of institutional incentives. They must be ready to reason and act from the perspective of that common centre.

What should we say about the editorial group, however, if it just fails to get its act together and lives on the wild side, exposed to a constant danger of bankruptcy? Should we think that it is not a group agent after all but only a collection of individuals, like those who live in the same zip code, who should not be held to expectations of consistency and the like? Or should we think of it as a group agent that is failing on this front and that can be held up to criticism for the failure?

Clearly we would think the latter; and, equally clearly, we should do so. The reason we would hold it to expectations of consistency is that the tasks entrusted to the group, and embraced by it, mark it off from the zip-code population, giving it the cast of an agent. And the reason we should hold it to these expectations is that

even while it breaches a constraint of rationality, it can acknowledge the relevant standards, like any symbol-using, reasoning creature, and can display its agency by doing so. We saw earlier that the reasoning human being can vindicate his or her agency, unlike the simple animal, even while failing to conform to certain standards of rationality. The same is true with the reasoning groups that human beings form.

### *Conclusion*

Jenann Ismael (forthcoming) makes a useful distinction between self-organizing and self-governing systems and it may be useful, in conclusion, to summarize the upshot of the argument in these terms.

Consider the sort of entity that generates a global, agential profile as a result of the locally stimulated responses of its members or parts or subsystems. Take for example the insect colony in which individual insects combine to produce a field of chemical stimuli such that when individuals respond to that field – for each insect, to the local stimuli provided by that field – the result is a coherent, self-organizing pattern of action. At any moment in the life of the colony the insects act together in different roles to advance a coherent set of goals in a coherent way. And the inputs they make in doing so generate a field of stimuli that elicits suitable responses at the next stage and continues into the future to support the agential profile. A system of this kind is self-organizing. Its behaviour as a whole, which displays a striking coherence, is a function of the behaviour of the sub-units. Yet those units produce that behaviour as a result of each of them marching to its own drum, without any of them processing system-level feedback.

The self-governing agent, in contrast with this merely self-organizing system, gives some sub-units a special, regulatory role. They keep track of the performance of the system – this is at least in the spirit of Ismael's account – and intervene at certain junctures in order to ensure that it displays the character of a rational agent.<sup>18</sup> They monitor the system as a whole, gathering feedback on its dispositions to behaviour, whether by observation or by using a predictive or simulative device (Grush 2004). And they manage the system by making interventions that correct for ways in which it may be disposed to act out of rational type.

The lesson of this paper is that in more or less standard conditions a collectively rational and individually responsive group agent will have to be self-governing rather than self-organizing. This lesson holds, more specifically, for conditions where logically connected issues present themselves for resolution,

<sup>18</sup> The systems considered here are agents but not all self-organizing or self-governing systems have to be agents. The free market in which individually self-seeking agents sustain equilibrium prices is not an agent, yet it ordinarily counts as a self-organizing system. Nor is the command economy an agent, though it should probably count as a self-governing system.

rational agency requires a complete, consistent set of answers, and it is important that those answers are rationally sensitive to the overall evidence available. Whether it has an assembly or a network character, the group agent will have to organize itself for such conditions so that some or all of its members can keep track of its accumulating judgments and take steps to guard against the onset of inconsistency. The lesson may not have the full-dress credentials of an a priori necessity but it is as safe a bet as we are likely to be able to identify in this area.\*

## REFERENCES

- ANSCOMBE, G. E. M. 1957, *Intention*. Oxford: Blackwell.
- BILGRAMI, A. 1998, 'Self-knowledge and Resentment', in: B. S. C. Wright, and C. Macdonald, eds, *Knowing Our Own Minds*, Oxford: Oxford University Press, pp. 207–41.
- BRATMAN, M. 1999, *Faces of Intention: Selected Essays on Intention and Agency*, Cambridge: Cambridge University Press.
- BROOME, J. 2007, 'Wide or Narrow Scope?' *Mind* **116**, pp. 359–70.
- BURGE, T. 1998, 'Reason and the First Person', in: B. S. C. Wright, and C. Macdonald, eds, *Knowing Our Own Minds*, Oxford: Oxford University Press.
- CARROLL, L. 1895, 'What the Tortoise said to Achilles', *Mind* **4**, pp. 278–80.
- CLARK, A. 1999, Leadership and Influence: The Manager as Coach, Nanny and Artificial DNA, in: J. Clippinger, ed., *The Biology of Business: De-coding the Natural Laws of Enterprise*, San Francisco: Jossey-Bass, pp. 46–66.
- DAVIDSON, D. 1984, *Inquiries into Truth & Interpretation*, Oxford: Oxford University Press.
- DENNETT, D. 1994, 'Learning and Labelling: A Commentary on A. Clark and A. Karmiloff-Smith', *Mind & Language* **8**, pp. 540–48.
- DENNETT, D. C. 1969, *Content and Consciousness*, London: Routledge.
- DIETRICH, F. and LIST, C. 2006, 'Arrow's Theorem in Judgment Aggregation', *Social Choice and Welfare* **29**, pp. 19–33.
- EVANS, G. 1982, *The Varieties of Reference*, Oxford: Oxford University Press.
- FODOR, J. 1975, *The Language of Thought*, Cambridge: Cambridge University Press.
- GRUSH, R. 2004, 'The Emulation Theory of Representation: Motor Control, Imagery, and Perception', *Behavioral and Brain Sciences* **27**, pp. 377–96.
- HARMAN, G. 1986, *Change in View*, Cambridge, MA: MIT Press.
- HYLLAND, A. and ZECKHAUSER, R. 1979, 'The Impossibility of Bayesian Group Decision Making with Separate Aggregation of Beliefs and Values', *Econometrica* **47**, pp. 1321–36.
- ISMAEL, J. forthcoming, 'Selves and the Limits of Self-organization', *Minds and Machines*.
- KOLODNY, N. 2005, 'Why be Rational?' *Mind* **114**, pp. 509–63.
- KORNHAUSER, L. A. and SAGER, L. G. 1993, 'The One and the Many: Adjudication in Collegial Courts', *California Law Review* **81**, pp. 1–59.

\* This was presented as *The Dialectica Lecture 2006* at the *Gesellschaft fuer Analytische Philosophie* meeting in Berlin, September 2006. It also had an airing at the Inter-University Center, Dubrovnik, the University of Sydney, the Massachusetts Institute for Technology, Brown University, Queen's University, Belfast and Stanford University. My thanks for the many useful comments I received in discussion at those events, in particular to respondents and co-symposiasts: Katie Steele, Jenann Ismael and John Sutton in Sydney, Avi Tucker in Belfast. The paper is related to ongoing work with Christian List and apart from the debt that I owe to our incorporated, shared self, I am greatly indebted to him for detailed criticisms of an earlier draft.

- LEVI, I. 1991, *The Fixation of Belief and its Undoing: Changing Beliefs through Inquiry*, Cambridge: Cambridge University Press.
- LIST, C. 2004, 'A Model of Path-Dependence in Decisions over Multiple Propositions', *American Political Science Review* **98**, pp. 495–513.
- LIST, C. 2006, 'The Discursive Dilemma and Public Reason', *Ethics* **116**, pp. 362–402.
- LIST, C. and PETTIT, P. 2002, 'Aggregating Sets of Judgments: An Impossibility Result', *Economics and Philosophy* **18**, pp. 89–110.
- LIST, C. and PETTIT, P. 2005, 'On the Many as One: A Reply to Kornhauser and Sager', *Philosophy and Public Affairs* **33**, pp. 377–390.
- LIST, C. and PETTIT, P. 2006, 'Group Agency and Supervenience', *Southern Journal of Philosophy* **45** (Spindel supplement), pp. 85–105.
- LIST, C. and PETTIT, P. forthcoming, 'Agency Incorporated: The Possibility, Design and Status of Group Agents'.
- MCGEER, V. and PETTIT, P. 2002, 'The Self-regulating Mind', *Language and Communication* **22**, pp. 281–99.
- PAULY, M. and VAN HEES, M. 2006, 'Logical Constraints on Judgment Aggregation', *Journal of Philosophical Logic* **35**, pp. 569–585.
- PETTIT, P. 1993, *The Common Mind: An Essay on Psychology, Society and Politics*, paperback edition 1996, New York: Oxford University Press.
- PETTIT, P. 2001a, *A Theory of Freedom: From the Psychology to the Politics of Agency*, Cambridge and New York: Polity and Oxford University Press.
- PETTIT, P. 2001b, 'Deliberative Democracy and the Discursive Dilemma', *Philosophical Issues* (supp. to *Noûs*) **11**, pp. 268–99.
- PETTIT, P. 2002, *Rules, Reasons, and Norms: Selected Essays*, Oxford: Oxford University Press.
- PETTIT, P. 2003a, 'Akrasia, Collective and Individual', in: S. Stroud and C. Tappolet, eds, *Weakness of Will and Practical Irrationality*, Oxford: Oxford University Press.
- PETTIT, P. 2003b, 'Deliberative Democracy, the Discursive Dilemma, and Republican Theory', in: J. Fishkin and P. Laslett, eds, *Philosophy, Politics and Society Vol 7: Debating Deliberative Democracy*, Cambridge: Cambridge University Press, pp. 138–62.
- PETTIT, P. 2003c, 'Groups with Minds of their Own', in: F. Schmitt, ed., *Socializing Metaphysics*, New York: Rowan and Littlefield.
- PETTIT, P. and SCHWEIKARD, D. 2006, 'Joint Action and Group Agents', *Philosophy of the Social Sciences* **36**, pp. 18–39.
- RAIFFA, H. 1968, *Decision Analysis*, New York: Addison-Wesley.
- SEARLE, J. R. 1983, *Intentionality*, Cambridge: Cambridge University Press.
- SMITH, M. 1987, 'The Humean Theory of Motivation', *Mind* **96**, pp. 36–61.
- STEARNS, M. L. 2000, *Constitutional Process: A Social Choice Analysis of Supreme Court Decision Making*, Ann Arbor: Michigan University Press.