3 The reality of group agents

Philip Pettit

Introduction

Human beings form many sorts of groups but only some of those groups are candidates for the name of agent. These are groups that operate in a manner that parallels the way that individual agents behave. They purport to endorse purposes, to form representations and to act for the satisfaction of those purposes according to those representations. And, building on those purported capacities, they make commitments and incur obligations, they rely on the commitments of others and claim rights against them. As candidates for group agents of this kind we might cite the partnership or the corporation, the church or the political party, the university or the state.

But are such entities truly agents? Or are they mere simulacra of agents? Do they replicate the agency of individual human beings? Or do they merely simulate it? That is the question I address in this chapter.[1]

The chapter is in three sections. In the first section I set out the requirements that systems of any kind must fulfill if they are to count as agents. In the second I look at the way in which individuals might seek, on the basis of shared intention, to form a group agent. And then in the final section I show how the sort of entity they construct in that way can meet the requirements given and count as a genuine agent.

[1] The question is the third of three questions that I take to be crucial in social ontology. The first is the question between individualism and non-individualism and bears on how far social regularities undermine the agential autonomy that we ascribe in folk psychology to individual human beings. The second is the question between atomism and non-atomism and bears on how far the psychology of individual human beings non-causally or superveniently depends for some of its important features on those individuals having social relations with one another. Those two questions are addressed in Pettit 1993, where I argue for individualism but against atomism. The question considered here divides singularism from non-singularism, as we might call the rival approaches, and bears on how far groups of individuals can constitute agents on a par with individuals.

1 The requirements of agency

1.1 Agential behavior

It is possible, on the face of it, for something that is not strictly an agent to display agential behavior. We can imagine finding evidence in the behavior of a system that it is an agent, but then overruling that evidence on the basis of further information. So what is it that we should expect in a system's behavior if that behavior is perfectly agential: if it is to do as well as possible in constituting evidence, however defeasible, that the system is an agent?

This is probably the easiest question in the theory of agency, for almost all sides are agreed that behavior manifests agency to the extent that it instantiates what we may describe as a purposive–representational pattern. Let the behavior of a system be understood, not just as the behavior it actually manifests, but as the behavior that it displays across the fullest range of possible scenarios, actual and counterfactual: that is, as the behavior it is disposed to display in such scenarios. That behavior might consist in an entirely random collection of behavioral pieces, without any rhyme or reason to them. But if it is agential in character, then it will be patterned in a way that links it to certain purposes and certain representations (Dennett 1991).[2]

The candidate purposes of the behavior will be revealed by the outcomes that it reliably achieves. And given an assignment of purposes, the candidate representations of the system will be revealed by the adjustments it makes in pursuit of those purposes, as it registers the nature of the different situations it confronts. The behavior of a system will display a purposive–representational pattern, and exemplify agential behavior, to the extent that there is a suitable set of purposes and representations – ideally, a single set – such that the behavior promotes those purposes according to those representations (Stalnaker 1984). The behavior involves the adoption of means for realizing those purposes that will tend to be effective if the representations, being suitably responsive to situations, are correct.

In explaining the notion of a purposive–representational pattern, I abstract from the extent to which the pattern is enriched or impoverished. It should be clear that agents may differ in how far the purposes they pursue, and the representations they form, relate to the here and now as distinct from the spatially and temporally distant; refer to the

[2] The spirit of this chapter is broadly congenial to the views with which Dennett is associated. For a more explicit connection between those views and a realistic model of group agency, see Tollefsen 2002.

presumptively actual as distinct from the counterfactual and possible; engage with a limited, as distinct from an open, range of particulars and properties; quantify over abstract entities like numbers as well as more concrete items; and bear on particular matters of fact as distinct from generalities and laws. In the remainder of the discussion I shall continue to abstract from this issue, since the argument goes through independently of how rich the domain of agential behavior happens to be.

While abstracting from the richness of the purposive–representational character of agential behavior, I focus, without using the word, on the rationality of the pattern. A pattern of behavior will be agential to the extent that, in ordinary terms, it is rational. The purposes and representations must make sense in an attitude-to-evidence dimension, being responsive to the different features of the situations it confronts; they must make sense in an attitude-to-action dimension, being organized to generate whatever interventions are instrumentally required by the purposes of the system according to its representations; and for a mix of evidential and instrumental reasons, they must make sense in an attitude-to-attitude dimension, being more or less consistent with one another, for example, and even perhaps mutually supporting.

Given a conception of agential behavior we can now ask after what we should expect of a system that is to count as an agent. Presumably a system will count as an agent just to the extent that it relates in a certain way to an agential pattern of behavior. But what precisely is the relationship required?

1.2 Agential behavior and agency

The simplest theory of agency would say that a system is an agent just to the extent that it instantiates an agential pattern in its behavior. There are purposes and representations that it is independently plausible to ascribe to the system – this constraint may be variously interpreted[3] – and the behavior of the system generally promotes those purposes according to those representations. This is classical functionalism or dispositionalism. Let a system be disposed, no matter on what basis, to display a plausible, agential pattern of behavior. Or, to be more realistic, let it be disposed in general to display such a pattern; naturalistic limitations are bound to make for occasional failure. To the extent that the

[3] Not only are there different views as to what is required to remove mystery on this front; the views also differ on how restrictive the requirements are. Those in the "teleosemantic" camp, for example, hold that ascriptions of reprentations have to satisfy requirements of an evolutionary kind and that these are quite demanding. See Millikan (1984).

agent displays such a disposition, it will count as a center of agency. To be an agent, on this approach, is simply to function as an agent: to pass as an agent on the behavioral front.

There are three broad alternatives that compete in the current literature with a purely functionalist theory of agency. Each would add a further clause to the behavioral condition. And, typically, each would downgrade that first condition in the process: it would allow that in the presence of the further clause, a system may count as an agent without fully satisfying the behavioral condition.

The first of these alternatives would stipulate that in order to count as an agent proper, a system has to be composed of the sort of stuff or substance or material out of which paradigmatic agents – perhaps human beings, perhaps humans and other animals – are composed. It would suggest that only systems that are made up of that same stuff, or perhaps stuff of a broadly similar sort, can constitute agents. A good example is the Cartesian account that takes human beings to be composed of a non-physical kind of thinking substance, and that makes the presence into a prerequisite of agency.

The second alternative would stipulate that in order to count as an agent, a system does not just have to instantiate the dispositions that constitute relevant purposes and representations. Those dispositions have to be realized within the system in a certain psychological pattern: according to a certain architectural design, for example, or with a certain conscious, qualitative feel. The requirement is not just that the dispositions have to evolve in interaction with the environment, and not be rigged in advance (Block 1981); that is already guaranteed by the assumption that representations should be responsive to situational features. It is the much more problematic assumption that any genuine agent has to display something like the architecture of classical computing: or the conscious life of a biological organism like one of us (Searle 1983, Fodor 1975).

Finally, the third alternative would stipulate that in order to count as an agent a system has to have the capacity, not just to conform broadly to a pattern of agential behavior, but to achieve a critical, ratiocinative perspective on that pattern (Davidson 1980). The system has to be able to identify some of the demands imposed by the pattern as regulative or normative requirements, and to let the identification of those demands reinforce conformity and underpin the recognition of non-conformity as a failure. It has to be able to do the sort of thing that we do when we reason.

When we reason theoretically we don't just form representations in accordance with *modus ponens* or any such rule; we don't just come to

believe that q, on coming to believe that p and that if p, then q. We also recognize that "p" and "if p, then q" are true and that their truth ensures the truth of "q". And then this further recognition reinforces an independently triggered belief that q makes up for the lack of an independent trigger, prompting the belief to form for the first time. Again, when we reason practically we don't just form intentions in accordance with the means-end principle; we don't just come to intend that X, on coming to form the intention to G and the belief that the only way to G is to X. We also recognize that G is a goal to be realized, and that realizing G entails realizing X. And then this further recognition reinforces the independently triggered intention to X or makes up for the absence of an independent trigger. Let a system have the capacity to reason and it will be able to recognize some of the demands imposed by agential pattern – say, the demand to intend any necessary means towards an intended goal – and this recognition will play a role in supporting compliance with that demand.

If we impose the first or second of our three extra conditions on agency, then we cannot admit the reality of group agents. Group members may act and speak as if a single representational and purposive mind lies at the origin of the group's actions and utterances. They may even manage to display a more or less perfect form of agential behavior. But on either of the first two alternatives, what the members achieve together can only be a show of agency, not its substance. The behavior will not be produced by the appropriate stuff or according to the appropriate sort of processing.

I do not think that this need not concern us unduly, since both of these alternatives to pure functionalism seem dubious. We ascribe agency to one another in light of our behavior and without giving any obvious thought to the basis on which that behavior is produced. Thus the conception of agency that we deploy in mutual interpretation – and, plausibly, in the interpretation of many other animals – does not necessarily presuppose anything about the stuff or the process in which agency materializes (Jackson and Pettit 1990a). It is more purely functional than either of those approaches suggests.

What should we say about the divergence between pure functionalism and the third alternative? Here it is possible to be ecumenical (Pettit 1993, chs. 1–2). The system that instantiates an agential pattern of behavior, at least in the absence of perturbation, can count as a regular or non-ratiocinative or non-critical agent. The system that is capable, in addition, of recognizing and responding to the demands of agential behavior can count as a special, ratiocinative or critical agent. This ecumenism is attractive because it enables us to countenance many

non-human animals as agents, which we surely have reason to do, and yet at the same time to acknowledge an important gap between such animals and human beings like you and me. Most of the time we human beings may operate as agents in the unthinking manner of other animals. But we sometimes adopt the ratiocinative pose exemplified by Rodin's sculpture of *Le Penseur*. Moreover, we are always ready to resort to such a perspective when the red lights go on. And we can rely on the possibility of such resort to help keep us in line with the demands of agential pattern: critical reflection can guard us against incautious or sloppy processing at the more spontaneous level.

The groups whose claim to agency we will be exploring, as we shall see in the next section, are groups that adopt a critical perspective on agential pattern, like individual human beings and unlike other animals. These groups are capable of going through something like a process of reasoning and of using that process to guard against failure. This means that even those who refuse the title of agent to systems that are not capable of a critical perspective on agential pattern can find the argument that follows congenial; nothing in the argument presupposes the truth of the view they oppose.

1.3 Testing for agency

The upshot of this discussion is that if we are to test a system for whether it is an agent, then we have to look to see if it broadly displays a purposive–representational pattern of behavior, whether on a critically informed or uninformed basis. What issues should we focus on when determining whether a system really does display such a pattern? I distinguish three core sub-questions that are crucially involved in the general question. I describe these respectively as the issues of systematic perturbability; contextual resilience; and variable realization.

Systematic perturbability. There are two quite different ways in which a system might be perturbed in the display of an agential pattern. There might be systematic factors such that in their absence, there is little or no noise; most perturbation derives from those factors and only a little materializes as random disturbance. On the other hand more or less all of the perturbation to which the system is subject might appear as random disturbance: as an unpredictable breakup of the pattern, like the spasmodic trembling of an otherwise steady hand.

Systematic perturbability is easy to square with agency, unsystematic perturbability not. If an entity behaves like an agent, subject to random, unsystematic departures from form, it is easy to think that the appearance of agency may be an illusion. But if there are established,

independently intelligible sources of perturbation and in their absence the system behaves like an agent, with only a very little failure, then it is harder not to take the appearance of agency seriously. There will be no temptation to think that the system is an aleatoric device that just happens to project a broken image of agency in a sequence of chance events. Given the systematic character of the perturbers, it will be natural to take them as constraints under which the system was designed or selected for fidelity to a purposive–representational pattern.

1.4 Contextual resilience

A pattern may be very reliable under certain boundary or contextual conditions. But those conditions can be very demanding, so that even if the pattern is almost certain to obtain so long as the conditions are fulfilled, it will break up under even a slight variation in those conditions. Consider the sequential pattern generated under John Conway's game of life by the initial figure in which, roughly, there are four squares placed close to one another to form a larger square on a grid; this is called the exploder pattern.[4] The kaleidoscopic, explosion-implosion sequence that is generated from that starting figure is entirely reliable. But it is fragile or non-resilient, in the sense that it will fail if even a single box in the grid, adjacent to the starting figure, is also filled in.

We cannot be very confident that a system is an agent if it is inflexible and fragile in this manner. A system will be an agent insofar as it is disposed, on an uncritical or critical basis, to display a purposive–representational pattern of behavior. But as this disposition is tied more closely to a suitable context – as it becomes the disposition to display that pattern in context x and only context x – it becomes more and more implausible to describe the pattern generated in terms that abstract from that context. The disposition will amount to nothing more than a reactive habit that is tailored to cues provided in that particular situation.

Consider the Sphex wasp that Daniel Dennett (1979) discusses. This wasp brings its eggs to the edge of a hole that it has found or dug, enters the hole to make sure that it is still provisioned with the paralysed prey that it has previously deposited there, then comes up and takes the eggs back into the hole. But it turns out that if the eggs are moved even a little bit away from the edge while the wasp is in the hole, then the wasp goes through the whole routine again and that it can be forced by this intervention to repeat the exercise an indefinite number of times. The failure here prompts us to recognize that the wasp is not displaying the

[4] See www.bitstorm.org/gameoflife/

pattern of ensuring that its eggs are placed in a suitable hole, as if it were focused in an agential way on that abstract purpose. The pattern that it is displaying is tied inflexibly to the context in a way that makes that ascription of purpose unwarranted.[5]

1.5 Variable realization

Suppose that a system displays fairly systematic perturbability and a high degree of contextual resilience in the generation of a purposive–representational pattern. Will it tend to count, then, as an agent? Not necessarily. For one further condition that we expect agents to satisfy is that they generate a purposive–representational pattern, not just over variations in surrounding context, but also over variations in how precisely they are compositionally organized and in how, therefore, the generative dispositions are realized within them.

The relevance of this issue is illustrated by the fact that a simple system for controlling the temperature in a room or building might otherwise count as an agent. This system generates a purposive–representational pattern of behavior, albeit one that is impoverished to the point where only one purpose is in play – to keep the temperature in a certain range – and only one sort of representation: that which registers the ambient temperature in the relevant space. The system might generate this behavior with only systematic perturbability and with a high degree of contextual resilience: no matter how the space is cooled or heated beyond the set range, for example, the system will restore the temperature to that range. But we would not for a moment think of the system as an agent. It would be quite extravagant to do so.

The reason why this is so, I suggest, is that while a heating-cooling system might be designed on this or that basis, any given system manifestly operates on the basis of a single, simple mechanism; this is going to be manifest even where it is unclear which particular mechanism is in operation. The agential pattern will be realized without exception or variation in that mechanism, then, and the causal relevance of any given agential configuration – say, the system's registering a drop in temperature – is going to be undermined by the causal relevance of the simple mechanism. There will be no information given about the genesis of the ensuing adjustment by tracing it to that configuration over and beyond the information given by tracing it to the state of the mechanism that uniquely realizes that configuration.

[5] The requirement of contextual resilience is close to John Searle's (1983) requirement that an agent satisfy "the background" condition of having sufficient skills to be able to adjust appropriately under situational variation.

Things would be very different if there were a variable pattern of realization for the agential configuration. In that case the appearance of that configuration would be causally relevant to the ensuing adjustment, since it would ensure the appearance of the adjustment, regardless of the state of the mechanism by which it is realized. If we think of the state of the mechanism as producing the adjustment, we might think in this case of the agential configuration as programming for that production: ensuring, regardless of its mode of realization, that the adjustment occurs (Jackson and Pettit 1988; Jackson and Pettit 1990b and Jackson and Pettit 1992).[6] In this case there will be extra information given about the genesis of the ensuing adjustment by tracing it to that configuration over and beyond the information given by tracing it to the state of the mechanism that uniquely realizes that configuration. The extra information will be that the adjustment is ensured, not just in the world where the actual realizing state is present, but also in those worlds where different mechanical states play the realizing role.

2 Candidate group agents

2.1 The transparent group

If agency requires just the display of purposive–representational pattern – specifically, in a way that satisfies systematic perturbability, contextual resilience and variable realization – then it is at least logically possible that a group of people might be an agent without the members of that group recognizing the fact; they might mediate the agency of the group in the unthinking, zombie-like manner in which, on a naturalistic picture, my neurons mediate my agency. Equally, it is at least logically possible that a group of people might be engineered into constituting an agent, while only one or two members recognize the fact; those in the know might recruit others to suitable roles, without revealing the agent-constituting point of the roles. And furthermore, it is certainly possible that a group of people might constitute an agent under a procedure that gives them differential roles and that makes the workings of the group relatively opaque to those in lesser offices, if not opaque in quite the same measure as in the other possibilities.

[6] The language of producing and programming should not suggest that there is a difference of kind between the way causality is exercised at the two levels. Considered in relation to the subatomic states that realize it in turn, the mechanical state can also be described as programming for the adjustment. For all that need be presupposed, there may be an infinite number of levels of this kind, and no bottom level at which causality is exercised in a different fashion.

In considering groups from the viewpoint of the question about real agency, I shall concentrate on more transparent possibilities of group formation. Specifically, I shall consider only groups in which there is full and equal awareness of the aspiration to agency among members, and full and equal participation in the attempt to realize that aspiration. This strategy makes sense. The existence of such transparent groups is not open to empirical doubt, so that it will be a significant result if we can establish that they are real agents. And if we can establish that such transparent groups are real agents, then there can be little hesitation about ascribing agency to variants in which the crucial factors remain fixed, but transparency is reduced.

Let us consider the case of groups, then, in which the members have a shared intention or commitment that they form a group agent; they jointly intend that together they operate in a way that parallels the manner in which an individual agent might behave (Searle 1995; Tuomela 1995; Bratman 1999; Gilbert 2001 and Miller 2001). There are many analyses of what shared intention requires but for our purposes here, we need not endorse any one of those analyses rather than others; all we have to assume is that there is good sense in the idea of shared intention.[7]

In order to emphasize the transparent character of the groups envisaged, let us assume in addition that the content of the shared intention is this:

- that the purposes and representations of the group be formed on the basis of member views – in effect, votes – as to which attitudes ought to be adopted;
- that the deputies who enact such purposes and representations on behalf of the group are selected on the basis of member views about selectional procedures;
- that in this formation and enactment of attitudes members are treated equally, having the same group roles, or the same chance of playing group roles, as others.

2.2 The transparent group

The sort of group to which these stipulations saliently direct us is the association or partnership or assembly in which members gather to discuss and vote on decisions about matters that engage them collectively and agree to be bound by those decisions, authorizing suitably chosen

[7] Pettit and Schweikard (2006) argue that an analysis that is broadly in the spirit of Bratman works well for a theory of group agency.

deputies to enact the decisions in their name. The standard image of such a body is classically associated with the image of the democratic assembly presented by Hobbes (1994, ch. 16) and Rousseau (1973, Bk 4, ch. 2) and, in less explicit mode, by Locke (1960, Bk 2, ch. 8.96). In this image, members unanimously agree to be bound by the majority vote of the assembly. They authorize the assembly, and those who are chosen to act for the assembly, as figures by whose agreed words and actions they are bound or committed.

Despite the distinguished heritage, however, this tradition is mistaken in suggesting that an assembly might operate as an agent on the basis of majority voting. Suppose that the assembly has to resolve logically connected issues, whether at the same time or over a period of time. No matter how deliberative and democratic the assembly is, and no matter how consistent the individual members are, majority voting may generate an inconsistent set of resolutions on such issues. And no assembly can be expected to function properly as an agent if the representations and purposes it endorses are inconsistent and incapable of being realized together. Assuming that the assembly will only vote on matters that are near the coal-face of action, and not for example, on abstruse issues of metaphysics or theology, any inconsistency in the representations or purposes is liable to affect its capacity to act; the attitudes will guide it at once in different directions. And in any case the endorsement of inconsistent representations or purposes will mean that other agents, including its own members, cannot think of it as a potential partner in reasoned exchange; no one can take seriously the commitments of an agent that does not care about the inconsistency of the positions it endorses.

The unreliability of majority voting is revealed by the discursive dilemma (Pettit 2001a, ch. 5), a problem that generalizes the doctrinal paradox in juridical theory (Kornhauser and Sager 1993). For an illustration of the dilemma consider the way a group of three members of a political party, assuming they endorse a balanced budget, might vote on whether to increase taxation, increase defense spending and increase other government spending. The members, A, B and C might each vote in a consistent pattern on these issues, yet the group view of A-B-C, as determined by majority voting, might involve an inconsistency. The possibility is registered in Table I.

There are many variations in which the discursive dilemma appears, all suggesting that no group can expect to function as a proper agent if it insists on forming its representations or purposes on the basis of majority voting (List 2006). But the problem is not restricted to majority voting. It turns out that making a group responsive to its individual members in the manner that is exemplified by majority voting, but not only by

Table 1

	Increase taxation	Increase defense spending	Increase other spending
A	Yes	Yes	Yes
B	No	Yes	No (reduce)
C	No	No (reduce)	Yes
A-B-C	No	Yes	Yes

majority voting, rules out an assurance that the group displays collective rationality. Specifically, it rules out an assurance that, if the group faces logically connected issues, then it can resolve them completely and consistently. This is a significant result. Every group will tend to confront connected issues, at least over time. And since these issues will typically be restricted to questions that the group needs to resolve in order to pursue its purposes, a failure to resolve them completely or consistently will undermine its agential capacity. A failure of consistency will leave the group unable to decide between rival courses of action; a failure of completeness will leave it without any purpose or representation to act on.

There are broadly three respects in which we might expect that a paradigmatically transparent group agent to be responsive to its membership. First, it should be robustly responsive to its members, not just contingently so; the group judgments should be determined by the judgments of members, independently of how the members judge. 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, and with its attitude to any proposed goal being determined by the attitudes of the members to that goal.

There are a number of possible voting procedures under which a group would be responsive to members in a robust, inclusive and issue-by-issue way, and majority voting is only one example. It can be shown that under a variety of interpretations, however, any suitably responsive group will tend to fail the requirement of collectively rationality. If it is faced with logically connected issues that it is required to resolve, then it is liable to endorse resolutions that are inconsistent with one another. Thus there will be a hard choice for the group, as in the discursive dilemma, between

endorsing individual responsiveness and aspiring to collective rationality. One example of a result that demonstrates this general problem is proved in List and Pettit 2002, and others have since followed.[8]

2.3 The transparent group

But if the majoritarian version of the assembly is not going to give us an example of a presumptive group agent, there are variants that certainly can do so. One obvious variant would be to have the assembly follow the sequential priority rule (List 2004). This would order issues so that whenever a group faces an issue on which its prior judgments dictate a resolution, voting is suspended or ignored and the judgment recorded on that issue is the one that its existing judgments dictate. The most salient ordering might be a temporal one.[9] Consider our A-B-C group in the earlier example and imagine that it took votes on the taxation issue first and then on the issue of defense spending. Under a sequential priority rule, the group would suspend or ignore the voting on the issue of other spending, for the first two votes would have mandated a decrease in such spending. This rule might be followed on the basis of reflection about what existing resolutions require. But equally, at least in formal domains, it might be followed on a mechanical basis, with a computing device registering the entailments from existing resolutions that dictate the response to new issues.

The reason why the sequential priority rule would enable the group to be consistent 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 a majority of members on

[8] See for example (Pauly and Van Hees (forthcoming) and Dietrich and List (forthcoming)). Notice that the three dimensions of responsiveness are not always reflected in a one-to-one fashion by three exactly corresponding conditions. The List-Pettit result demonstrates the problem under the following interpretation 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);
- Issue-by-issue responsiveness: the group judgment on each issue is fixed in the same way by member judgments on that very issue (systematicity).

[9] 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 and List 2004).

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 such a group could scarcely count as a rationally satisfactory agent. 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 will be unable to abide by them; it will not be reliably sensitive to the demands of evidence.

The evidential insensitivity of the sequential priority rule is apparent from the path-dependence it would induce.[10] 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.

2.4 The straw-vote assembly

Happily, however, there is a variation on the sequential priority rule that would enable a group to escape the conflict between individual responsiveness and collective rationality, without forcing it to be evidentially inflexible. Under this variation, the assembly would consider different issues in turn, depending on the order in which they arise, and with every issue that arises the assembly would determine whether existing resolutions dictate a resolution of that issue. But at that point it would take a different approach from the sequential priority rule.

Rather than automatically endorsing the resolution dictated in such a case – rather than suspending or ignoring a vote on the most recent issue – the assembly would adopt the following sequence of steps: take

[10] For ways of mitigating the effects of path-dependence see ist (2004).

a straw vote on the new issue; if the vote gives rise to an inconsistency, identify the other resolutions that combine with that straw resolution to generate the problem; and then vote on which of the problematic resolutions to reject in order to restore consistency. Under this straw-vote procedure the upshot might be the same as under the sequential priority rule: a rejection of the straw vote. But equally it might be the revision of some of the group's existing commitments. Both possibilities are open and so the group that follows this line can avoid the path-dependency, and the associated inflexibility, of the sequential priority rule.

Consider, then, how the A-B-C group might operate under the straw-vote constitution. They will register that their existing commitments on taxation and defense spending require them not to increase – in fact to reduce – other spending. And then, if their vote goes in favor of increasing other spending, they will register the inconsistency and reflect on which of their existing and proposed resolutions to drop. They may decide to reduce other spending, as the sequential priority rule would require them to do. But equally they may decide to revise their commitment not to increase taxation or to increase defense spending. All three options are open.

The assembly that follows the straw-vote procedure is as standard a candidate for group agency as the assembly that cleaves to the strategy of determining every issue by majority vote or that follows the sequential priority rule. And it has the virtue, in similar measure, of being wholly transparent. If we can show that this sort of group should count as a real agent, then we will have shown that a very plausible sort of group can display real agency. Moreover, we will have shown that real agency is a prospective feature for other sorts of groups that depart from it in ways that are not crucial to the argument provided in support of real agency.

3 The real agency of the straw-vote assembly

Does the straw-vote assembly introduced in the previous section count as a real agent, by the criteria of agency emerging from our discussion in the first section? The general question is whether it can display a broadly agential pattern of behavior, however subject to perturbations. The more specific questions are whether the perturbability of the pattern is systematic rather than unsystematic; whether the pattern is relatively resilient across different contexts; and – most important, as we shall see – whether it is variably realized across the contributions of individual members.

3.1 The general question

There can be little doubt about the capacity of a straw-vote assembly in general to display a broadly agential pattern of behavior. The members of such an assembly will collectively endorse certain purposes, perhaps revising them from time to time, and they will ensure that the path taken to the realization of any endorsed purpose is determined by the representations that they also collectively endorse, and no doubt revise as occasion demands. The guiding representations will bear on a variety of matters such as the opportunities available for satisfying their purposes, the relative importance or urgency of those purposes, and the best means at their disposal for realizing one or another purpose.

Why does it seem so natural to ascribe a purposive–representational pattern of behavior to a straw-vote assembly? The reason may be that the members are required to reason as a group and to develop a critical perspective on the demands associated with that pattern. For it is hard to think that a group which reasons about what is demanded under a certain pattern of purposes and representations, and which regulates itself for fidelity to those demands, might not actually display such a pattern, at least in broad outline. This feature of the straw-vote assembly marks it off dramatically from any simpler arrangement, such as the assembly that operates by majority vote (Pettit 2007a).

In the majoritarian assembly the members share an intention that they together perform as an agent but under the majority rule they need never reflect on what is demanded of the group agent in view of its existing commitments. All they each have to do, at least in the formation of attitude, is to play their local part, voting as required on the different issues posed and trusting in the majoritarian constitution to assemble their individual contributions into a collective, sensible whole. Under such a mode of organization the members as a group would never conduct anything akin to reasoning in sustaining the performance of the group. They would each follow a personal rule of voting: say, that of voting according to their individual judgment on any issue of collective purpose or representation. And blind adherence to that rule is all that the group would require of them; in sustaining the group they might each be as unreflective as the ants that sustain a colony or indeed the neurons that sustain an individual agent.

But the members of an assembly cannot rely blindly on a majoritarian constitution to ensure that their individual contributions are assembled into a sensible group profile, whether in the space of purposes or representations; as we saw, majority voting might lead the group to endorse an inconsistent set of attitudes. Nor can the assembly members rely

blindly on a sequential priority rule – say, one applied by a computer – that would automatically discount any vote that generates inconsistency with prior resolutions; such a rule might lead them, path-dependently, to endorse an evidentially unsupported set of attitudes. That is why we resorted to the straw-vote assembly in order to identify a plausible candidate for group agency.

In the straw-vote assembly, however, members are expected to reason. They do not let the group attitudes be generated blindly, as under the majoritarian constitution. Nor do they conform blindly to *modus ponens*, as the sequential vote procedure would have them do. They have to consider those propositions that have already been endorsed as purposes and representations; they have to determine what those propositions imply, if anything, for the answer to any issue that is currently up for voting; and in the event of the vote going contrary to such implications, they have to decide on which of the conflicting resolutions to drop. This means adopting precisely the sort of critical stance on the demands of a purposive–representational pattern that individual human beings embrace when they question their spontaneous processes of attitude formation.

The fact that the members of the straw-vote assembly can recognize and respond to the demands associated with purposive–representational behavior should give us confidence that they will generally display that sort of behavioral pattern. It means, after all, that while their behavior may sometimes drift away from that pattern, there are correctives available that should serve to guard against this. But even when those correctives fail to work in a given case, the critical character of the group may give us grounds for continuing to ascribe agency; it may mean that the actual display of suitable behavior is less important than it would have been with a non-critical agent. For if a straw-vote assembly is truly sensitive to purposive–representational demands, then in a case where its behavior drifted away from the required pattern it can presumably be made to recognize the fact and to acknowledge it as a failure. To the extent that it can do this, and can use the recognition of the failure to guard in some measure against repeat failures, we can be much more confident that this is a system that should count as an agent. Whatever past lapses from suitable behavior, it apparently has the capacity to guard against similar lapses in the future.

But now I should turn to the more specific questions that may be raised about the claim of a straw-vote assembly to count as a real agent. Those questions bear on the systematic perturbabilty, the contextual resilience and the variable realization of the agential behavior that an assembly is likely to display.

3.2 The specific questions

Systematic perturbability. Every natural agent, human or animal, individual or collective, is bound to fall away from the demands of the purposive–representational pattern, whether on the attitude-to-evidence, attitude-to-action, or attitude-to-attitude front. There is a great difference, however, between two sorts of perturbability. A system may be subject to random perturbations whose origin remains opaque, or it may be subject to perturbations that derive from identifiable factors. It may be subject to unsystematic or systematic perturbability.

In the unsystematic case, we may certainly say that the system approximates the profile of an agent. While many of the responses it makes do not have any agential sense, they are few enough in number to count as noise in a system that is otherwise constructed to display agential pattern. But approximating the profile of an agent is not necessarily being an agent; it may just mean simulating agency in a more or less imperfect manner. In the systematic case, things are different. Being able to identify the sources of perturbation, we may be able to see them as constraints on the operation of the system that its history of selection or design had to take as given.

We have no hesitation in identifying systematic sources of perturbation in our own performance as individual agents, given that we reason with one another about the demands imposed by purposive–representational pattern. We assume that we can target the same purposes, the same representations, and the same requirements of rationality. When we have access to the same information, therefore, but diverge from one another in relevant judgments – say, judgments on what the evidence supports, on whether certain purposes or representations are inconsistent, on what means are required for a given purpose – we assume that at least one of us has been subject to perturbation. And so we have a heuristic available for identifying perturbers: we look for factors such that their presence tends to generate judgments – and, presumably, corresponding behaviors – that are discrepant from those of others. This heuristic has produced a wealth of folk knowledge on the perturbing effects of perceptual obstacles and interpersonal pressures, of bias and passion and inattention, and of paranoia and compulsion and other pathologies. And this body of knowledge has been greatly expanded with psychological studies of cool and hot irrationality.

What is true of individual human beings, performing as individuals, will presumably carry over to human beings in assemblies that reason in the manner of the straw-vote group. While every such assembly is going to fail as an agent in various respects, there is surely ground for expecting

that the failures will be traceable to systematic sources of perturbation of the kind with which we are familiar with individual human beings. There may also be sources of perturbation that operate on assemblies and other groups but do not affect human beings in isolation. But again these will tend to be more or less familiar or identifiable, such as the contagion effect whereby panic or pack-behavior can be generated among people in crowds, or the hierarchy effect whereby some group members may mindlessly defer to others.[11]

3.3 Contextual resilience

The first lesson of our earlier discussion was that that the perturbability of a would-be agent should be systematic rather than unsystematic and we have found that it is borne out with the straw-vote assembly. The second lesson was that equally, the purposive–representational pattern displayed by the system should be contextually resilient. It should not be tied to such a specific context of performance that it is misleading to think of the system as aiming at a more abstract goal; the lesson was illustrated by Dennett's Sphex wasp.

As the critical, reasoning character of the straw-vote assembly ensures that its perturbability is as systematic as that of ordinary human subjects, so that character makes it natural to ascribe a high degree of contextual resilience to the pattern of behavior that it displays. Let the group behave in a given context after a certain purposive–representational pattern. Should we expect it to be able to adjust so as to continue to realize the same purposes, as the context changes? Or should we expect it, like the Sphex wasp, to be capable of displaying the pattern only in a stereotypical, context-bound way?

If the straw-vote assembly is capable of reasoning then it has to be capable of recognizing the abstract goals it targets, and the different means that may be appropriate for realizing them in different situations. And if it is capable of reasoning then it has to be capable, equally, of responding to that recognition and choosing the appropriate means. But the presence of those capacities means, then, that the purposive–representational pattern it displays is more or less bound to enjoy a high degree of contextual resilience. The group will be robustly disposed to

[11] Our earlier discussion shows why we should be loathe to ascribe agency when the perturbability is unsystematic. The majoritarian assembly might pass as an agent that is unsystematically perturbable, since it acts in a purposive–representational way, subject to the random, ramifying disturbance introduced by discursive dilemmas and the like. But it would be wrong to think of such an assembly as an agent, since it lacks a crucial, agential resource: the capacity to register and remedy attitudinal inconsistency.

pursue relevant purposes according to appropriate representations, not disposed to do so only under fixed situational parameters.

3.4 Variable realization

But now we come to a third, more problematic issue. When a straw-vote assembly operates as an agent does the realization of that pattern by individual members vary in such a way that we have to see the agential configuration at any point – say, the group's endorsing such and such a means of achieving such and such a goal – as causally relevant to what ensues? Or is the variation so limited that that configuration has no causal relevance over and beyond the relevance of the realizing set of attitudes that is present in individual members? Does the group configuration program for the ensuing behavior over a range of possible ways in which it might be realized at the individual level? Or is there no significant variation at the individual level and no ground for assigning a programming role to the group attitudes?

This issue may seem more problematic than the other two, precisely because the straw-vote assembly is a critical, reasoning body in which individuals do not play their parts blindly, like ants in a colony or neurons in the brain. Individual members monitor where the group is going, and rely on issue-by-issue voting only when there is no issue of consistency between the different purposes or representations they endorse. And so it may well seem that while that reasoning character made it easy to give appropriate answers to the questions of systematic perturbability and contextual resilience, it makes it difficult to defend an appropriate answer to the question of variable realization. If individuals play the critical part required under the straw-vote model, why not think of the configuration that programs at any point for action as the configuration of individual attitudes? Why ignore individual attitudes and invoke the group-level configuration as the causally relevant antecedent?

When an assembly operates with the straw-vote procedure, it is certainly true that what the group does it does with the full endorsement of members; nothing happens behind their backs. But that does not undermine the possibility that group attitudes are variably realizable and that they program for group responses independently of how they may be realized in the dispositions of members. On the contrary, the reasoning character of the straw-vote assembly actually increases the variability with which the group attitude on any issue is likely to be realized at the individual level. And so here, as with the other two issues, it argues for an answer that supports taking that sort of group as a real agent.

Consider the majoritarian group that constructs its purposes and representations on the basis of majority vote. In this case there will be one obvious source of variation in the way the constructed attitude is realized at the individual level. The majority may be any of those sets of members that have one more member than the set of other members; any of those sets of members that have two more members than that other set; and so on. And so to know that the group is committed to a given purpose or representation is not to know much about how in particular the individual members are disposed. It is only to have knowledge of the statistical, aggregative fact that a majority supports that purpose or representation.

This source of variation as between individual and group levels is expanded in the case of the sequential priority group by a further factor. If such a group holds by a certain purpose or representation, that may be because of a majority of members support it, as in the majoritarian case. But it may also be because, while a majority reject that purpose or representation, the group is required in consistency to endorse it, given prior majority support for certain other purposes or representations. Thus the ways in which individuals may be disposed, consistently with the group endorsing that purpose or representation, are greater in number than the ways in which members may be disposed in the counterpart case with the majoritarian assembly.

The two sources of variation that are relevant with the sequential-priority group remain in place with the straw-vote assembly. But at this stage, a third source of variation enters as well, for there is a further way in which individuals may adjust so as to ensure that the assembly endorses a certain purpose or representation. This is the sort of adjustment that members make when they revise an earlier commitment and endorse a certain group attitude, because that adjustment is taken by them to be the best way of responding overall to the demands of evidence.

When we take all of these sources of variation into account, it becomes hard to see any reason why we should balk at treating a straw-vote assembly as a real agent. Not only can such a body display a purposive–representational pattern of behavior under a systematic mode of perturbability and with a high degree of contextual resilience. It is more or less bound to display this pattern in a way that is radically discontinuous with the attitudes of the individual members who constitute it. We should have no hesitation in looking to the group attitudes at any point as causally relevant factors that program for what the group goes on to do. For that complex of attitudes will program for the group response over an indefinite range of variations in how it is realized in the dispositions of individual members. The contrast with the simple heating-cooling system could not assume a starker profile.

4 Conclusion

The criteria I proposed as tests of agency are hard to question and the straw-vote assembly that I identified as a candidate for group agency is hard to dismiss as an institutional possibility. Yet by those criteria it is demonstrable that that candidate does indeed count as an agent. Thus there is every reason to conclude that groups can be real agents. Nothing but prejudice can stand in the way.

But prejudice on this matter is in no short supply. It comes in two major forms, one associated with an epistemological presupposition, the other derived from a complex of metaphysical and normative fears.

The epistemological presupposition that may block the admission of group agency is the assumption that if group agents are real, then they must have a mental life of their own, in particular a mental life that is not accessible to other agents. Thus they may have to draw on the conceptual and ratiocinative abilities of their members in order to operate properly; but they must have a consciousness that individuals as such do not access. They must operate, in some sense, behind the backs of their members.

Group agents in the straw-vote mould certainly do not operate behind the backs of their members; they exist by virtue of the monitoring and management that those individuals exercise. But agents are distinguished by the sets of attitudes that they embody, and by the principles of development to which those attitudes are subject, not by the extent to which their attitudes are inaccessible to others (Rovane 1997). And even though the attitudes of a straw-vote assembly are fully accessible to its members, being intentionally formed and enacted by the membership, they constitute a developing set that bears no systematic relationship to the attitudes by which the individual members are each characterized. Group attitudes have to satisfy criteria of rationality in order to support a unified pattern of agency and their being robustly rational means, as we saw, that they cannot be systematically responsive to the attitudes held by individual members.

There is also a complex of metaphysical and normative fears that may stand in the way of admitting the reality of group agency. Thus, people will balk at the admission on the grounds that it makes groups mysteriously emergent, that it would raise again the totalitarian specter that was banished in the last century by Popper's attack on holism, or that it is liable to introduce group rights as restrictions on the rights of individuals.

These sorts of fears, however, are baseless. The straw-vote assembly has an agential profile that may go with any of a variety of individual profiles but it is nothing and it does nothing except on the basis of the

contributions of its members; it is superveniently dependent on how those members are disposed to behave (List and Pettit 2006). The state may well count as a real agent, by extension from the case of the straw-vote assembly, but that does not mean that the state has to be given the totalitarian role that it assumed amongst fascists and communists (Pettit 2003). And while group agents of different kinds, associational, commercial and political, may be given certain rights under the law, the rights they are given should surely be constrained by how well and how far they serve individuals; they are the tail, we are the dog.

This is not to say that ascribing reality to group agents has no important normative and explanatory implications. On the normative side it means, as I have argued elsewhere, that group agents should be held responsible for programming for certain actions, even though it may also be appropriate to hold members responsible for enacting their programmed roles (Pettit 2007b). And on the explanatory side it means that there is good reason to seek explanations at a level where group agents are treated as agents in their own right without always exploring the nuts and bolts of individual contribution; the refusal to go to the fine grain of causal mechanism may be crucial for the pursuit of certain explanatory purposes (Pettit 1993, ch. 5). More generally, recognizing the reality of group agents opens up an enormous range of questions as to how such entities can and should be designed, both in general and in certain political or other contexts. It places an important research program on the agenda of explanatory and normative social theory.[12]

REFERENCES

Block, N. 1981. "Psychologism and behaviorism", *Philosophical Review* **90**: 5–43.

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

Davidson, D. 1980. *Essays on Actions and Events*. Oxford: Oxford University Press.

Dennett, D. 1979. *Brainstorms*. Brighton: Harvester Press.

1991. "Real patterns", *Journal of Philosophy* **88**: 27–51.

Dietrich, F. and C. List (forthcoming). "Arrow's theorem in judgment aggregation", *Social Choice and Welfare*.

[12] This chapter owes an enormous debt to my collaboration with Christian List on a book about group agents; in that book we detail and illustrate the sort of research program at which I only gesture here. I benefited greatly from discussion at two presentations of the material: one, in June 2007, at Witten/Herdecke University and the other ("Collective Intentionality VI" in July 2008), at the University of California, Berkeley).

Fodor, J. 1975. *TheLanguage of Thought*. Cambridge: Cambridge University Press.

Gilbert, M. 2001. "Collective preferences, obligations, and rational choice", *Economics and Philosophy* **17**: 109–120.

Hobbes, T. 1994. *Leviathan*. E. Curley (ed.) IN: Hackett.

Jackson, F. and P. Pettit 1988. "Functionalism and broad content", *Mind* **97**: 381–400; reprinted in F. Jackson, P. Pettit and M. Smith, 2004. *Mind, Morality and Explanation*, Oxford: Oxford University Press.

1990a. "In defence of folk psychology", *Philosophical Studies* **57**: 7–30; reprinted in F. Jackson, P. Pettit and M. Smith, 2004. *Mind, Morality and Explanation.* Oxford: Oxford University Press.

1990b. "Program explanation: A general perspective", *Analysis* **50**: 107–17; reprinted in F. Jackson, P. Pettit and M. Smith, 2004. *Mind, Morality and Explanation.* Oxford: Oxford University Press.

1992. "In defence of explanatory ecumenism", *Economics and Philosophy* **8**; reprinted in F. Jackson, P. Pettit and M. Smith, 2004. *Mind, Morality and Explanation*. Oxford: Oxford University Press.

Kornhauser, L. A. and L. G. Sager 1993. "The one and the many: Adjudication in collegial courts", *California Law Review* **81**: 1–59.

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

2006. "The discursive dilemma and public reason", *Ethics* **116**: 362–402.

List, C. and P. Pettit 2002. "Aggregating sets of judgments: An impossibility result", *Economics and Philosophy* **18**: 89–110.

2006. "Group agency and supervenience", *Southern Journal of Philosophy* **4**: 85–105.

Locke, J. 1960. *Two Treatises of Government*. Cambridge: Cambridge University Press.

Miller, S. 2001. *Social Action: A Teleological Account*. Cambridge: Cambridge University Press.

Millikan, R. 1984. *Language, Thought and Other Biological Categories*. Cambridge, MA, MIT Press.

Pauly, M. and M. Van Hees (forthcoming). "Logical constraints on judgment aggregation", *Journal of Philosophical Logic*.

Pettit, P. 1993. *The Common Mind: An Essay on Psychology, Society and Politics*. Paperback edition 1996. New York, NY: Oxford University Press.

2001a. *A Theory of Freedom: From the Psychology to the Politics of Agency*. Cambridge, New York, NY: Polity and Oxford University Press.

2001b. "Deliberative Democracy and the Discursive Dilemma", *Philosophical Issues supp to Nous* **11**: 268–299.

2003. "Deliberative Democracy, the Discursive Dilemma, and Republican Theory." In *Philosophy, Politics and Society Vol 7: Debating Deliberative Democracy*. J. Fishkin and P. Laslett (eds.). Cambridge: Cambridge University Press, pp. 138–162.

2007a. "Rationality, reasoning and group agency", *Dialectica* **61**: 495–519.

2007b. "Responsibility incorporated", *Ethics* **117, pp. 171–201**.

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

Rousseau, J.-J. 1973. *The Social Contract and Discourses*. London: J.M. Dent & Sons Ltd.
Rovane, C. 1997. *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton, NJ: Princeton University Press.
Searle, J. 1995. *The Construction of Social Reality*. New York, NY: Free Press.
Searle, J. R. 1983. *Intentionality*. Cambridge: Cambridge University Press.
Stalnaker, R. C. 1984. *Inquiry*. Cambridge, MA, MIT Press.
Tollefsen, D. 2002. "Organizations as true believers", *Journal of Social Philosophy* **33**: 395–410.
Tuomela, R. 1995. *The Importance of Us*. Stanford, CA: Stanford University Press.

3 – Comment
A Note on Group Agents

Diego Rios

The primary aim of Pettit's chapter is to provide a general framework detailing the conditions for ascribing agential status to groups (such as political parties, assemblies, churches, states, etc.) in a way that parallels the attribution of agency to individuals. We normally use an intentional vocabulary to refer to the behavior of complex collective organizations, and we implicitly assume that these organizations behave as true agents. We say, for instance, that the aim of Parliament at the moment of a vote on new legislation is to reduce poverty by increasing welfare allocations; or, we say that the objective of the government with this or that measure is to reduce unemployment. In both cases, Parliaments and governments are conceived as agents having specific goals and objectives that they attempt to promote. These ascriptions of intentional and purposive behavior have both normative and explanatory consequences. From the normative point of view, they are used to create obligations and other commitments: we say, for instance, that such and such an assembly has promised to do this or that, and we can criticize it for failing to honor its self-imposed obligations. From the explanatory point of view, ascribing purposes and goals to organizations and groups is a way to account for their behavior: we explain, for instance, the behavior of governments and states by ascribing goals to them and assuming that they attempt to satisfy – with different degrees of success – such goals. I take as uncontroversial the *existence* of this kind of agential talk about groups and organizations. The problem is how literally we are ready to interpret this agential vocabulary.

Group agency has sometimes been looked at with some scepticism within the social sciences. There are two major sources for this scepticism. One source is associated with social choice theory. Social choice literature has isolated systematic failures at the moment of aggregating individual preferences, generating an array of well-known social paradoxes (Riker 1982). A paradigmatic example is the case of majority voting, that could generate – at the group level – an incompatible set of social preferences. Even when voters have consistent sets of preferences,

majority rules may not be able to deliver a properly consistent set of social preferences. The prospects for granting agential status to collective bodies governed by this inconsistent set of preferences looks rather dubious. The group or assembly in question could, after all, endorse incompatible courses of action and push through resolutions that violate elementary rationality constrains. This analysis has been enlarged to cover a wide range of paradoxes (List 2004 and 2006; List and Pettit, 2002 and 2004; Kornhauser and Sager 1993; Pettit, 2001 and 2001b).

The other source of scepticism about group agency has its roots in the old philosophical tradition associated with individualism. According to Hayek and Popper, groups are not true agents. They would probably concede that we sometimes *speak* as if they were intentional agents, ascribing them intentional features; nevertheless, this is just a *facon de parler*. There might be different ways to state what these authors understood by individualism, but they seem to have amalgamated individualism with singularism (Gilbert 1989): the idea that only individuals are agents. Early individualist literature seems to have rejected both the emergent status of groups and their agential standing, maybe assuming – without much discussion – that granting agential status to groups inevitably leads to conceiving them as emergent. Independently of how appropriate this amalgamation is, it seems to be true that Popper and Hayek were emphatically committed to the idea that only individuals are agents, and that talk about group agency is purely metaphorical.

These are then two possible sources of scepticism for the project of granting agency status to groups. Pettit's objective is to provide a general framework to dispel some of the assumptions underlying this scepticism. His strategy is developed in three steps. First, he sets the conditions that should obtain in order for it to be possible to ascribe agential status to a given system. Pettit' s reasoning on this issue has a strong Dennettian flavor: a system will count as properly agential when it exhibits the disposition to display, in a wide range of contexts – actual and counterfactual – a purposive–representational pattern. The second step consists in showing how this framework could be applied to fully transparent groups – groups where all its members have full and equal awareness of the collective goals. The third – and last – step consists in generalizing what has been said about fully transparent groups to less than fully transparent ones: the main line of the argument is that once the agential status of fully transparent groups has been granted, it is natural to extend the claim to cases where transparency is reduced.

The first part of the chapter discusses the conditions that must be met for a system to count as an agent. One common objection to the instrumental theory of agency is that it is too generous at the moment

of granting agential status: too many systems count as agents. Some of these difficulties are avoided by introducing further conditions – systematic perturbability, contextual resilience, variable realization – restricting the set of systems that could eventually be granted agential status. Nevertheless, I am not sure that all our intuitive judgments are captured by this apparatus. Compare Dennett's Sphex wasp and a Coke machine: they will both exhibit similar scores when submitted to the perturbability, resilience and variable realization tests. Intuitively I would say however that Dennett's Sphex wasp might count as an agent, while the Coke machine cannot be one. I am not sure that we can capture this difference. It could be argued that it need not be so: it is perhaps enough if it is able to make sense of most intuitions, without taking into account some dubious or borderline cases where our own intuitions may not be precise enough.

In the second part of the chapter, Pettit dispels some of the reasons for being sceptical about group agency. Although majoritarian rule could give rise to inconsistent sets of preferences, more demanding voting procedures – like the straw-vote procedure – might help the members of the group to rule out manifest inconsistencies in group resolutions, dispelling at least one important source of trouble. Note that Pettit's rehabilitation of group agency does not imply that groups are emergent entities. The behavior of the group *supervenes* on the behavior of its members: the group, even if conceptualized as a true agent, is still dependent on individual behavior. Every variation at the group level will be accompanied by at least one change or variation at the individual level. Individuals are always the underlying causally efficacious elements responsible for group behavior. This line of thought makes good sense when looked at with other of the author's major contributions to the field – the idea of program explanations (Jackson and Pettit 1990, 1992). Group agents could be conceived as *programming* individual behavior: in a way they contribute to *canalizing* individual behaviors along specific lines. Although not causally efficacious, group agents are nevertheless causally *relevant* in the production of social outcomes: they raise the frequency of certain types of outcomes and contribute to directing the behavior of individuals (Pettit 1993: 258). They are then an important part of the causal history of a given social event.

The primary aim of Pettit's paper is to reconsider fully transparent and quasi-transparent groups as potential agents. I find his analysis on this topic very convincing. Pettit is not concerned in this chapter with fully opaque groups; nevertheless he leaves open the possibility of conceiving them as agents. It might be interesting to know whether this analysis could be extended to cover the case of fully opaque groups. In

stark contrast with fully transparent groups, the fully opaque ones are characterized by the fact that *none* of the individual members is aware of the purposive global outcomes of the group. Pettit briefly mentions this possibility:

> If agency requires just the display of purposive–representational pattern – specifically in a way that satisfies systematic perturbability, contextual resilience and variable realization – then it is at least logically possible that a group of people might be an agent without the members of that group recognizing the fact; they might mediate the agency of the group in the unthinking, zombie-like manner in which, on a naturalistic picture, my neurons mediate my agency.

This paragraph can be interpreted in different ways. The parallel to zombie-like neurons suggests that there is room at the group level for true agency, even when *all* the individual realizers are unaware of the high-level agential goals and purposes of the group. In the case of fully opaque agents, the parallelism between the individual and the collective agency would be strong: the constituent elements – neurons, in the case of an individual agency; individuals, in the case of group agents – produce high-level purposive outcomes via purely blind, mechanical interactions. The idea is that none of the individuals – as in the case of the neurons – is aware of the global outcomes that its own behavior is helping to promote.

How plausible is this option? Some could argue that fully opaque groups are exactly the examples that critics like Popper and Hayek had in mind when they criticized the reification of groups as agents. Note that transparent or semi-transparent groups are easier to tackle: it is not impossible to imagine a group designed by some of its members in such a way as to behave as a group agent; the designers need just to *canalize* the behavior of the other individuals. Although unaware of the global consequences of their actions, these individuals nevertheless contribute by their own behavior to produce the collective purposive outcome. None of the members of the group – except the designers – need know about the general purpose of the group. In the case of partially transparent groups, the explanation of the purposive outcome will be given in terms of those constituents of the groups that are transparent – the designers. This move however cannot be made when the group is fully opaque, because, by definition, fully opaque groups lack individual designers: *all* the constituents of such a group are blind about the ultimate goals of the group.

The most serious problem connected to granting agential status to fully opaque groups is that the ultimate global outcomes of the system are left totally *unexplained*. Many times in the social sciences, fully opaque

groups have been described as agents, without any clue having been provided about the mechanism generating the purposive outcome. I am sceptical about this strategy. Maybe selectional considerations could be introduced to explain the emergence of group purposive outcomes at a group level, even when all the members of the group are unaware of the collective goals of the organization to which they belong. This selectional move might contribute to explaining the global purposive outcome. This is an intriguing option that opens complex philosophical issues about the units of selection; unfortunately, I am not competent to assess this suggestion.

Two further notes. The first concerns the scope of the theory of agency when applied to fully opaque groups. Pettit argued in favor of a functionalist interpretation of agency that could – eventually – be enriched with thicker conditions requiring the existence of a ratiocinative capacity. The straw-vote model left room for this ratiocinative process to take place: the members of the group were "forced" to reason as a group and to take a critical attitude to the potential intentional profile of the group. Obviously this enriched conception of agency as involving sensitivity to rational requirements can only be applied to fully transparent or semi-transparent group agents. Fully opaque group agents cannot be critical agents in the same way: they lack the internal resources to be so. In order to be to able to take an evaluative stance toward the intentional profile of the group, the member must be aware of the purposes of the group as such – this condition is absent in fully opaque group agents. In a way, fully opaque group agents can only be part of the thin functional theory of agency, not the thick ratiocinative one. This makes fully opaque group agents much less interesting, especially when considered from the perspective of institutional design.

The second point generalizes what has been said before concerning the relationship between singularism and individualism. One of the important points of the chapter is that granting agential status to groups need not violate individualist constraints. A critic might think that this would not be an option when dealing with fully opaque groups. Nevertheless this objection is not correct. All social outcomes will be traceable to individual behavior: social outcomes supervene on individual behaviors. This is so independently of the degree of opacity of group agents. The two issues are conceptually different. Granting fully opaque groups agential status need not imply a commitment to a reified conception of groups. Even fully opaque group agents would be the result of the interaction of individual constituents: granting agency to them does not necessarily entail conceiving them as mysteriously emergent players in the social world.

To sum up: Pettit's chapter is a powerful contribution to one of the fundamental problems in the philosophy of the social sciences – to wit, the possibility of enlarging the set of social actors to include not only singular individuals but also groups. This opens for the near future a rich philosophical agenda, raising new questions and challenges for the entire domain. Apart from quite obvious implications for social theory in general, the issues raised in this paper have the potential to generate a fresh look at old practical problems, concerning how to design institutions in order to help the members of a group to police and filter the consistency of the organization's resolutions and attitudes. There might also be important normative consequences for the way we assess the responsibility of individuals for the actions of the group they belong to. The claim that some types of groups, but not others, contribute to making individuals *think* about their goals as a group – a process that amounts to properly collectivizing reason – promises to be of paramount importance in the near future.

REFERENCES

Gilbert, M. 1989. *On Social Facts*. Princeton, NJ: Princeton University Press.

Jackson, F. and P. Pettit 1990. "Program explanation: A general perspective", *Analysis* **50**: 107–117.

1992. "In defense of explanatory ecumenism", *Economics and Philosophy* **8:** 1–21.

Kornhauser, L. and L. Sager 1993. "The one and the many: Adjudication in collegial courts", *California Law Review* **81**: 1–59.

List, C. 2004. "A model of path-dependence in decisions over multiple propositions", *American Political Science Review* **98**(3): 495 – 513.

2006. "The discursive dilemma and public reason". *Ethics* **116**(2): 362–402.

List, C. and P. Pettit 2002. "Aggregating sets of judgments: An impossibility result", *Economics and Philosophy* **18**: 89–110.

2004. "Aggregating sets of judgements: Two impossibility results compared", *Synthese* **140**(1): 2007–235.

Pettit, P. 1993. *The Common Mind: An Essay on Psychology, Society and Politics*. Oxford: Oxford University Press.

Pettit, P. 2001a. *A Theory of Freedom* Polity, 2001.

2001b: "Deliberative democracy and the discursive dilemma", *Philosophical Issues*, **11**: 268–299.

Riker, W. 1982. "*Liberalism against populism: A confrontation between the theory of democracy and the theory of social choice*", San Francisco, CA: Freeman & Co. Ltd.