4

How to Tell if a Group is an Agent

Philip Pettit

Introduction

I take a human group to be a collection of individual human beings whose identity as a group over time, or over counterfactual possibilities, need not require sameness of membership. The typical group can remain the same group even as its membership changes, with some members leaving or dying, others joining or being born into the group. As we envisage the possibility of changes in the membership of such a group, even ones that are never going to materialize, we think of them as changes in one and the same, continuing entity.[1]

This conception of a group distinguishes it from a set or collection, where a change of members necessarily entails a change of set. But it still encompasses a generous range of social bodies, since it says nothing about the basis on which we individuate a group over time or possibility. It allows us to take almost any property, whether of origin or ethnicity, belief or commitment, career or hobby, even height or weight, to fix the identity of a group. The Irish, the Catholics, the lawyers, the stamp collectors and the obese can constitute groups. Thus while groups may vary in how important their individuating property is, and in how far it is socially significant for members or non-members, the information that a collection constitutes a group is no big news.

Among groups in this common, downbeat sense, however, some stand out from the crowd. These are groups that perform as agents, incorporating in a way that enables them to mimic the performance of individual human beings. They make judgments, form commitments, plan initiatives and, relying on members who act in their name, undertake actions in any of a range of domains. Examples are the partnerships and companies that operate in commercial space, the associations and movements that characterize civil society, and the churches and states that shape the lives of people throughout the world. Such entities certainly involve collections of individuals in coordinated relationships, and they certainly count as groups since they are individuated by their common acquiescence in what is done in their name. But their capacity to act, and more generally to perform as agents, marks them out. They are a class apart.

This claim is not uncontentious, however, since there are many instances where we ascribe agent-like features to groups without any real suggestion that they count as agents proper. Thus we say that the bond market is unsettled by the indecisiveness of the Eurozone leadership, or that the X-generation has lost its affection for video games, or that the sun-worshippers on a beach acted courageously in helping to save a swimmer in difficulty. Yet most of us will agree that the bond market is just a network of bond traders, each with his or her own goals; that the X-generation is just the collection of people born between about 1965 and 1980, allegedly characterized by certain shared traits; and that if those on the beach acted courageously, that just means that they each played a part in a

coordinating plan, not that they formed a distinct agent. In none of these cases is there a serious candidate for the role of a group agent.

If I am to support my belief in group agents, then, I have to be able to give an account of how we can tell bona fide group agents apart from mere pretenders like these. That is what I try to do in this paper, building on work done jointly with Christian List (2011). I begin with a discussion of agency in general, distinguishing between non-personal and personal agency, and argue that we have a special way of detecting personal agency: by the direct experience or indirect evidence of interpersonal engagement. And then I try to show that under plausible epistemic scenarios such experience or evidence is necessary for the ascription of agency to a group.

The paper is in four main sections. In the first I provide a general account of agency, concentrating on simple cases. In the second I argue for the distinctions advertised between different modes of agency and of agency-detection. And in the third I use this material to argue that in our ordinary practice only direct or indirect evidence of interpersonal engagement provides a warrant for ascribing agency to groups. This means that the only group agents that we generally recognize are agents of a personal kind and in the final section, connecting with the work done with Christian List (2011), I argue that such group agents count as real, non-fictional agents.

4.1. The conception of agency

A system is an agent insofar as it is organized to instantiate a set of goals and a set of representations and to pursue those goals in accordance with those

representations. This notion of an agent is best introduced with a simple example. Imagine that you return home one evening to find that your whiz-kid sister has set up a knee-high robot in the kitchen, which she invites you to observe. You see that it has bug-like eyes that appear to scan the room, wheels on which it can move about, and arms suited to lifting and adjusting objects up to its own size. Your sister drops a can on the floor and, to your surprise, the robot moves towards it, lifts it in an awkward embrace, then takes it to a trash bin in the corner and deposits it there. Amazed, you check for reliability by dropping an orange on the ground and, once again, the robot makes its way to the orange, lays hold of it with its arms, and takes it to be deposited in the bin. You double-check on its capacity by moving the robot, the trash bin and the orange to another room and, as before, you find that the robot performs up to par. You carry on with similar checks and it turns out that with only a few exceptions the robot performs quite reliably to this pattern.

There is a clear sense in which this system gives evidence of being a goal-seeking, representation-guided system and makes a claim to count as an agent: it more or less reliably acts to realize a certain goal or purpose according to more or less reliable representations. The goal is that things on the floor are put in the bin; this counts as a goal insofar as it is a condition whose non-fulfillment prompts rectifying action on the robot's part. The representations are states in the agent that provide information about the environment; these count as representations insofar as they come and go with the presence and absence of the conditions on which they provide information. The robot is so organized that, depending on

whether or not its representations indicate that there is an object on the floor, it will act or not act; and depending on where the representations locate the object relative to robot and bin, they will guide its movements and other adjustments.

Or at least the robot is so organized that it will perform to this standard when independently plausible conditions of functioning are satisfied: when the lights are on, it is not misled by pictures of objects, its batteries are not run down, and so on. Assuming such conditions are met, the robot functions quite reliably in representing the environment and in acting for its goal in accordance with those representations. It is constituted so that in the absence of factors that plausibly impede its functioning, it reliably moves any objects on the floor to the bin area, operating on the basis of its reliable representational faculties. It displays that behavior in actual circumstances and in a range of variations on the actual circumstances where the goal remains relevant and attainable and its functioning—its forming and acting on its representations—is not impeded. The robot is marked out as an agent by the evidence of this robust, if conditioned pattern of behavior.[2]

Is such evidence sufficient in itself to ensure that the system counts as an agent? Not strictly, since the system may not be organized, as I put it earlier, so as to behave in the purpose-driven, representation-guided mode described; it may do so under purely external rigging. It may turn out to be following instructions, for example, from a spatially distant controller like a marionette (Peacocke 1983). Or it may be conforming to a look-up tree, implanted by a temporally distant controller who foresaw every situation the system might confront and pre-

programmed its response (Block 1981). But if the system is enduringly organized within itself so that it displays the required pattern of behavior, then there can be no room for doubt about its agential status (Jackson and Pettit 1990a; Jackson 1992).

The robust pattern of behavior displayed by the toy system of our example is about as simple as it is possible to imagine. But we can see that a similar story can hold as we go to more and more complex patterns, and more and more complex agents. The purposes pursued by agents may be multiple, and variously ordered. The representations formed by agents may extend into a number of sensory modalities, they may assume the form of memories as well as current representations, they may become abstract or propositional as well as concrete or perceptual, and they may include representations of how things might be as well as of how things actually are. And those representations, as well as the purposes they serve, may be endorsed in degrees, as well as in the on-off manner envisaged so far. The variations and developments possible are legion and are evident across the spectrum from simple to complex robots, from simple to complex non-human animals, and from non-human animals to our own kind. Still, despite complexities of these kinds, the category of agency retains its common form across those variations. Each of the systems imagined, no matter how complex, is organized so as to reliably promote certain goals or purposes under the guidance of reliable representations, when there are no factors present that impede its functioning (List and Pettit 2011, Ch 1).

This discussion of the nature of agency and the evidence for agency leaves one question unanswered. How robust does the conditioned pattern of behavior that is characteristic of agency have to be? Absent factors that impede its functioning, what range of variations ought to make no difference to the performance of an agent?

An extreme line on this question might be that no such variation ought to make any difference to the reliability of the agent in responding to evidence and executing its actions. But this is likely to be unrealistic with naturalistic, essentially limited subjects and I shall only assume that there is some threshold in variations, perhaps sensitive to context, such that it is enough for agency that a system proves to be evidentially and executively reliable beyond that threshold.

I do not have anything to say on where that threshold might lie but, wherever it lies, there are two fronts, internal and situational, on which any system must display the required degree of evidential and executive reliability (Pettit 2009). I defend two claims, bearing on these two forms of robustness. First, if the agent does not achieve internal robustness, there will be no reason to trace its behavior to states like representations and purposes as distinct from the many possible neural or electronic realizers of those states. And, second, if it does not achieve situational robustness, then the states to which we trace it, even if they are multiply realizable, will not be fit to count as representations and purposes.

The first claim is that the purposive and representational states or attitudes that are taken to prompt the agent's behavior must do so over a range of possible variations in how they are realized within the system. For example, the kitchen

robot does not have to be given evidence of an object on the floor that strikes its bug-like eyes from just one particular angle, making one particular retinal impression and triggering one particular computational process. It responds appropriately no matter what the angle of vision and no matter what the retinal impression and computational process that realizes its representation of the object. In the absence of robustness over variations in the internal, physical realizers of representations and indeed purposes, there would be no reason to posit representations and purposes at the origin of the behavior; there would be no reason to posit anything other than the physical realizers themselves. The claim of purposive and representational attitudes to causal relevance consists in their programming for behavioral responses: that is, in their leading to the responses over variations in how they are realized at lower levels (Jackson and Pettit 1990b; List and Menzies 2009). Absent internal robustness, there would be no grounds for taking them to have any such causal relevance to the behavior produced.

The second claim is that an agent must display a purposive-representational pattern of behavior over situational as well as internal variations. Assume that impeding factors are absent. In order for a state to count as a representation that p it must form in response to evidence that p and unform in response to evidence that not p. And in order for a state to count as a purposive state of seeking to X, it must prompt different behaviors under different representations as to the opportunities and means of X-ing. This means that the representational attitude must form and unform in response to evidence, even when there are other variations in situation, and that the purposive attitude must

prompt suitable initiatives over parallel variations; otherwise they would not count respectively as representational and purposive. Absent impeding factors, then, to take a system to act for a certain purpose according to a certain representation is necessarily to assume that it would do so over suitable situational variations: that is, variations in which the representation continues to be supported and the purpose continues to be capable of implementation.[3]

4.2. Two modes of agency and agency-detection

4.2.1. Personal and non-personal agency

While the considerations in the last section introduce the basic conception of agency with which I shall be working here, they ignore the fact that there are two modes of agency that stand in deep contrast with one another. On the one side there is what I shall describe as the non-personal agency exhibited by the toy robot—and, I suspect, all other robots and all non-human animals. And on the other there is the personal agency that we human beings generally display. Non-personal agency, as should be clear already, comes in many varieties and appears at many distinct levels of sophistication; there is a deep gulf between even the family pooch and the kitchen-cleaning robot. But variegated as it is, non-personal agency still contrasts in the deepest possible fashion with our own personal form of agency.

In order to bring out the distinctive character of personal agency let me rehearse in a set of dot-points certain things that you—and human beings in general—can more or less clearly do but which no animal or robot can approximate (Pettit 1993; McGeer and Pettit 2002). These points are inevitably

telegraphic, given restrictions of space, but I hope that they are intuitively clear and plausible.

- Like robots and other animals, you can form purposes and representations, beliefs and desires, relying on your non-intentional, usually unconscious processing—for short, your sub-personal processing—to guide the formation of those attitudes under the flow of incoming evidence, perceptual and otherwise. But you can also do much more.
- You can assent to, dissent from, or suspend judgment on sentences or propositions that express attitudes you hold or might hold; and you can do this in light of considering the evidence for and against those propositions. A proposition expresses a certain attitude when acting as if the proposition were true—acting as if things were as it says they are—amounts to acting according to that attitude. In this sense 'p' expresses the belief that p; '"q" is attractive' expresses the desire that q (as well as the belief that "q" is attractive); and 'I will do X' expresses the intention to X (as well as the belief that you will do it).
- In passing judgment in this way on a proposition, you reveal—or perhaps make it the case for the first time—that you hold the corresponding attitude. Exercising judgment over propositions is a way of forming or revealing attitudes that is distinct from the spontaneous, sub-personal way of forming attitudes associated

with basic agency and the behavioral mode of revealing the attitudes that you form in that way.

- Many of your attitudes will be spontaneously formed, of course, and perhaps never become associated with judgment. But on pain of not being an interpretable agent—even an interpretable agent for yourself—there had better be a general coherence between the attitudes that form spontaneously within you and the attitudes formed or confirmed via the exercise of judgment. In particular, the attitudes you hold spontaneously ought to expand or contract or alter in response to the judgments you make. Were they to come regularly apart, then you would have two inconsistent profiles as an agent.
- While coherence is likely to be generally assured by your subpersonal make-up, judgment may come apart from attitude in particular cases. You may make your judgment without sufficient attention to the evidence and your spontaneously formed attitudes, being better attuned to evidence, may not vary as a result of the judgment. Or you may make your judgment thoughtfully, as when you come to reject the gambler's fallacy, but your spontaneously formed attitudes may not fall in line: you may forget yourself at the casino table (McGeer and Pettit 2002). But you can guard against this occasional incoherence between assent and attitude by taking

measures to ensure greater care in making judgments and greater caution in acting on related beliefs.

- Assuming coherence between judgment and attitude, the fact that you assent to ‘p’ or dissent from ‘p’ will indicate that you hold the attitude that ‘p’ or ‘not-p’ expresses: the act of assent or dissent will induce or perhaps reveal that attitude within you, ensuring the presence of a disposition to manifest associated patterns of behavior. And, assuming coherence, the fact that you suspend judgment on whether or not p will indicate that you hold neither attitude: you have an open mind.
- All of this being so, you are able to prompt the formation of attitudes in any area, or at least reveal their presence—that is, bring them to consciousness—by resort to judgment: by seeing whether or not the available evidence leads you to give assent to relevant propositions. You can intentionally make up your mind, as we say, passing judgment on whether the weather is improving or Pythagoras’s theorem is sound; on whether it would be fun to go to town or whether to take a break.

It should be clear that the capacities at which I gesture here mean that you and human beings in general are very different from other sorts of agents. The regular agent is at the mercy of the beliefs and desires and intentions that happen to form within it—and at the mercy of how sensitive they are to evidence—acting under the ebb and flow of their influence. But as a human being you are able to have a

sort of intentional control over whether or not you form certain beliefs and desires in a given area, over how well the attitudes you form are faithful to available evidence and over whether they satisfy related conditions: whether they are consistent with one another, and whether they are closed under entailment.

What desires are likely to guide you in the exercise of such intentional control? You will want to form beliefs and other attitudes in any domain where you are required or otherwise motivated to act; this will be necessary for shaping what you do. And as a prerequisite of satisfactory agency you will want to form beliefs and other attitudes that are faithful to the evidence, consistent with one another and even to some extent closed. Any failure in such regards is liable to limit your capacity to perform as an agent, your ability to act effectively for whatever purposes you happen to embrace.

The control you can exercise on these lines is essentially epistemic, allowing you to determine the matters on which you form beliefs and other attitudes and to promote the broadly evidential quality of the attitudes you form. But there is also another sort of control, evaluative rather than epistemic in character, which your expressive and judgmental capacities as a human being ought also to make possible. This is control over what purposes you embrace rather than control over how you pursue those purposes.[4]

On the picture sketched you are able with any desire you have—say, the desire that p—to register and assent to the proposition that ‘p’ is an attractive prospect; you can judge and believe that that is so at the same time as you are attracted to the prospect. But suppose, plausibly, that experience gives you a basis

for judging that while the p-prospect is attractive here and now, it is not reliably or robustly attractive. Like the gratification of a passing impulse, it is a prospect that you will wish you hadn't sought as you look at what you chose from the perspective of a later self or a perspective that you share with other people. If you can now predict and privilege your standpoint as an intertemporally enduring, interpersonally engaged self, forming beliefs about what is robustly attractive, it will be rational to take a critical attitude towards your current desire. And human experience suggests that by taking a critical attitude—by forming the belief, under epistemic control, that the prospect is not attractive in a suitably robust way—you can exercise a distinct evaluative control over your desires and purposes; the beliefs you form may provide you with the means of disabling offensive desires or prompting more satisfactory alternatives (Smith 1994).

Assuming that you have control over how far the attitudes you form on the basis of available evidence are epistemically and perhaps evaluatively satisfactory, you will meet standard conditions for being fit to be held responsible—fit to be praised or blamed—for the formation of relevant attitudes and for the deeds they lead you to enact (Pettit and Smith 1996). Faced with the issue of whether or to form a belief that p, it will be intuitively up to you whether you are attentive to the evidence; you will have a capacity to promote such attention, even if you fail to exercise it. And faced with the issue of whether to form a desire that q or that not-q, it will equally be up to you whether you form a desire that accords with your beliefs about the robust attractiveness—the desirability—of the prospects; again you will have the required capacity, even if

you fail to exercise it. In each case, then, you will be fit to be held responsible for performing well or badly by epistemic and evaluative standards in the attitudes you embrace or fail to embrace.

In ordinary parlance, this is to say that you will be personally responsible for the attitudes you form—or fail to form—as distinct from just being causally responsible for them. Even the simple robot or animal is causally responsible for the attitudes it forms, since it is the sub-personal processing of the system that produces those attitudes, updating in response to the evidence it confronts. But you will be personally responsible for relevant attitudes insofar as you can be called to book for them: you can be exposed to praise or blame for what you do or do not believe or desire or intend—and for how you consequently act or fail to act—whether on an epistemic or evaluative basis. It is this dimension of personal responsibility that leads me to describe the sort of agency that you and other human beings display as a personal form of agency, distinguishing it from the non-personal agency of simpler systems like robots and other animals.[5]

4.2.2. Two ways of detecting agency

In determining whether a simple system like our robot is an agent we rely on induction from the evidence of how it interacts with what we may describe as an impersonal environment: how it performs in the limited range of cases that we explore as we put different objects at different places on the floor, or as we observe its reactions to differences introduced by other hands. It will give evidence of being an agent just insofar as it is disposed to act after a certain pattern in an indefinite range of possible scenarios, of which the limited range

explored is a subset. The limited range offers an inductive basis for ascribing the wider disposition. The existence of the disposition, realized in its internal organization, offers the best explanation for why it behaves as it does in the cases actually investigated.

We know from long-established psychological studies that we human beings have a powerful tendency to look for agency, being prompted to ascribe it even in cases where the systems involved—for example, the geometrical shapes in a simple, cartoon movie (Heider and Simmel 1944)—are manifestly incapable of agency. We are hair-triggered to move from the most slender behavioral evidence to the postulation of the robust capacities that agency requires. It's as if we are pre-programmed to be animists. We worry about overlooking any agents that may inhabit our world and, for the sake of avoiding that possibility, we routinely run the risk of taking many non-agential systems to be agents proper.

But despite this readiness to leap to ascriptions of agency, we don't primarily rely on induction from interaction with an impersonal environment when we ascribe agency to other human beings. As human beings we are personal agents. And as personal agents we have a special basis for recognizing the agency of other personal agents, at least to the extent that we share expressive resources. What we mainly rely on in ascribing agency to other human beings is induction from their interpersonal interaction with other persons, whether we engage directly in that interaction ourselves or have indirect evidence of the interaction in their relations with others.

In order to see how we can gain access to your agency, recognizing the presence of suitable attitudes, I add some dot-points to the list already constructed. These register ways in which we, as engaged interlocutors, can come to determine the agential presence and operation of attitudes of belief, desire, intention and the like. They reflect capacities that are more or less clearly within the capacity of any normal human being and within your capacity in particular.

- Assuming that you can make up your mind on certain propositions, determining your own attitudes, you can know your mind on those matters other than by reviewing yourself introspectively. You can test yourself on your response to an arbitrary proposition and depending on how you judge, you can know whether or not you believe the proposition—and in relevant cases hold or do not hold a corresponding desire or intention.
- On those matters where you make up and know your mind, you can speak for yourself by making up your mind and displaying that knowledge publicly in assertion. If you assert that ‘p’ then, assuming sincerity, that will manifestly communicate that you have knowingly made up your mind that p and that you believe that p; it will amount to avowing the belief, as we say.
- Communicating by avowal that you believe that p—or have any other attitude—contrasts with communicating your belief by reporting that you believe that p: it seems to you, as you might put it, that you believe that p. With a report you can excuse a later

failure to act as if p in either of two ways: by explaining that the introspective evidence on your belief misled you; or by explaining that you changed your mind, say by discovering new perceptual or other evidence against ‘p’.

- In avowing a belief that p, communicating that you have made up your mind, you will communicate at the same time that you cannot excuse a later failure to act as if p by the claim that the introspective evidence on your believing that p was inadequate or misleading; that would be inconsistent with your having made up your mind which, in traditional terminology, gives you a maker’s rather than a reporter’s knowledge of your attitude. The change-of-mind excuse will remain available but not the misleading-evidence excuse.
- With any belief and desire or intention on which you can make up your mind, you have a choice between avowing and reporting it.[6] The fact that you manifestly and voluntarily avow it rules out excusing a failure to display the attitude by appeal to misleading evidence on your attitudes, as a reporter might excuse such a failure; it amounts to a commitment not to try to escape that cost, should it be incurred.
- Since the avowal of an attitude is more costly than reporting the attitude—reporting it as you might report the attitudes of another—it is also more credible than a mere report; and being more credible

it is likely to be more appealing: it will have a better chance of shaping the expectations of your audience and coordinating with them to your mutual benefit.

- You can make your ascriptions of future actions more credible and more appealing in a parallel way, by strengthening an avowal into a promise. Like the avowal, the promise rules out the excuse of having been misled about your attitudes when you fail to act according to an attitude previously avowed. But it also rules out the excuse of having changed your mind since making the avowal. Promise that you'll meet me at the theater and you cannot claim in later excuse either that you got your intention wrong or that a better opportunity presented itself and you dropped the intention.[7]

What these points emphasize is that if you are a personal agent, then there are very exacting expectations to which we, your interlocutors, will hold you. We will expect you in suitable areas to be able to speak with authority to what you believe and desire and intend; to be willing to make commitments to us— avowals or promises—and not just to report on yourself as you might report on another; to prove capable of living up to those commitments in the general run, displaying the beliefs and desires, the intentions and actions, to which your words testify; and, where you occasionally fail to live up to those words, to be able to recognize the failures and to be willing to make excuses or apologies that suggest a determination to improve. Being a personal agent, you will be expected to prove

yourself a conversable agent too: someone we find it possible to reach in the realm of words and to engage to our mutual benefit.

The fact that we tie your agential status to such a rich array of expectations means that if you are not an agent—or at least not an agent for whom your words speak—then that will show up very quickly. And the fact that those expectations are very exacting means that as you begin to meet the expectations, it will quickly become plausible that you are an agent.[8] Induction plays a central role in the exercise of establishing that you are a personal, conversable agent but the exercise is very different from the inductive procedure that we have to follow with the robot in our earlier example. To put the difference in a slogan, it involves induction from interpersonal interaction rather than induction from impersonal interaction.

We see that you are a personal agent in virtue of probing your attitudes, eliciting avowals and promises, and finding that you do not let us down: that is, in virtue of vindicating your status in interaction with us. Or we see that you are an agent by learning of the pattern of interpersonal interaction that you enjoy with third parties. We rely just on induction from evidence of impersonal interaction in the case of a non-personal agent like the robot, whether this be a form of interaction we sponsor in experiment or register as mere observers: whether in that sense it be direct or indirect. But in the case of personal agents like you and any other human being we can also rely on induction from evidence of interpersonal interaction, whether this be interaction in which we directly engage or interaction with third parties that we learn about indirectly.

4.3. Recognizing group agents

The discussion so far suggests that if groups are agents, then they may be non-personal agents like robots and animals or personal agents like you and me. And it suggests that whether they count as agents of one or the other kind will correlate with the sort of evidence we find appropriate for establishing their agency. I argue for two theses in this section, one positive, the other negative. First, that we can certainly establish the agency of some groups by finding them conversable in the manner of personal agents: that is, by interacting with them interpersonally or having evidence of such interaction with others. And second, that we cannot plausibly establish the agency of any group just on the basis of evidence, direct or indirect, of an impersonal form of interaction. The upshot is that the only groups we can plausibly expect to count as agents are groups that succeed in attaining conversability.

4.3.1. Ascribing group agency on the basis of interpersonal interaction

Most of the groups that make a persuasive claim to count as agents speak for themselves in the manner of individual human beings, having individuals or bodies that serve as corporate spokespersons. In claiming to speak for the group—that is, for all the members—on any issue, such spokespersons lay claim to an authority, based on the individual acquiescence of members to live up to the words they utter on the group's behalf. Thus the agential status of the group will be manifest in the fact that the declarations that the spokespersons make are ones that other members honor, acting as the words require of them, now in this situation, now in that. As we deal with the group through its spokespersons, we

find that it vindicates its status as an agent by how it interacts with us. And we find no difficulty in this, since the authority claimed and manifested by spokespersons testifies to explicit or implicit commitments on the part of members—presumably capable of confirmation in individual interaction with them—to abide by the utterances of suitable representatives.

The spokespersons for any group may be individuals or assemblies of individuals and while no group need have the same spokesperson on every issue, different spokespersons must speak with a single voice, ensuring by whatever means that the avowals and promises they make on behalf of the group form a coherent whole. The group's accepted mode of organization and decision-making will usually ensure this coherence among spokespersons, as it will ensure that members know how they are required to behave by the utterances of such authorities (French 1984; List and Pettit 2011). A group agent may fail on occasion to live up to those utterances, of course, as an individual may fail too. But the mode of organization ought at least to make it capable in such a case of proving responsive to complaints about the breakdown, enabling it to recognize when an excuse is available, or an apology due, and to act accordingly.

What sorts of declarations do spokespersons make on behalf of a group agent? They avow the beliefs of the group, as when the church outlines its tenets of faith, the political party presents its analysis of the economy, or the corporation explains why its profits fell in a recent quarter. They avow equally the wishes and values and intentions of the group as when the church expounds what it stands for, the party embraces certain principles or policies, and the corporation endorses

a strategic statement and a statement of medium-term tactics. And they promise future action in one or another domain, as when the church promises greater openness about priestly abuse, the party commits itself to one or another initiative in government, and the corporation enters contracts with its suppliers and customers.

Is it excessive to take the declarations of spokespersons to be avowals and promises? Absolutely not, for the declarations are taken in common usage to rule out excuses of misleading evidence or change of mind in the way that is characteristic of avowals and promises. Suppose a group fails to live up to a belief or value ascribed by an authorized spokesperson. It will not do for that spokesperson to excuse what was said on the grounds of having mistaken the evidence about what the group held. The spokesperson's only recourse will be to resign from the role assigned by the group or, maintaining that role, to try to offer another excuse for the failure or to make an apology on the group's behalf. Or suppose a group fails to live up to a promise that the spokesperson made on its behalf. In this case the spokesperson can invoke neither the misleading-evidence excuse nor the change-of-mind excuse. Again the alternatives will be as stark as before: resign, excuse on other grounds, or make an apology.

Let us assume that a group designates unique spokespersons in different domains, then, and that it robustly lives up to the words of its spokespersons. And let us assume, as this implies, that the voice supported by the different spokespersons is reasonably coherent, offering a self-consistent, if developing story of the group's attitudes. Or let us assume at least that, when the voice fails to

be coherent, the spokespersons respect the demand to speak with one voice, making amendments that restore coherence. If such conditions are fulfilled, then there can be little doubt about the grounds for treating the group as an agent. The words of the spokespersons project a robust pattern of goal-seeking, representation-guided action and the group is systematically organized to live up to those words and keep faith with the projected pattern.

More specifically, the words of the spokespersons project that robust pattern—that pattern of evidentially and executively reliable performance—on the two fronts, internal and situational. On the internal front, they give us evidence that the members of the group will perform appropriately, living up to what the group demands of them, across a raft of variations in their personal attitudes: any variations, at any rate, that are consistent with their remaining committed to the group. And on the situational front, they indicate that the members of the group will perform appropriately as circumstances change, giving rise to a change in what attitudes are supported or what action would be appropriate for enacting the group's attitudes.

We have grounds at least as solid as in the robot case for treating such a collectivity as an agent. Moreover, indeed, we have grounds that entitle us to treat it as a personal agent that can be held responsible for its attitudes, given the capacity it must have to take account of epistemic or evaluative critiques of the attitudes it embraces. The spokespersons that speak for the group, whether these be individuals or assemblies, will presumably be as capable of responding to such challenges when they act for the group as they are when they act for themselves.

They may refuse on occasion to answer a particular challenge but the pressures of credibility on any group that claims to support coherent attitudes, and to invite relationships with individuals and with other groups, will argue for not making a habit of such refusal. Within its domain of operation, it must purport and prove itself to be a conversable subject: an entity capable of being reached and engaged in speech.

The conditions identified in these observations are satisfied over and over in the social world. Our societies teem with commercial, ecclesiastical, associational and political groups, each with its own mode of organization, its own way of generating a single, self-representative voice and its own way of guaranteeing fidelity in action to the words uttered in its name. As the law of incorporation has grown over the last century or so, these entities have become ever more powerful, gaining a capacity to act in different areas, to change their area of action as they will, to adopt and amend the goals that they pursue there, and to do all of this on the basis of resources that are strictly corporate, with the liability of members for group bankruptcy being severely limited.

4.3.2. Ascribing group agency from evidence of impersonal interaction

Evidence of interpersonal interaction, direct or indirect, would clearly be sufficient for thinking of certain groups as agents: specifically, as conversable agents. But is evidence of interpersonal interaction necessary for establishing the status of a group as an agent? Or might the evidence of impersonal interaction alone suffice to establish a group's claim to agency, and presumably to agency of a non-personal kind? Might we be reasonably led to cast a group as an agent just

by finding that it displays an agential pattern of interaction of broadly the kind illustrated by the robot? In particular, might we be reasonably led to do this without making assumptions that can only be confirmed by recourse to evidence of interpersonal interaction? I argue that the answer is, no.[9]

Every group, in the nature of the case, is made up of individual human beings, each with a mind of their own. Thus whatever goal-seeking patterns are postulated at the group level, they have to emanate from individual actions: the actions whereby some or all of the members do their bit, whatever that is, in sustaining the group-level patterns. And whatever representations are supposed to guide the group in its fidelity to those patterns, they have to be formed on the basis of representations formed in some or all of its members: the members, after all, are the group's eyes and ears. If we are to treat a group as an agent, then there has to be good ground for expecting that it will robustly display any goal-seeking, representation-guided patterns we postulate. And that means that there has to be good ground for expecting that it will do this over possible variations in how, independently of the group, members individually see things and are individually disposed to act. It has to be evidentially and executively reliable, as we put it earlier, over certain variations on the internal front: that is, in the individuals who make it up.

The dependence of the behavior of a group on the intentional profiles of its members generates a dilemma for anyone who thinks that observing the impersonal interaction of a group might be enough on its own to provide adequate evidence of group agency. Suppose that we come across a group such that its

interaction with an impersonal environment—and such impersonal interaction only—suggests that there is a purpose or set of purposes that it is pursuing in light of representations it forms about the opportunities and means of action at its disposal. Either the behavior of members of the group will be intelligible just in light of their individual profiles—their group-independent beliefs and desires. Or the behavior of the group will not be intelligible on that basis. And in neither case are we likely to think that the interaction of the group with its impersonal environment provides sufficient evidence for casting it as an agent.

If the behavior of the group is intelligible in light of the group-independent, individual profiles of members, then there will no reason to postulate a group agent, since the pattern displayed by the group as a whole will not be robust over relevant sorts of internal variation within the group: that is, variations in the group-independent profiles of the members. Take the example of a market in some domain of commodities, which advances the purpose of establishing the relative prices at which those goods can be successfully cleared, in light of information about—and, we might think, a representation of—the level of aggregate demand. This apparently purposive-representational pattern ought not to lead us to think of the market as an agent. The pattern is only as robust as the group-independent desires of members to trade with one another at maximal returns to themselves.[10]

Let us turn now to the second possibility, that the pattern displayed in a group's interaction with an impersonal environment—a pattern like that illustrated in the market—is not intelligible in light of the group-independent,

individual profiles of its members. Might the evidence of such a pattern suffice on its own to establish the agency of the group? I do not think so. We would hardly find that pattern compelling unless we had some explanation as to why members should support it, given that they may be disposed by their group-independent attitudes to act in an unsupportive manner. And the only explanation that would have any plausibility in such a scenario would require confirmation, direct or indirect, by reference to interpersonal interaction. This explanation is that the members are committed to live up to avowals and promises made in their name: that in that sense their behavior is determined by group-dependent, not group-independent, attitudes.

Imagine that you are hovering in a helicopter and watching the rush hour traffic clog the main highway out of town. And suppose you notice that the line of traffic is systematically blocking an ambulance from crossing that highway. All the crossings give priority to the highway and you see the ambulance being blocked, now at this crossing, now at another, now at a third. You might think of the traffic as a group agent that aims at frustrating the ambulance; after all, the evidence of its impersonal interaction with the ambulance suggests that that's what it is doing. But could you sensibly reach that conclusion just on the basis of such evidence? I think not.

The problem is that, whatever the group-level evidence, you are bound to assume that individual drivers each have group-independent attitudes of their own; you are hardly going to take them to be automatons or zombies. And under that assumption it would be a miracle—a cosmic accident—if the attitudes

robustly fell in line with the requirements of the alleged group goal. The only basis on which you could reasonably conclude that the traffic had an agential character, with the frustration of the ambulance as a goal, is the belief that the individual drivers are committed, as under a rule of authorized spokespersons, to the service of that group-level goal. You might not be clear about how they could be guided by a spokesperson and might even be forced to postulate channels of hidden electronic communication. But any such postulate, no matter how unlikely, would be more reasonable than taking them to constitute an agent, without reliance on the possibility of confirmation by direct or indirect evidence of interpersonal interaction.[11]

4.3.3. The bottom line

The considerations in this section suggest that under plausible epistemic scenarios, the only evidence that we can take as determinative of the presence of a group agent is evidence of interpersonal interaction. That means that the only sort of group we are ever likely to recognize as an agent is a conversable body. Such a group will have spokespersons that maintain a single voice and a mode of organization that gives credibility to their words, prompting other members to keep faith with those words when they act in the name of the group. It will count as a personal rather than a non-personal agent.

This line ought not to be surprising in view of the history of the concept of group agency. The concept emerged in medieval Europe, where guilds and orders and other novel entities flourished, and it quickly gained a wide currency. It applied to any group of people who united together in such a way that collectively

they appeared in law, and figured in the courts, in the manner of an individual subject. The paradigm example was the group that could own property and enter contracts, sue others and be sued in turn, and operate legally in the manner of an individual agent. What struck the legal theorists of the time was that in an entity like a guild or parish or town certain individuals or assemblies were authorized to speak for the corporate body, avowing the judgments or purposes of that body on the basis of the authority vested in them, and being entitled to speak for the body in promising to take one or another action. Such spokespersons were expected to maintain a coherence of voice, not holding by inconsistent claims or plans. And ordinary members of the body were required under the rules of incorporation to keep faith with the words given in their name, living up to the avowals and promises that their spokespersons made.

Congenially with the view developed here, the authorities or spokespersons in this image were generally cast as playing a representative role, and groups were held to perform as agents just to the extent that the members rallied behind the words of their representatives. Thus in 1354, Albericus de Rosciate could say that a collegial agent, although it is constituted out of many members, is one by virtue of representation: *collegium, licet constituatur ex pluribus, est tamen unum per representationem* (Eschmann 1944, 33, fn 145). The theme dominates the work of legal theorists of the time like Bartolus of Sasseferrato and his pupil, Baldus de Ubaldis, who make much of the way a suitably represented group, in particular the represented people of a city, could figure as a corporate agent or person (Woolf 1913; Canning 1983). Arguing that

the *populus liber*, the free people of a city republic, is a corporate person, Baldus explains that this is because the council—the representative, rotating council—represents the mind of that people: *concilium representat mentem populi* (Canning 1987, 198).

This conversability criterion makes clear why churches and political parties and commercial firms are certainly group agents. But, to go back to earlier examples, the model makes equally clear why there is no temptation to ascribe group agency to the bond market, or the X-generation, or even the group of people who coordinate their efforts to save a swimmer.

There is no purpose pursued as such by the bond-market or the X-generation, no pressure on them to agree on representations to guide the pursuit of that purpose, and so no basis for expecting them to perform robustly as agents in their own right. But what of the beach group? The members in this sort of group do have a shared purpose, and do agree on the means of furthering it, and there may seem to be a better case for treating it as an agent.

On reflection, however, it should be clear that this sort of group will not constitute a group agent either. There is no reason to expect such a group to display the internal or situational robustness that we associate with agency. We might be able to predict on the basis of the individual character of the members that faced with a similar crisis in the mountains, they would almost certainly respond in some equally altruistic way. But there would be nothing about the group as such—nothing about its authorization of spokespersons or the mode of

its organization—that would support such extrapolation across changes of situation, let alone changes in its internal make-up.[12]

4.4. The reality of group agents

Despite the fact that the argument provided appears to support the reality of conversable group agents—and only indeed of group agents of that kind—a common approach suggests that still such agents should count only as fictions. They may perform as if they were agents but really they are not. They are merely the projections of the individual agents who make them up; they are the fronts or avatars behind which the members, who are the only real agents, operate for their own purposes.

This sort of fiction theory goes back to Thomas Hobbes (1994, Ch 16), in particular to his discussion of how a group of individuals can authorize a spokesperson to speak for them, thereby constituting a conversable body, capable of making and living up to commitments (Skinner 2010). Hobbes, a seventeenth-century philosopher, stands out among his predecessors for insisting that the spokesperson that speaks for a group has to be capable of performing as a pre-existing agent or agency. His idea is that a corporate agent will form just insofar as such a pre-existing agent or agency takes on the role of spokesperson for members of the group. ‘A multitude of men are made one person, when they are by one man, or one person, represented’. But their spokesperson or representative may be a committee, not just an individual, provided that the committee forms its judgments by majority vote, making suitable accommodation for ties: ‘if the

representative consist of many men, the voice of the greater number must be considered as the voice of them all'.

Hobbes assumes that the spokesperson for any group agent exists prior to the formation of that entity as an independent individual or committee and provides unity for the group agent insofar as its words can be treated as the words of the group, not 'truly', but 'by fiction'.[13] Thus in order to deflate the representation whereby group agents form—he often describes this as personation—he insists that it involves nothing more than the representation whereby an individual may speak for a wholly inanimate object, as in asserting its rights. 'There are few things that are incapable of being represented by fiction. Inanimate things, as a church, a hospital, a bridge, may be personated by a rector, master, or overseer'.

Does the account given here support the sort of fiction theory that Hobbes espouses and that continues to be espoused in contemporary circles, particularly among economists and economically minded lawyers (Grantham 1998)? No, it does not. It is possible in principle for a group agent to form around a single, dictatorial spokesperson, as Hobbes envisages, but this would be a degenerate case of group agency; it might be as well cast as an example of an individual agent with a multitude of helpers. And, even more importantly, it is not possible for a group agent to form around a single, majoritarian committee, whether this be an elite committee or a committee of the whole.

Hobbes assumes, as many assumed before and since, that a committee can function like an individual agent, mechanically generating its judgments and

purposes from the bottom up via majority voting. That is why he thinks that a committee can serve like a dictator to speak for a group and establish it as a conversable agent, capable of entering and keeping commitments. But it turns out that he is quite mistaken about that, as the discursive dilemma makes clear (Pettit 2001, Ch 5; List 2006). A majoritarian committee cannot reliably function like an independent agent, in the way Hobbes envisages, because majority voting among individually consistent members can generate inconsistency in the group judgments on various connected issues.

Suppose that you, Bloggs and I want to form a group agent and that we must decide on the attitudes of the group on three propositions, 'p', 'q' and 'p&q'. You and Bloggs may vote for 'p', I against, and Bloggs and I for 'q', you against. How then will we cast our votes on 'p&q'? You and I will vote against and only Bloggs vote for. Thus our majority voting pattern will lead us as a group into embracing, incoherently, the package: p, q, not-p&q. We will then face a discursive dilemma. Let that package stand and we must reject the aspiration to collective rationality. Alter the package so as to ensure collective rationality and we must reject the aspiration to individual responsiveness.

This simple observation shows that the majoritarian committee cannot be recruited to the role of a spokesperson, allowing the group for which it speaks to count as an agent. In order for the three of us to establish a group agent we have to follow a procedure that targets the requirements for such an agent to exist, ensuring in particular that it is reliably consistent and coherent. Thus we might follow a straw-vote procedure under which every attitude supported by a majority

vote is checked for consistency with other attitudes adopted; if it is consistent, we endorse it; and if it is not consistent, as in the case illustrated, we make a decision on which member of the conflicting subset to reject, whether that be the new candidate or something accepted in the past. This might lead us as a group to endorse the claims ‘p, q, p&q’, as it might have led us to endorse ‘not-p, q, not-p&q’, ‘p, not-q, not-p&-q’. In any such event it will enable us to preserve collective rationality, and to allow us to form a group agent, but require us to sacrifice individual, majoritarian responsiveness; there will be at least one proposition we endorse as a group that a majority of members individually reject.

The group agent that we might form in this way, via the straw-vote procedure, is not an agent that we, an independently existing agency, go through the motions of representing, giving it a fictional existence. No, it is a group agent that comes into existence by dint of our individual efforts, in particular our efforts to ensure that the conditions for the existence of the group agent are met. As individual, bottom-up voting proceeds, we gather feedback on the emerging pattern of judgments and purposes that this would generate for the group and, when necessary, we act top-down to ensure that that pattern is fit for agency: we suspend the effect of a vote and revise the overall results of voting, past and present, so as to ensure our coherence as a group agent. Before the appearance of that group agent, we exist as individuals, of course, being required to bring the group into existence. But before its appearance there is no other agent or agency—nothing like the dictatorial spokesperson—such that by contrast with that entity the group agent created is merely a fiction.

The discursive dilemma shows that one particular pattern of bottom-up responsiveness to member votes— that which majority voting would ensure—is ruled out by the requirement of collective rationality and so that a majority committee could not play the agential, representative role that the fiction theory of group agents requires. But could a group agent be represented in any other bottom-up way—say, under another other voting system—by a single committee or indeed by a network of committees with complementary tasks? No, it could not. The recent impossibility theorems on judgment-aggregation generalize the lesson illustrated by the discursive dilemma and support that negative line (List and Pettit 2002; List and Polak 2010). They show, roughly, that when individuals construct a group agent—a reliably rational entity—the exercise will be effective only if the judgments and purposes they assign to the group are not constrained to be a reflection, majoritarian or otherwise, of the corresponding attitudes of the members. And that means, as in the straw-vote case, that the individuals have to construct a group agent *de novo*: they have to construct an agent such that there is no pre-existing agency—no pre-existing spokesperson—in comparison with which it might look like a fiction.

And so to the denouement. It may be, as we saw earlier, that the only plausible basis for ascribing agency to groups is evidence of interpersonal interaction, and that only groups whose members organize to make them conversable have a claim to constitute agents. But still, so our concluding observations suggest, such a group agent is not just a fiction or pretense: a dummy agent that reflects only the voices of a ventriloquist master, in the way in which

the dicatatorial agent would reflect the voice of the dictator. Any group agent will be the same collection as the set of its members at or over time, since it does not have any existence apart from them. But it will not be the same agent.[14] Indeed, prior to the formation of the group entity, the collection of individuals who construct it will not be an agent of any kind; as Hobbes would say, it will be merely a multitude.[15]

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.

Canning, J. (1987). *The Political Thought of Baldus de Ubaldis*. Cambridge, Cambridge University Press.

Canning, J. P. (1983). "Ideas of the State in Thirteenth and Fourteenth Century Commentators on the Roman Law." *Transactions of the Royal Historical Society* 33: 1–27.

Eschmann, T. (1944). "Studies on the Notion of Society in St Thomas Aquinas: St Thomas and the Decretal of Innocent IV Romana Ecclesia, Ceterum." *Medieval Studies*: 1–42.

French, P. A. (1984). *Collective and Corporate Responsibility*. New York, Columbia University Press.

Gilbert, M. (2001). "Collective Preferences, Obligations, and Rational Choice." *Economics and Philosophy* (17): 109–120.

Grantham, R. (1998). “The Doctrinal Basis of the Rights of Company Shareholders.” *Cambridge Law Journal* 57: 554–88.

Heider, F. and M. Simmel (1944). “An experimental study of apparent behavior.” *American Journal of Psychology* 13.

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

Jackson, F. (1992). Block’s Challenge. *Ontology, Causality, and Mind: Essays on the Philosophy of David Armstrong*. K. Campbell, J. Bacon and L. Rhinehart. Cambridge, Cambridge University Press.

Jackson, F. and P. Pettit (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.

Jackson, F. and P. Pettit (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.

List, C. (2006). “The Discursive Dilemma and Public Reason.” *Ethics* 116: 362–402.

List, C. and P. Menzies (2009). “Non-reductive physicalism and the limits of the exclusion principle.” *Journal of Philosoophy* 106.

List, C. and P. Pettit (2002). “Aggregating Sets of Judgments: An Impossibility Result.” *Economics and Philosophy* 18: 89–110.

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

Comment [OUP-CE1]: AU: Please provide page range for the reference. 243-59

Comment [OUP-CE2]: AU:Please provide page range for the reference. 475-502

List, C. and B. Polak (2010). “Symposium on Judgment Aggregation.” *Journal of Economic Theory* 145 (2).

McGeer, V. and P. Pettit (2002). “The Self-regulating Mind.” *Language and Communication* 22: 281–99.

Peacocke, C. (1983). *Sense and Content*. Oxford, Oxford University Press.

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

Pettit, P. (2001). *A Theory of Freedom: From the Psychology to the Politics of Agency*. Cambridge and New York, Polity and Oxford University Press.

Pettit, P. (2007). “Rationality, Reasoning and Group Agency.” *Dialectica* 61: 495–519.

Pettit, P. (2009). The Reality of Group Agents. *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice*. C. Mantzavinos. Cambridge, Cambridge University Press**:** 67–91.

Pettit, P. and D. Schweikard (2006). “Joiint Action and Group Agency.” *Philosophy of the Social Sciences* 36: 18–39.

Pettit, P. and M. Smith (1996). “Freedom in Belief and Desire.” *Journal of Philosophy* 93: 429–49; reprinted in F.Jackson, P.Pettit and M.Smith, 2004, Mind, Morality and Explanation, Oxford, Oxford University Press.

Searle, J. R. (1983). *Intentionality*. Cambridge, Cambridge University Press.

Skinner, Q. (2010). A Genealogy of the Modern State. London.

Smith, M. (1994). *The Moral Problem*. Oxford, Blackwell.

Comment [OUP-CE3]: AU: Please provide page range for the reference. Not relevant. Whole issue

Tuomela, R. (1995). *The Importance of Us*. Stanford, CA, Stanford University Press.

Woolf, C. N. S. (1913). *Bartolus of Sassoferrato*. Cambridge, Cambridge University Press.

---

[1] As Chad McIntosh has reminded me, a group might be defined so that it is required to have certain individuals as members. Hence the cautious phrasing about the difference between collections and groups.

[2] We cannot invoke impeding factors, it should be noticed, in a free or undisciplined manner. There has to be reason to posit a contingent factor that gets in the way of the operation of the system. Suppose that the robot performed the part described but only on a random basis. In that case we would have little reason for recognizing it as an agent, unless there were evidence that a particular perturber was randomly getting in the way.

[3] The requirement of situational robustness 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.

[4] It may be more appropriate to speak of checking rather than controlling in the epistemic and indeed the evaluative context. In a given case your belief as to whether the evidence argues that p may be incorrect and your spontaneously formed belief that p correct rather than the other way around. But the capacity to form the evidential belief puts a check on

spontaneous belief-formation, making it more likely that you will end up satisfying epistemic ideals (Pettit 2007).

[5] Under the argument presented, of course, the domain of personal responsibility will be restricted to attitudes that are capable of being expressed in our common language. But that is not a particularly problematic constraint. If you are fit to be held responsible for forming or acting on attitudes engaging matters for which you and we have resources of expression—say, matters to do with the nature of the liquid in the glass before you, the position of that cup, and the desirability of drinking from it—then it will not matter that we have no words in which to express other presupposed attitudes: say, the sub-personal representation of the precise size of the glass, and its orientation from your body, that presumably plays a role in guiding your arm and the grasping motion of your fingers. We can hold you responsible for drinking the gin, even though we don't hold you responsible for the precise way in which you grasp and raise the glass.

[6] If you choose to report that you have a certain attitude—say, that you believe that p—then you cannot help but avow a distinct attitude: your belief that you believe that p. Although you can avoid avowal with any particular attitude, then, you cannot avoid avowing some attitudes.

[7] The fact of having avowed a belief does not give you a new reason for believing it; should the evidence change, you can excusably change your belief. But the fact of having promised to do something does give you a new reason

for desiring and acting accordingly: it puts your reputation at stake and constrains any changes of mind.

[8] Notice, as registered in (List and Pettit 2011, Ch 1), that this extra evidence may serve in the case of a conversable agent to override the evidence of behavioral failure that might lead us to doubt the agency of an impersonal system. If you fail to behave according to the attitudes of which we have independent evidence but admit the failure, perhaps even apologizing for it, then that will provide an assurance that you are an agent that would be hard to attain in the case of a non-personal agent.

[9] Questions naturally arise about what to say of groups where the evidence from interpersonal interaction is mixed—for example, where would-be spokespersons are in conflict—but the evidence of impersonal interaction is strong: for example, it suggests that some of the spokespersons are reliable, others not. I do not address such questions here but stick for simplicity to the purer cases.

[10] And even if that were not thought to be an objection, canons of parsimony would argue against invoking group agency to explain a pattern that is already explicable by the group-independent profiles of the group's members.

[11] You might drop the belief in the agential status of the individuals, as certain radical ontologies would do. But this would be a resort of radical despair. For a critique of the option see Chapter 3 of my book *The Common Mind* (Pettit 1993).

[12] In the beach case there is certainly a joint action on the part of the participants, sponsored by a joint intention that they form. This might materialize insofar as it is manifest to each that they all want to save the swimmer, that they can do so only together, that the salient way of doing so is to link arms and form a chain into the water, and that if anyone starts the chain then others will join up. Joint intention is certainly required for the formation of a group agent, as that is described here; it is implicit in the acquiescence of members in the identification of spokespersons and in the authorization of their words. But necessary as joint intention may be for the formation of a group agent, it is not sufficient on its own to ensure the presence of such an agent (Pettit and Schweikard 2006; List and Pettit 2011) There is a large literature on what occurs when people form and act on a joint intention; see for example (Tuomela 1995; Bratman 1999; Gilbert 2001). The account that fits best with my comments here is probably Bratman's.

[13] This fiction theory is important to Hobbes, since it undermines the idea that the commonwealth—for him, the supreme group agent—might be formed on the basis of a mixed, republican constitution that requires different spokespersons to agree in determining the voice of the state. He thought that such a constitution would create civil war, rejecting it on the grounds that it would create 'not one independent commonwealth, but three independent factions' (Hobbes 1994, Ch 29, s 6).

[14] A further consideration in support of this view is that a given collection of individuals might constitute one group agent, with its own commitments, in one context, and a different group agent, with different commitments, in another. The town council might have just the same members, for example, as the hospital board. To hold that either group was the same agent as its members would be to imply, absurdly, that the town council is the same agent as the hospital board.

[15] My thanks for the many helpful comments I received at a number of events where a version of this paper was presented: at an American Philosophical Association meeting in Chicago and at workshops in the University of Vienna, University College, Dublin and the University of Copenhagen. The paper draws heavily on my joint work with Christian List, of course, and I am deeply indebted to him. I am grateful to Rachael Briggs and Jennifer Lackey, who provided very helpful comments on earlier drafts.