

Christian List and Philip Pettit

Group Agency: The Possibility, Design and Status of Corporate Agents.

Oxford and New York: Oxford University Press 2011.

ix + 238 pages

\$45.00 (cloth ISBN 978-0-19-959156-5)

Wolfgang Prinz

Open Minds: The Social Making of Agency and Intentionality.

Cambridge, MA: MIT Press 2012.

xvi + 340 pages

\$40.00 (cloth ISBN 978-0-262-01703-9)

The notion of collective agency is at the centre of a growing literature in philosophy, of which much – though not all – draws on work in a number of other disciplines such as political science, law, economics, and developmental psychology. Perhaps unsurprisingly, philosophers with interests and influences as disparate as this list suggests do not always seem to be engaged in the same set of conversations. The two books under review, both excellent in their own way illustrate the situation well: despite their titles, there is almost no overlap in subject matter, orienting questions, or a shared background literature.

Despite the subtitle of his book, Prinz's main concern is not with collective agency, but with the ways in which the development of our capacity for agency at the individual level depends on our awareness of the agency of others. The following claim, which occurs early in the book, sets the stage for much of what follows: 'First we understand our own minds; and only then and thereby do we come to design our own minds accordingly' (33). We can distinguish between two separate strands in Prinz's thinking here. One is what we might call the 'outside-in' picture of self-understanding, on which our understanding of the mental states of others is developmentally and conceptually prior to our understanding of the minds of others. The second is what we might call the 'mind design' picture of agency, according to which 'what people *believe about the workings of their minds* can come to *affect the actual workings* of their minds'. Prinz suggests that as well as being interesting in its own right, this approach to agency has a philosophical pay-off: it can free us from the temptation to adopt a Cartesian view of the mind and it can help to dissolve philosophical problems about free-will.

It is impossible not to be impressed by Prinz's empirical contribution to our understanding of topics such as imitation, and the common-coding account of the representation of actions (and the role of mirror neurons within it) which he has pioneered certainly helps us to understand how a naturalistically plausible version of the 'outside-in' story about self-knowledge is possible. Still, some of his philosophical claims seem oversold. Like many psychologists, Prinz tends to overestimate the extent to which contemporary philosophy is in thrall to a Cartesian picture of the mind. Thus, he attacks advocates of a simulationist account of folk psychology as being committed to an 'inside-out' picture of our understanding of the minds of others, on which our understanding of the mental states of our own minds is prior to our understanding of the minds of others. This might be regarded as fair (though unoriginal) criticism of one very prominent simulationist, namely Alvin Goldman. However, other

simulationist authors such as Robert Gordon, Jane Heal, and Adam Morton have developed versions of the view which seem to be either compatible with an ‘outside in’ picture (as with Heal and Morton) or even explicitly to require it (in the case of Gordon).

Prinz’s claims about the impact of his view of agency on our understanding of long-standing problems about free-will also seems hard to sustain. He argues that we should see free-will as being a ‘societal institution’ (176) but goes on to say that since ‘social facts are no less real and efficacious than natural facts’ we can conclude that ‘free-will is not an illusion, but a real thing: those who have appropriated it actually have it.’ (176) This suggests that what is on offer will be a new and exciting view. However, the view Prinz actually presents does not seem to be one that would trouble an incompatibility determinist much: it is one on which our practice of attributing free-will has a social significance, and other people’s attribution of free-will to us plays a role in developing our capacity for agency. But while it is almost certainly true that our practice of treating one another as responsible agents plays an important role in our development of capacities for agency, it is not clear how this on its own can show that the beliefs which underpin this practice are true. (It might well be that much here depends on exactly how the notion of free-will is being understood: Prinz’s discussion seems to rely on a rather vague and intuitive understanding of the notion, and an analysis which drew on a more developed account, such as that presented in Philip Pettit’s 2001 book *A Theory of Freedom*, might allow Prinz’s claims here to be developed in a more convincing manner.)

A further observation seems worth making. We might want to distinguish between two ways in which agency and intentionality are socially made. On Prinz’s view there is no capacity for agency and self-knowledge without interaction with and observation of others. But at least as far as the theory is concerned, these others remain other individuals. Another kind of view, rather different from Prinz’s, would be one on which our capacity for individual agency depended on our capacity for doing things *together*. Prinz’s conception of the social is one on which even the possibility of a view of this sort cannot come into focus.

List and Pettit’s book *Group Agency: The Possibility, Design and Status of Corporate Agents* can be seen as an extensive discussion of the normative and metaphysical issues raised by what the authors call the ‘Discursive Dilemma.’ The Discursive Dilemma, or DD for short, is a generalization of an issue relating to collective decision-making first noted in the academic law literature by Kornhauser and Sager (‘Unpacking the Court’, *Yale Law Journal* 96, 82–117). Put at its simplest, the issue is that it seems impossible to guarantee that a group of individuals, trusted with making a decision that involves complex reasoning on the basis of more than one premise and involving a simple decision rule (such as majority decision-making) on all questions at issue can be guaranteed to come up with logically coherent responses.

The problem can be illustrated by the use of examples involving groups of 3 judges forming a judgment on a complex proposition whose truth depends on two simpler propositions. For example, suppose a regulatory framework decrees that a corporation is in breach of a regulation if and only if given action was both performed recklessly and caused substantial damage; and suppose a panel of three judges is empowered to make decisions as to whether substantial damage was caused, whether the corporation was reckless, and whether there was breach of the regulations occurred. As Kornhauser and Sager noticed, the following state of

affairs seems possible: Judge 1 rules that the corporation was reckless, but that substantial damage did not occur, and so the corporation is not in breach; Judge 2 rules that the corporation was not reckless, but that substantial damage did occur, so again the company is not in breach; and Judge 3 rules that substantial damage did occur, and that the corporation was reckless, so that the corporation is in breach. If the judges put each of the propositions before them to a majority vote, they will come to the following conclusions, each by a majority of 2-1: i) the corporation was reckless; ii) substantial damage did occur; iii) the corporation is not in breach of the regulations. But this is incoherent: given the regulatory framework, if i) and ii) are true, then iii) cannot be.

One might suppose that problems of this sort can only arise in artificially simple cases. However, an important result, proved by List and Pettit elsewhere, and reported on in the book under review, is that the problem can be generalized. There is no way of coming up with a rule or function which starts with a set of individual attitudes to a set of propositions on an agenda which are both complete – in other words, every individual has an attitude to every proposition under consideration – and consistent and which produces results which satisfy a set of constraints, ones which List and Pettit label ‘universal domain’ (any possible set of propositions can be taken as input); ‘anonymity’ (the results of the procedure do not depend on which individual has which attitude); ‘systematicity’ (the same rule holds for each proposition on the agenda, and the results for one proposition do not depend on the attitudes agents have to other propositions); and collective rationality (the output provides a result for every proposition on the agenda, and the overall results are consistent).

If synchronic rationality is a necessary condition for agency – a view which Pettit has endorsed in earlier work – then both Kornhauser and Sager’s point and List and Pettit’s more general result appears to undermine the notion of collective agency, since it suggests that putative collective agents are distressingly prone to lapses in synchronic rationality. However, List and Pettit’s key move, one which is of striking dialectical elegance, is to turn the apparent *modus ponens* on which this line of thought turns into a *modus tollens*. In effect, they argue that where we do have collectives whose attitudes are complete and consistent, they cannot be derived from the attitudes of the individuals who make up the collective by means of a well-behaved aggregation function. Seen in this way, the impossibility result has anti-reductionist implications, at least to the following extent: the most initially promising routes by which one might hope to reduce collective attitudes to those of the individuals making up the collective are ruled out. However, this need not be seen as making collective agency ‘spooky’ in a way which is incompatible with a naturalistic world-view. Even if straightforward recipes for reduction fail, collective agents might still have attitudes which supervene on the attitudes of individuals.

This line of argument falls short of showing that there are any collectives with attitudes that meet the conditions for agency. To reach that conclusion would require one to look at some actual instances of collective decision-making. So the subtitle of the book – which refers only to the ‘possibility’ of group agents and whose modal modesty has drawn puzzled comment from some reviewers – seems entirely appropriate. However, one kind of question which the book raises, and which would appear to repay further investigation, is whether the sorts of collectives to which we often ascribe intentionality (such as states, business corporations, scientific communities, and so on) do display rationality of this sort, and what sorts of mechanisms sustain

it in practice. List and Pettit do not pursue questions of this sort, at least in part because it would appear to require a set of social scientific tools which are very different from the ones which they are interested in deploying.

Instead, List and Pettit try to tease out some of the philosophical and practical implications of this and other formal results about judgment aggregation. While much of this is of great philosophical interest, the discussion often makes reference to mathematical results proved in the rapidly-growing literature on judgment aggregation but without giving full-scale proofs of them. While doing so would probably have made the book longer and less readable, it would also have made it more philosophically satisfying. (An appendix of mathematical results might have been a worthwhile compromise here.) List and Pettit's work on epistemological topics, such as Chapter 5's discussion of circumstances under which information pooling by individuals improves the ability of group agents to form collective beliefs which track the truth, struck me as having an interesting bearing on issues in the philosophy of science relating to ways in which the design of scientific institutions can enhance our collective capacities for gaining knowledge which have been explored by Philip Kitcher, Michael Strevens, and others. While it would have been interesting to see some of these issues explored by the authors, perhaps this is something for further, collective, investigation by our epistemic community.

Bill Wringe

Bilkent University