Methodological Individualism, the We-mode, and Team reasoning

PREPRINT

in a Symposium on Raimo Tuomela's *Social Ontology*,
forthcoming in *Protosociology* (2016).

Kirk Ludwig

1. Introduction

Raimo Tuomela is one of the pioneers of social action theory and has done as much as anyone over the last thirty years to advance the study of social action and collective intentionality. *Social Ontology: Collective Intentionality and Group Agents* (2013) presents the latest version of his theory and applications to a range of important social phenomena. The book covers so much ground, and so many important topics in detailed discussions, that it would impossible in a short space to do it even partial justice. In this brief note, I will concentrate on a single, though important, theme in the book, namely, the claim that we must give up methodological individualism in the social sciences and embrace instead irreducibly group notions. I wish to defend methodological individualism as up to the theoretical tasks of the social sciences while acknowledging what is distinctive about the social world and collective intentional action.[1]

Tuomela frames the question of the adequacy of methodological individualism in terms of a contrast between what he calls *the I-mode* and *the we-mode*. He argues that we-mode phenomena are not reducible to I-mode phenomena, and concludes that we must reject methodological individualism. I will argue that the irreducibility of the we-mode to the I-mode, given how the contrast is set up, does not entail the rejection of methodological individualism. In addition, I will argue that the three conditions that Tuomela places on genuine we-mode activities, the group reason, collectivity, and collective commitment conditions, if they are understood in a way that does not beg the question, can plausibly be satisfied by a reductive account. Finally, I will argue that the specific considerations advanced in the book do not give us reason to think that a reductive account cannot be adequate to the descriptive and explanatory requirements of a theory of social world.

2. Methodological Individualism

What is methodological individualism? The basic idea of methodological individualism in the social sciences is that we can understand social action and individual action in the social context ultimately in terms of concepts we deploy in understanding and explaining individual action. Tuomela divides methodological individualism into three components,

[1] Of course, what follows is not a full defense of methodological individualism but a partial defense of it against the main arguments in *Social Ontology*.

paraphrased here (Tuomela 2013, p. 10, all parenthetical citations to page numbers alone are to this book):

(1) We can understand individual action in terms of the individual's own attitudes and reasons (which may refer to other agents and their attitudes and reasons, and to reducible group attitudes and reasons) together with physical and nonsocial facts.
(2) We can explain individual action on the basis of the individual's own attitudes and reasons (which may refer to other agents and their attitudes and reasons, and to reducible group attitudes and reasons) together with physical and nonsocial facts.
(3) Social ontology consists solely of the actions and activities of individuals and their relations and interactions, and reference to groups and group properties is reducible to the bases in (1) and (2).

Methodological individualism does not, as I understand it, deny the existence of *groups*. Groups are simply collections of objects, like the white pawns in a chess set, or the stars in Orion's belt, or the people sitting on benches in Central Park in New York. We can refer to groups, in particular, groups of agents, and their properties and relations (so long as they aren't irreducibly social), compatibly with methodological individualism. But what we don't allow are notions in our theoretical and explanatory practices that cannot be understood in terms of notions that apply in the first instance to individuals and groups neutrally described with respect to whether they form a group oriented toward joint intentional action.

3. The Irreducibility Thesis

How should we understand the thesis of irreducibility? Tuomela explains the thesis in terms of the irreducibility of we-mode concepts to I-mode concepts.

> The social world can be adequately understood and rationally explained only with the help of we-mode concepts expressing full-blown collective intentionality and sociality in addition to I-mode concepts. We-mode thinking and reasoning is not conceptually reducible to I-mode reasoning; i.e., it is not definable by, or functionally constructible from, I-mode notions, nor does it seem fully explaining explainable in terms of the I-mode framework. (p. 15)

What is the contrast between I-mode concepts and we-mode concepts? The we-mode is introduced first, though in part in relation to the I-mode:

> the intuitive idea [is] that the acting agent in central group contexts is the group viewed as an intentional agent, whose members are engaged in *we-reasoning from the group's point of view* (e.g., "We will do X" and "What does our doing X require me to do?"). Conceptually, the individual agent is not the primary agent (as in the I-mode approach) but rather a representative acting for the group—although ontologically, in the causal realm, individuals are the only initiating "causal motors". In general, the conceptual and justificatory direction for theorizing and conceptual

> construction in the we-mode account is “top down” rather than “bottom up,” as in the I-mode account. (loc. cit.)

There is an initial puzzle. In this passage, the we-mode is defined as distinct from, and in contrast to, the I-mode. This theme is developed more fully in later discussion (see the definitions on pp. 70; 147-9).[2] Irreducibility to I-mode notions appears to be built in definitionally. If so, and if by I-mode notions we mean generally those at play in our understanding of individual action, then the question whether we-mode concepts are reducible to I-mode concepts would not be substantive. This would shift the substantive question from whether the we-mode is reducible to whether it is needed or applies to the social, and whether we operate with we-mode concepts.

However, I think we can understand how the idea of the we-mode can be introduced so as to secure that we operate with we-mode concepts, while leaving open the question whether we can understand it in individualistic notions. Tuomela identifies the three main concepts distinctive of we-mode activity as *authoritative group reason*, *collectivity*, and *collective commitment*. A group acts in the we-mode only if it acts for a *group reason*, and so meets the *collectivity* and *collective commitment* conditions. Provided that these can be understood in a way that doesn’t beg the question, we can proceed to ask whether an account of group activities that meets these conditions can be given within the constraints of methodological individualism.

This will, however, require us to distinguish between the claim that the we-mode account can be reduced to the I-mode account (or that we-mode concepts can be reduced to I-mode concepts) and the claim that methodological individualism is true. For insofar as I-mode notions are characterized definitionally by contrast with notions that can be used to meet the three conditions, even when characterized so as not to beg the question against methodological individualism, the irreducibility of we-mode to I-mode notions would leave open whether we-mode notions can be reduced to notions that are at work in our understanding of individual action and agency.

What is a *group reason*? A group reason is a reason promoting the group’s interests (p. 38). An *authoritative* group reason is one that the members of the group “are normatively bound to comply” with (p. 115). As long as we leave open that we may give a reductive reading of ‘group interests’ (and allow a sense of ‘normatively bound’ that is at least *prima facie* neutral), we beg no questions against reducibility in requiring this of we-mode activity. For example, we might understand group interests in terms of a goal that the group members have collectively agreed upon, at least tacitly (where we leave open how to analyze ‘collectively agreed’). What is the *collectivity condition*? This requires constitutively that a group’s intention “is satisfied” for one member of the group qua member iff it is satisfied for every member qua member, and a group’s intention is satisfied

[2] Given the definition of ‘we mode intention’ (WMI) on page 68, it is clear that the I-mode notions on page 70 are defined so as to exclude the we-mode, that is to say, to act or intend in the I-mode definitionally excludes acting or intending in the we-mode. Thus, there appears to be no hope of reducing acting in the we-mode to acting in the I-mode.

iff it is satisfied for every member qua member (pp. 40-1). Examples make clear what the idea is. If we act in the we-mode as a group, we each intend to do our parts in the group doing something. Meeting the collectivity condition comes to it being required for each of us to satisfy his or her intention that each other member of the group satisfies his or her intention, and similarly for the group to satisfy its intention, each member has to satisfy his or her intention. So long as we leave open whether participatory intentions and shared intentions are understandable in terms of notions at play in the understanding of individual action, this condition does not beg the question against reducibility. Finally, what is *collective commitment*? Collective commitment is tightly connected with joint intention (pp. 43-5; 82). If we jointly intend to lift the piano, then we are collectively committed to doing so. Being collectively committed to something entails group-based obligations toward one another to do it. Since the guiding idea is joint intention (and joint intention is sufficient for collective commitment), as long as we have a way of locating what we have in mind by joint intention without stipulating that it is irreducible, requiring collective commitment as a condition on we-mode activity will not beg the question. We can locate the relevant sort of intention by reference to the collective reading of ordinary plural attributions of intention. For example, it is the sort of thing we have in mind in talking about our intending to meet to have lunch together.

4. Reductive Satisfaction of the We-mode Criteria and Construction of Thick Group-centric We-mode Concepts

Prima facie, these three criteria for the we-mode can be satisfied by accounts of joint intention and we-intentions that are reductive in character. For example, Bratman's well-known account of shared intention (in cases of modest sociality) seems to satisfy the three criteria (Bratman 2014, 1992, 1999). First, since our grip on collective commitment goes by way of our prior understanding of joint intention, as long as there is no independent objection to Bratman's analysis (or any other), we cannot cite failure to secure collective commitment against the account. Second, on Bratman's account we share an intention to *J* *only if* each of us intends that we *J* by way of meshing subplans associated with our respective intention that we *J*. Thus the account clearly satisfies the collectivity condition, for the participatory intentions of each member of the group cannot be satisfied without those of other members of the group being satisfied, and this is also a condition on the joint intention being satisfied. And finally since jointly intending to *J* involves at least tacitly agreeing that we will have as our goal *J*-ing (we are functionally so to speak all on the same page), we secure also that we act for a group reason in the sense of acting for group interests.[3] So far as I can see, the same thing goes for my own account of our jointly

[3] Tuomela discusses Bratman's account explicitly, but mischaracterizes it: "In his [Bratman's] account the basic cooperative intention has the form "I-intend that we perform joint action X," where X instantiates an individually shared goal G (recall (CIM)). His account does not make use of the constitutive feature of collective acceptance nor the other central elements of the we-mode framework. Accordingly, he deals with a weaker notion of cooperation than my notion of we-mode cooperation. His account seems to deal rather with "pro-group I-mode" cooperation and would seem to be a special case of my (CIM) account" (pp. 158-9). But, first, the characterization of the basic cooperative intention is incorrect. For Bratman, for one to we-intend, one must intend (at least) that the group J by way of the intentions of each that the group J by way of meshing subplans of their intentions that they J (in his (2014) account this is secured in modest sociality by the mutual

intending to *J*, on which each one of us intends to bring it about in accordance with a plan he has that there is a plan (one and the same plan) in accordance with which each of us contributes to bringing it about that we *J* (Ludwig 2007).

What about the requirement that collective commitment entails group-based obligations toward one another to do what the group is committed to doing? If one thinks of these as *sui generis* obligations that attach to joint intention *per se* (as in (Gilbert 2006, ch. 7)), then that would introduce a conceptually irreducible element. But this is itself a contested issue, and so one cannot enter it as an objection, without further argument, to accounts like Bratman's and mine, which see what obligations there are to the other participants in joint action as derived from external requirements, such as those of morality, applied to the conditions of joint intention. This will also secure a sense in which group reasons are authoritative, though here we may also cite group reasons as authoritative simply because they come into existence as a result of all members of the group adopting a settled commitment to act as a group: that is, they share an intention to do something (all we-intend they do it), and it is characteristic of such commitments that one acts forthwith in accordance with them without revisiting the practical reasons for them absent special reasons.

With this as background, we could, as Tuomela does (p. 78), characterize thicker notions of a we-intention, conceived of as a slice in a joint intention. For illustration, I will characterize three notions of we-intentions of increasing "thickness." To distinguish these notions from the basic idea of a participatory intention, I will put '**we-intention**' in boldface. First, we can say: x **we-intends** that we J iff x is one of us and we all we-intend that we J. Then having a **we-intention** would entail the existence of the group and that the group has a joint intention. Second, we can add features on top of this to strengthen the social glue, requiring (i) x to have a true belief that the group g has a joint intention to J and that this be x's main reason for having a participatory intention, (ii) x to have a true belief that preconditions for success in carrying out the joint intention are met, (iii) x to truly believe there is a mutual belief among members of g that preconditions for success in carrying out the joint intention are met, and that (iv) x have his participatory intention and belief that the group has the joint intention in part because of (ii) and (iii). Finally, we could, if we liked, add that members of the group have a strong enough commitment to

responsiveness condition). That is a much more powerful condition than simply that I intend that we perform joint action X. This is what secures that the collectivity condition is met. Furthermore, one cannot object here that the collective acceptance condition is not met unless one can *otherwise* object to the account. For if the account of shared intention that rests on this account of the content of a participatory intention at least provides sufficient conditions for joint intention, since joint intention entails collective commitment, he has thereby secured that the collective commitment condition is met. And for similar reasons one cannot object that Bratman's account is not adequate to the idea of authoritative group reasons without further argument. Tuomela raises a question about whether Bratman's account (and the same question can raised about my account) would be circular if we instantiate 'J' to an essentially intentional collective action type, for example, *playing chess*, or *shaking hands*. But if these types of actions can be analyzed into a component that is neutral with respect to whether it is being instantiated intentionally and the requirement that it be instantiated jointly intentionally, then there is no harm in embedding the concepts in the content of joint intentions, for then the participants know that to execute the intention to do something intentionally together, they need merely we-intend that they instantiate the neutral component. See (Ludwig 2014, n. 11).

acting as members of the group (acting on their we-intentions) to override private interests. We could then say a group has a **joint intention** to J when every member of it has **we-intention** to J (with three corresponding senses of increasing thickness). In this way we build up strong group centric notions of participatory intention and joint intention. None of this, however, introduces new *sui generis* notions as it enriches the content of '**we-intention**' and '**joint intention**'. Then we can characterize the notion of x's having an intention in the *we-mode* with collective content P (p. 68) as a member of a group g as a matter of x's being prepared to function as a member of g (in the sense of doing his part in the group's doing something it **jointly intends** to do in the sense(s) above), x's intention presupposing the agents in g collectively accept P as the content of their **joint intention** for satisfying the interests of g (where both these conditions are secured already by the group **jointly intending** P), x's intending to participate in the satisfaction of g's **joint intention** (by we-intending), and x's presupposing the we-mode criteria are satisfied for the participants (secured by the thick notion of **joint intention**). If the basic notions we start out with can be understood in terms of concepts in play in our understanding of individual action, then all these thicker notions can be introduced on their basis in turn, to secure a foundation for understanding the we-mode that is consistent with methodological individualism.

These definitions are not those that Tuomela gives, of course, since those appeal ineliminably to the notion of joint intentional action in the analysans (pp. 76-7). The suggestion is, however, that *prima facie* we can construct analogs that are compatible with methodological individualism as long as we can give a reductive account of the more austere notions of a we-intention and joint intention that satisfy the three criteria on the we-mode. On a neutral understanding of those criteria, it appears that we can do this.

5. Objections to Reducibility

Despite the *prima facie* compatibility of these reductive accounts with the three criteria, there may be further reasons to reject the reduction. With that in mind, I want to look at the arguments offered in *Social Ontology* for the irreducibility of we-mode concepts to concepts already in play in our understanding of individual action. The main locations of arguments for irreducibility are chapter 3, section 5, and chapter 7.[4] I will restrict attention to these arguments, and give special attention to the arguments in chapter 7 based on considerations involving we-mode reasoning in the context of collective action dilemmas.

As mentioned above, I am not taking the question whether we-mode notions are reducible to be equivalent to the question whether we-mode states (or notions) are reducible to I-mode states (or notions). I rather take the question to be whether we-mode notions (identified neutrally in the way we have above) are understandable in terms of notions already at play in our understanding of individual action. If my account, or Bratman's account, is correct, then this is what is achieved. But it is not clear that either of these accounts show that we-mode states are reducible to I-mode states because I-mode states

[4] I think what I say below shows how the response would go to arguments in chapter 6.

appear to be defined explicitly so as to contrast with we-mode states. So I will be asking whether accounts along the lines of Bratman's or mine succeed in meeting the neutral desiderata on we-mode activities. If a reductive account is possible, then it will not disturb any of the claims made about the importance of we-mode groups or their functioning. It may still be maintained, so far as that goes, that there is a contrast between we-mode groups and pro-group I-mode groups, and that we-mode groups lead to more efficient forms of group action, solve or dissolve group action dilemmas, and so on.

In developing the "arguments for the irreducibility of we-mode concepts and states" (p. 91) in chapter 3, section 5, Tuomela begins with the thesis that is be established.

> (IRRED) Propositions containing predicates (concepts) that express we-mode collective attitudes and actions or other related we-mode collective intentionality properties and activities (e.g., cooperation, collective commitment) in general are neither conceptually nor explanatorily or ontologically reducible to propositions containing predicates (concepts) expressing private (i.e., I-mode) intentions and actions and what can be conceptually constructed out of I-mode resources. (loc. cit.)

For present purposes, I will understand "predicates (concepts) expressing private (i.e., I-mode) intentions and actions and what can be conceptually constructed out of I-mode resources" to mean "predicates (concepts) expressing concepts already in use in our understanding of individual action and intention." This will allow us to address directly the question whether the arguments suffice to show that we must give up methodological individualism without having to be concerned with issues that might arise from the way the I-mode appears to be contrasted definitionally with the we-mode.

There appear to be four main arguments sketched this section. The first is a précis for the arguments developed in chapter 7, so I will put it last in the list and address it in the context of chapter 7. The first three go as follows.

(a) The argument from collective commitment.

> ... in the purely private I-mode case a person is committed to herself to satisfying her intention, in the pro-group I-mode case she is committed to herself to participating in the satisfaction of the group's shared (I-mode) intention, and in the we-mode she is committed to the group to participating in the satisfaction of its intention as a group member. These differences in general entail dispositional differences concerning the behaviors to which the participants are committed, and again here the we-mode is seen to differ from the I-mode in a way suggesting that the we-mode is irreducible to the I-mode. (p. 93)

Reply: This argument appeals to a definitional distinction between shared I-mode intentions and genuine shared group intentions (of the sort expressed by 'We intend to meet to have lunch together') and so if doesn't already beg the question, it leaves it open that joint intention can be understood reductively along the lines of the accounts above which appear to meet the three criteria on we-mode activities.

(b) The argument from group reasons.

> … a group member may have an irreducible *group reason* to help other group members in a task, but he need not have a relevant, corresponding private reason—viz. one with the same content. We-mode group reasons accordingly need not supervene on I-mode reasons in this direct sense—in synchronic cases group reasons may change without a corresponding change in I-mode reasons—even in cases where the latter reasons are contingently group-based. (loc. cit.)

Reply: *Prima facie*, this begs the question by saying that a group member may have an *irreducible* group reason. That should be the conclusion of the argument, not its first premise. If we drop 'irreducible', however, and just think of a group reason as a reason promoting the group's interests, and think of the group's interests as some function of the utilities of members or some goal they have at least tacitly agreed to pursue in the sense of having formed a joint intention with that goal, then the reductive accounts above are adequate, even if the members of the group don't have individual goals, separate from their commitment to being members of the group, that aim at the same thing (that is, they wouldn't have that goal except insofar as they were committed to acting *as a member of the group*—where this is a matter of sharing in the joint intention).

(c) The argument from explanation.

> Furthermore, there is the simple general point discussed in chapter 1 that the social world cannot be understood and accordingly cannot be fully explained solely in terms of I-mode notions without change of topic, so to speak. (loc. cit.)

Reply: This thesis was stated in chapter 1 but not argued for there. But we can grant, as above, that I-mode notions, understood as definitionally contrasted with we-mode notions, are not adequate to the explanatory task, but still maintain that we-mode notions, which are needed, are constructible out of notions already at play in our understanding of individual action, in the way sketched above based on my or Bratman's accounts. *Prima facie* those accounts meet the three criteria for the we-mode, at least if they are stated in a way that does not immediately beg the question.

(d) The argument from Team Reasoning

I turn now to the argument from we-reasoning in game-theoretic contexts. The main idea of this argument is expressed in the following passage:

> In this chapter the we-mode approach to group-reasoning (or, which here amounts functionally to the same, we-reasoning) will be connected to the recent work by the economist Michael Bacharach on "team reasoning". My philosophical theory provides a conceptual framework that augments Bacharach's theory. Indeed, his mathematical results can be taken to support my claim about the irreducibility of the we-mode to the I-mode. In particular, it will be shown that group reasoning

> yields different results than pro-group I-mode theorizing, and in many cases it will be able to create more institutional order in the social world than I-mode theorizing. (p. 179)

As an initial remark, in line with what was suggested above, we could grant this, that is, that we-mode reasoning is distinct from pro-group I-mode reasoning (since pro-group I-mode reasoning is introduced explicitly to contrast with we-mode reasoning) but still allow that we-mode reasoning can be understood in terms of notions already at play in our understanding of individual action, at least so far as the criteria for identifying we-mode activities goes, as argued above. But it may still be the case that on reflection no reductive account of the we-mode would be adequate to intuitive problem that collective action dilemmas present and the empirical data which suggests that people engage in a form of reasoning that easily bypasses what look, from the standpoint of classical game-theoretic assumptions, to be collective action dilemmas. Therefore, we must consider directly the nature of the problems raised by collective action dilemmas.

The central question can be raised in connection with the Hi-Lo game and the Prisoner's Dilemma (PD).

In the Hi-Lo game, there are two ways to coordinate so each agent receives equal benefit, and coordination is better than not coordinating, but one way of coordinating yields a higher payoff for each than the other. For example, suppose that we agreed to meet for lunch but forgot to specify where. We always eat at one of two restaurants, one of which will be closer for both of us today, and this is common knowledge. Assuming they are otherwise equally good, and that closer is better, we face a Hi-Lo game. The payoffs might be represented in the following diagram. The first number at each intersection of choices is the payoff for Agent 1 and the second the payoff for Agent 2 for that combination.

Hi-Lo		*Agent 2*	
		Hi	Lo
Agent 1	Hi	3,3	0,0
	Lo	0,0	1,1

Classical game theory tell us that rational coordination involves finding a solution in which no one is better off by making a unilateral change.[5] This is a Nash equilibrium. But in the Hi-Lo game there are two Nash equilibriums: HiHi but also LoLo. So classical game theory doesn't single out one of the two as uniquely best.

In the Prisoner's Dilemma, two prisoners, held incommunicado, may cooperate with each other (refuse to confess), or defect (confess and implicate the other). If both cooperate, that is, refuse to confess, then they are convicted of a minor crime and spend a relatively small amount of time in jail. If one defects and the other does not, then the one who defects

[5] I will not enter here into whether this is correct, even for strategic reasoners. See (Risse 2000; Pearce 1984; Bernheim 1984) for some discussion.

is given a fine and released, but the other receives a long jail term. If both defect, they both receive medium length jail terms. If we assume that nothing matters to the participants except the external penalties imposed conditional on each choice, the relative payoffs can be represented in the following diagram. 'C' and 'D' represent 'cooperate' and 'defect', respectively.

Prisoner's Dilemma

		Agent 2	
		C	D
Agent 1	C	3,3	1,4
	D	4,1	2,2

In this case, there is a single Nash equilibrium, namely, DD. We may also appeal to a dominance principle. It looks as if the safest thing to do is to defect, since that is better for each no matter what the other chooses—the best reply (for each) in each case is to defect.

The puzzle is both empirical and conceptual. On the one hand, experimentally, and in everyday life, it appears that agents facing coordination problems which appear to have, or are designed to have, the structure of a Hi-Lo game frequently choose HiHi, and judge that to be rational. In the case of pairs of agents who face what appears to be or is designed to be a PD, while many choose DD, many also choose CC, and regard that as the obvious choice. However, classical game theory suggests that *rational* agents will defect in PD and provides no reason to favor HiHi over LoLo in the Hi-Lo game. On the other hand, conceptually, it seems intuitively as if it should not be hard or irrational for agents to choose HiHi, or even CC, but HiHi is not singled out by classical game theory as the uniquely rational choice profile, and CC in PD is not regarded as rational at all. The question is how to (and whether we can) understand the situations so as to make those choices the right ones, at least in many normal circumstances, from the point of view of the agents.

The two main strategies for resolving the puzzles are preference transformation and agency transformation. The first urges us to reconsider whether in the sorts of cases that we think have this structure the payoffs reflect fully agent preferences. For example, if each prisoner, in the situation described in the preamble to the PD matrix, places value on cooperating per se, and disvalue on defecting, then the values for CC might be (5,5), for CD (3,2), and for DC (2,3), and for DD (0,0), in which case the participants are not faced with a PD and CC becomes the obvious choice. The second urges us to think of the participants as reasoning as a team or group and not as strategic individuals. Team or group reasoners choose the option that maximizes, or maybe optimizes (see note 8), group payoff—intuitively, they play for the team, not for themselves. In PD that is CC, and in Hi-Lo that is HiHi. Tuomela focuses on the second approach.[6]

[6] One would not expect that being a team reasoner, as opposed to not, would leave one's individual preferences unaffected. Team reasoners are still individual agents, and they still have their own preferences. So the hypothesis that members of a group are team reasoners as opposed to individual strategic reasoners should leads us to reassess their preferences in putative cases of PD and Hi-Lo. If team reasoners have a commitment to maximizing or optimizing group utility, then that corresponds to a preference ranking in which maximizing or optimizing group utility *per se* is given a high value in each agent's preference ranking.

Tuomela's argument has two stages. In the first, he argues that we-mode we-reasoning, as opposed even to pro-group I-mode we-reasoning, yields the result that the rational choice in PD and Hi-Lo are CC and HiHi respectively. In the second, he argues that this makes it plausible that we can align we-mode we-reasoning with Bacharach's notion of team reasoning (Bacharach 1999, 2006), and align pro-group I-mode we-reasoning with Bacharach's notion of reasoning as a team benefactor—that is, someone whose individual preferences align with what's best for the group but who still reasons individualistically (as in the classical conception of Hi-Lo). He then argues that we can rely on Bacharach's theorem that team reasoning reduces the number of equilibria in games by eliminating Pareto-suboptimal equilibria over reasoning as a team benefactor. An equilibrium, as noted above, is a set of choices in which no one has a reason to change if others do not, and a Pareto-optimal equilibrium is one in which the benefit to one can't be increased without decreasing that of another.[7]

The most important stage for our purposes is the first, since if there is a place where we will find a difficulty for a reductive account of the we-mode, it will be in thinking about what has to be true of we-mode we-reasoning for it to solve or dissolve the collective action dilemmas.

In pro-group I-mode we-reasoning one asks "What should I do as a private person acting in part for the group?" and in we-mode we-reasoning one asks "What should our group do?" (181). (Note that these are characterized in a way that seems to require that they are distinct.) Let us first consider how the contrast between pro-group I-mode we-reasoning and we-mode we-reasoning is supposed to be reflected in how groups approach situations characterized Hi-Lo (see note 9 for PD). In the case of Hi-Lo, Tuomela says that in we-mode we-reasoning, the members of the group reason as follows (p. 187):

1. We intend to maximize group utility.[8]
2. Outcome HiHi uniquely maximizes group utility.

[7] Hi-Lo makes it clear how this works. Team reasoning ranks choice profiles by Pareto efficiency. One profile is more efficient than another if one agent's payoff can be increased without decreasing that of any other. In Hi-Lo, HiHi is the only Pareto optimal choice. Thus, in team reasoning, one choses Hi as one's part in HiHi. The team benefactor, whose ranking mirrors the group ranking, still reasons as an individual, and so, armed with only the resources of classical game theory, is at an impasse. Additional principles have to be added to resolve PD because there are three Pareto optimal choice profiles. Bacharach assumes the players in team reasoning prefer (strongly enough) both cooperating to one free riding (Bacharach 2006, pp. 168-9).

[8] It is not clear that, in these cases, we should think of we-mode we-reasoning as aiming at *maximizing* group utility rather than, say, aiming at the Pareto optimal solution if there is a unique one, or a solution which maximizes group utility relative to the requirement of certain minima for all participants, or aims for the greatest least inequitable distribution of goods. Maximization of group utility per se would potentially require participants to find it rational to make any personal sacrifice (giving his or her organs to save five other members of the group, e.g.) as his or her part in maximizing group utility. But if we, for example, say that the group is aiming at a Pareto optimal solution, and if there is more than one, then the one with the highest least inequitable distribution, it is clear that even in pro-group I-mode we-reasoning there is only one unique choice for Hi-Lo (where there is one Pareto-optimal solution) and for PD (where, while there are three Pareto optimal solutions (CC, DC, CD), one yields the highest least inequitable distribution (CC)).

Therefore,

3. I will perform my component in HiHi, that is, Hi.

This is contrasted with pro-group I-mode we-reasoning, which is said to lead to an impasse (loc. cit.):

1. You and I intend to maximize group utility.
2. If you choose Hi, my choosing Hi maximizes group utility.
3. If you choose Lo, my choosing Lo maximizes group utility.

Therefore,

4. I will perform what?

There is an initial puzzle here. If premise 2 is true in the first argument, then there is a unique outcome that maximizes group utility that is determined by the payoff structure in the matrix. In that case, premise 3 of the second argument is false, and premise 2 seems to have a false presupposition. Instead, we should replace 2 and 3 in the second argument with 2 in the first, and then it seems that the problem is solved, even if the participants can't communicate with one another. For if they know that both intend to maximize group utility, and there is a unique way to do it, they know what each needs to do, and know that the other knows that and so on. We could insist that 2 and 3 in the second argument are correct, but then 2 in the first would be false, and we-mode we-reasoning would be faced with the same problem. We might try to avoid this by saying that 'maximizing group utility' means different things in the first argument and in the second, but then we would not be comparing reasoning about the same goal. The problem here seems to me to be difficult to avoid. If you and I *intend* to promote some interest of the group, and we both know that, and there is a unique, or unique best, way to do it, then that determines our choice. If there is not, then we-mode we-reasoning is not going to help.[9]

The key to why C in the PD and Hi in Hi-Lo are singled out as optimally rational for each agent is that they each start out with a premise that states that they are both already committed to the same goal for the group. Each has to make only relatively weak assumptions about their capacity to reason in order for each to see that the other will conclude that there is only one pair of choices that will achieve the goal that they both know they intend. Given that, they can then each choose their part in what they do without further concern. The problem arises in cases in which the participants do not know (or

[9] What about PD? Let's take pro-group I-mode we-reasoning first. There doesn't seem to be any question about which combination of choices maximizes group utility—in our example. CC yields 6, DC and CD yield 5, and DD yields 4. Thus:

1. You and I intend to maximize group utility.
2. Choosing CC maximizes group utility.

Therefore,

3. I will perform my component in HiHi, that is, Hi.

A fortiori, we-mode we-reasoning gets the right result, but without a contrast.

presuppose) at the outset that they both intend to work toward the same goal. That is what makes the practical situation of we-reasoners starting with the assumption that they all intend to promote a group goal different from that of strategic agents reasoning under uncertainty. Nothing more, however, appears to be needed. This is present in both the case of pro-group I-mode reasoning and in the case of we-mode we-reasoning as originally presented, and so we do not appear to have established a functional difference between pro-group I-mode we-reasoning and we-mode we-reasoning in the context of collective action dilemmas.

It might be said that we should rethink how the reasoning goes in the pro-group I-mode case. It is not that each starts with the assumption that they both intend to maximize group utility. It is that each starts with the assumption that *he* intends to maximize group utility. Not knowing what the other is inclined to do, he will not be sure that he can maximize group utility, and perhaps will think that he should at least come as close as possible, and so choose Hi if the other does, and Lo if the other does. Then we would have a situation that would look more like an impasse.[10]

Contrast this with we-mode we-reasoning. Doesn't we-mode we-reasoning presuppose that other members of the group collectively accept the group goal (pp. 68, 78)? And in that case, it would not be possible (by definition) for we-mode we-reasoners to be in doubt about the others having appropriate intentions. That's fine, so far as it goes. We can define we-mode we-reasoning so that it builds in that the reasoners are not in doubt about the commitments of the others to pursuing the goal. But then if we want a fair comparison with pro-group I-mode we-reasoning, we need to allow in the comparison case that the pro-group I-mode we-reasoners are not in doubt either. If we were to say that that would transform pro-group I-mode we-reasoning into we-mode we-reasoning, then, since adding that they are not doubt about the other's commitment to the maximizing group utility is not to add any *sui generis* notion to the former, there would be no argument for strict irreducibility to notions at use in our understanding of individual agency.

If members of a group facing a putative PD or Hi-Lo game (i) all have a commitment to maximizing group utility, and (ii) each justifiably believes (i), and that (iii) the others believe (i), and (iv) that the others believe (iii), then they can easily rationally coordinate on CC and HiHi.[11] To the extent to which we-mode we-reasoning requires this, it will

[10] This is how Bacharach is thinking of team benefactors, who still reason like strategic individuals.

[11] I say 'justifiably' because if their beliefs are irrational or unjustified, even though they are able to reach the right result, we have not shown that they can do so by acting rationally. We-mode we-reasoning requires that members of the group believe the others will do their parts, but doesn't *prima facie* require their beliefs to be justified. We-mode we-reasoning could result in the right solution without the members of the group being rational in doing so, if their admittedly true beliefs that the others are participating and all the conditions for success are in place are not rational. So what is needed is not just we-mode we-reasoning (and the same goes for pro-group I-mode we-reasoning) but we-mode we-reasoning in which the participants are justified in engaging in that form of we-reasoning. To put it another way: if they are not justified in believing the premises of the arguments above, they are not justified in the practical conclusion they reach on its basis. So for we-mode we-reasoning (or pro-group I-mode reasoning) to be a rational approach to solving a collective action dilemma, the group members must have reason to think we-reasoning will lead to success. What reason do they have for this if all the information they have is that given by the standard game-theoretic

secure the conditions necessary for them to rationally coordinate on CC and HiHi. But this is compatible with we-mode we-reasoning being explained reductively because the condition secured is stated in terms already in play in our understanding of individual agency.

To the extent to which we have found no difference between we-mode we-reasoning and pro-group I-mode we-reasoning in their ability to resolve collective action dilemmas, the alignment of benefactor reasoning with pro-group I-mode we-reasoning and team reasoning with we-mode we-reasoning is undermined, and Bacharach's theorem does not support the irreducibility of we-mode to pro-group I-mode we-reasoning. But even if we stipulate that definitionally pro-group I-mode we-reasoning entails that each member of the group fails to satisfy conditions (i)-(iv) above, since these conditions don't involve any *sui generis* or irreducible group notions, we can construct a form of we-reasoning from the available materials that resolves collective action dilemmas without appeal to irreducible group notions.

6. Conclusion

The three criteria on the we-mode are that there be a group reason for acting, that the group meet a collectivity condition, and that the group have a collective commitment. If it is a substantive question whether the we-mode is irreducible, which is presupposed by offering arguments for that conclusion, then we need to understand these in a way that is at least *prima facie* neutral on the question. I have argued that when we construe these criteria so as to be neutral, a good case can be made that accounts like Bratman's or mine can satisfy them, and that out of these basic notions richer group centric notions analogous to Tuomela's can be constructed which still respect methodological individualism. I have further argued the arguments from collective commitment and group reasons in chapter 5 are not successful because they are used to show that the we-mode is not reducible to the I-mode, but this is secured definitionally in a way that leaves open that the we-mode is reducible to concepts used in the understanding of individual action and intention. In addition, I argued that the argument from team reasoning does not appear to succeed because it does not establish that pro-group I-mode reasoning is not adequate to the task of dissolving collective action dilemmas, since all that is required is that the members of the group have good enough reason to think all are reasoning in that mode. If pro-group I-mode we-reasoning were defined so that it were incompatible with having good enough reason to think that all group members are so reasoning, that would not rescue the argument because we can still specify conditions that are compatible with methodological individualism which secure the right results. I conclude, therefore, that methodological

payoffs? While we-reasoning leads in Hi-Lo to what we think of as the best choice, it looks reasonable only if one has reason to think the other is we-reasoning as well. In my example involve the two restaurants, this secured by a prior agreement to eat lunch together. In the case of situations involving no prior agreements or strangers, we face a problem. The empirical problem might be solved by saying that people we-reason by default. The conceptual problem is solved only if we add that they reasonably expect others to we-reason. Perhaps this is reasonable, but if so, it is so because we have a broadly inductive assurance that this is a default reasoning mode, which can be undermined in particular circumstances—which may explain why in PD we see both C and D chosen by many people.

individualism has not so far been shown to fail to suffice for “giving an adequate description and explanation of social facts and structures, which is the main task of social science” (p. 10).

References

Bacharach, Michael. 1999. Interactive Team Reasoning: A Contribution to the Theory of Cooperation. *Research in Economics* 53 (2):117-47.

———. 2006. *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton: Princeton University Press.

Bernheim, B. Douglas. 1984. Rationalizable Strategic Behavior. *Econometrica* 52 (4):1007-1028.

Bratman, Michael. 1992. Shared Cooperative Activity. *The Philosophical Review* 101 (2):327-341.

———. 1999. I Intend that We J. In *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge: Cambridge University Press.

———. 2014. *Shared Agency: A Planning Theory of Acting Together*. Oxford: Oxford University Press.

Gilbert, Margaret. 2006. *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*. Oxford: Clarendon Press.

Ludwig, Kirk. 2007. Collective Intentional Behavior from the Standpoint of Semantics. *Nous* 41 (3):355-393.

———. 2014. Proxy Agency in Collective Action. *Nous* 48 (1):75-105.

Pearce, David G. 1984. Rationalizable Strategic Behavior and the Problem of Perfection. *Econometrica* 52 (4):1029-1050.

Risse, Mathias. 2000. What is Rational about Nash Equilibria? *Synthese* 124 (3):361-384.

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