

Cedric Paternotte

Comment on Raimo Tuomela. Joint Action: How Rational? How Irreducible?

Abstract: In his ‘Cooperation as joint action’, Tuomela presents a we-mode account of cooperation, which he argues has several advantages over an individual account. This commentary examines to what extent this is true. In particular, I assess three related characteristics of we-mode joint action: its possible rationality, its greater efficiency, and its alleged irreducibility to purely individual properties, which are recurring points of the article.

The literature on joint action started blossoming about twenty-five years ago. For decades, philosophers have tried to identify the elements that are integral to human cooperation, and as a result a host of definitions have been crafted. What properties epitomize the fact that individuals act together, in unison, in agreement, or as one? For a while, it seemed that each theorist would have her own diagnosis, idiosyncratic and irreducible to that of others. In order to be part of a genuine joint action, maybe agents should have compatible plans of actions; or they should conditionally commit to do their part; or should have common goals; all of these suggestions make sense in a certain range of situations, and depending on what one considers to be a paradigmatic example of joint action. In other words, theorists have based their definitions on intuitive cases, and as a consequence have tended to identify only sufficient conditions for joint action, which turned up to be comparable to one another only with difficulty.

Against this background, Tuomela’s account stands out for several reasons. First, it has always strived to encompass all forms of joint action, from the weakest to the fullest, rather than privileging only one. Second, as can be seen in ‘Cooperation as joint action’, it is in essence compatible with the main intuitions behind competing accounts (such as Bratman’s, Gilbert’s or Miller’s). Third, and this is the gist of Tuomela’s present article, it now focuses on measurable properties of joint action, that is, on how efficient, functionally independent from and irreducible to simpler collective actions it is, for conceptual as well as empirical reasons. For too long, definitions of cooperation have ignored, or at least neglected, its explanations, that is, the various causes that are taken to lead people to cooperate (in a behavioural sense). In particular, theorists have

neglected the kinds of *reasoning* that can lead to cooperative behaviour, their normative properties and the empirical evidence that may support them. If joint action is defined by collective intentions based on a given mix of beliefs, goals, commitments and/or reasons to act, how can these be attained in the first place? To which kind of cooperative behaviour do they lead? Are they understandable from, generated by or reducible to individual mental properties? These are the questions that Tuomela now sets about answering. In the following discussion, I will focus on three related topics of we-mode joint action: its possible rationality, its greater efficiency, and its alleged irreducibility to purely individual properties, which are recurring points of the paper.

Usually, for a joint action to be rational is not considered as a necessary property, but at best as a welcome one. Suppose I define joint action as a set of intertwined individual beliefs, goals, intentions, etc. It may seem that the origin of these mental states should not matter, as long as the agents entertain them. However, definitions usually mention the causes that lead to these mental states. For instance, in Tuomela's definition of we-mode cooperation (CWM), it is not enough that the basic conditions be mutual belief among agents; the mutual belief must also be a reason for these conditions to hold in the first place. The idea is that not only must agents entertain the adequate mental states; they must have reached them for adequate reasons as well. Simply put, real cooperation exists when we all have a common goal and decide to act partly because we are aware of this commonality. So reasons to act enter the picture, although nothing is said about how they are assessed and combined within the agents' reasoning process. It is thus legitimate to ask whether it is rational for agents to cooperate when in the we-mode.

Tuomela answers positively. First, mutual beliefs must be rationally attained (74). Moreover, agents must

“rationally presuppose—and in general also believe—that the other participants will indeed perform their parts [. . .]. Basically, it is neither rational nor ‘group-socially’ beneficial for one of them to defect at least as long as the others do not defect. A participant cannot defect without being legitimately criticizable by the other.” (74)

In the we-mode, agents act for *group reasons* in order to realise a *collective goal*, to which they are *collectively committed*. The previous quote shows that these three conceptual pillars are actually supplemented by a fourth, hidden one, namely the requirement that agents reason collectively. A form of rationality becomes integral to the definition of cooperation, which may be problematic with weaker forms of cooperation, but not with the full-fledged forms that we-mode aims to encompass.

So what is collective rationality? It implies that one adopts the group's interest, but this is not enough. Agents could perfectly try to further the group goals by reasoning individually; this corresponds to what Tuomela calls the pro-group I-mode. To reason collectively entails that one adopt the group's point of view and does her part of what is the best global strategy for the group—for

instance what a group leader would decide¹ is better. This buys the we-mode group some ontological worth by bringing it closer to being an agent itself; individuals reason from the point of view of such an hypothetical agent.

However, the place of rationality in the we-mode raises two possible issues. First, it seems that the importance of rationality is inversely related to that of collective commitment. For to say that defection is irrational in the we-mode entails that any participant can legitimately criticize defectors, which in turn is just a way of expressing the fact that all participants are collectively committed to the goal and cannot unilaterally decide to stop furthering it. Indeed, in classical game theory, in which rationality reigns supreme, there is no need for the concept of commitment: a strategy is either irrational, in which case no commitment to it can be credible, or rational, and commitment to it is superfluous. As a consequence, collective rationality may at least join and at best replace collective commitment as a main conceptual pillar of the we-mode.

Another worry is that if agents in the we-mode genuinely adopt the group's interests and goals, and if "in principle, free-riding is an I-mode phenomenon" (83), then there should be no need for sanctioning agents or giving them incentive. However, such actions are expected from group leaders and participants (75; 76). Intuitively, a possible reason for this would be that when agents mistakenly consider two actions as equally beneficial with respect to the group goal and choose the worse one, others may resort to sanctions in order to steer them towards the better option. However, if an agent genuinely has the group's interests at heart, communication should suffice to correct her mistake. Overall, there is a dilemma: either agents can fully adopt the group goals and sanctions or incentives should be idle, or they never can and there is no we-mode.

I now turn to the second topic, namely the efficiency of the we-mode. According to Tuomela, "choice-theoretic matters [show] that sometimes the we-mode approach is functional-rationally better than the pro-group I-mode one" (77). As the end of the section (4) shows, Tuomela's case rests on situations of coordination such as the Hi-Lo game,² in which agents have to coordinate on one of two options, one Pareto-dominating the other (simply put, being preferable for all agents). Although it may seem obvious that both players should choose the action which is part of the dominating outcome (namely, H), according to game theory it is just as rational for individually rational players to choose the other one (L). What matters to rationality is only stability—I choose the best option given that the others do their part, and they do the same. Interestingly, in this situation, both Tuomela's we-mode and Bacharach's team-reasoning prescribe that H is the only rational solution. Thus they can be said functionally better, because they lead to strictly better outcomes.

¹ And in some cases *can* decide: Tuomela's account takes into account situations in which groups effectively have leaders.

² Typically represented by the following game matrix:

Player 1 \ Player 2	H	L
H	2,2	0,0
L	0,0	1,1

There is a difference though. Consider Tuomela’s remark: “The switch to group thinking (esp. group agency and we-reasoning) together with the assumption of Pareto-optimality (viz. ‘common interest’) [...] will often lead to group-rational mutual cooperation.” (82) Group thinking is collective reasoning; it leads agents to do their part of the group goal. Pareto-optimality, which is not conceptually part of the we-mode, is a condition which helps determining what this goal is. However, this mitigates the functional efficiency that is supposed to distinguish collective reasoning from I-mode pro-group reasoning. For the latter would *equally* lead to choose H if supplemented with a Pareto-optimality condition (it would suppress the outcome in which players both do L from the list of possible goals). But Bacharach proposed his model of team-reasoning precisely in order to get rid of such a condition; so the we-mode ‘group-thinking’ is not easily comparable to it. This damages the analogy between game-theoretic collective reasoning and the we-mode, and thus the claims to the irreducibility of the we-mode.

Moreover, in games other than coordination problems, the gap between collective and individual rationality narrows down even further. In a Prisoner’s dilemma (and social dilemmas in general), the we-mode can rationally explain cooperation *just as well as the pro-group I-mode can*. So even if they formally lead to different equilibria, the former does not lead to better cooperative outcomes and may even be considered as harder to attain. This point could be made formally; but intuitively it is reflected by the fact that the CWM definition necessitates much stronger conditions than the CIM one. Overall, the reasons for finding the we-mode ‘functional-rationally better’ are ambivalent. Note that the similar behavioural consequences of the we-mode and the pro-group I-mode in social dilemmas is precisely why, as Tuomela remarks, empirical studies do not provide a “clear and unambiguous evidential argument for the we-mode we-perspective” (85).

Finally, is the we-mode irreducible to the I-mode? Because it purports to define joint action analytically on the basis of apparently purely collective concepts, Tuomela’s account of joint action strikes a balance between individualistic approaches (exemplified by Bratman) and holistic but non-explanatory ones (such as Searle’s). But is it really irreducible to individual properties? Note that there are different ways to be ‘irreducible’ (ontologically, conceptually, functionally...).³ Here, it is useful to keep in mind Tuomela’s following claim: “The we-mode approach here is not reducible to the individualistic, I-mode framework because it necessarily differs from it with respect to two facts: it is based on a group agent as the basic agent and it involves we-reasoning instead of I-reasoning.” (82)

First, the foregoing discussion has shown that functional irreducibility may be dubious: differences between the we-mode and the pro-group I-mode are, or rather between collective and individual reasoning, are not clear, whether one considers coordination problems or social dilemmas.

³ “There are both conceptual, ontological, and functional differences between the we-mode approach and the pro-group I-mode approach.” (70)

Second, what about ontological irreducibility? While groups are obviously not individuals themselves as they do not have bodies, consciousness, etc., some of their properties may still be irreducible to that of individuals. One argument against irreducibility could be the following: since all collective properties involved in the we-mode are collectively built, that is, are the product of interactions between individuals, ultimately their origin can be explained in terms of individual properties. At best this could only mean that the we-mode is diachronically reducible, not that it isn't synchronically. Nonetheless, Tuomela does not discuss ontological irreducibility (Cf. 77).

I think the most promising thesis is that of conceptual irreducibility, which seems to be more fundamental.⁴ Conceptual irreducibility can be understood in two senses, depending on whether one talks of the collective concepts that define the we-mode (first sense), or of the concepts that *agents* resort to when acting in the we-mode are irreducible (second sense).

In the first sense, there are reasons to think that collective properties do not supervene on individual ones. For instance, Tuomela emphasizes that group interests are collectively accepted. Collective acceptance typically does not supervene on individual properties: given a configuration of individual beliefs, collective acceptance could be present or absent. This depends on the decision procedure employed by the group, which in turn depends on the group's history. For instance, that a high proportion of members believe something may lead to its acceptance if majority voting is used, but not if decisions are taken by an (previously) elected subgroup whose members happen to not share these beliefs.

Now for the second sense. Even if the group agent is not ontologically irreducible, it plays a prominent role in the participants' reasoning, because they do their part of what is best *from the group's point of view*. Even when this decision is not actually taken, it is as if the group agent, or a hypothetical leader of the group, first took a decision, which agents then follow. Taking into account a group agent of some sort is necessary to the we-mode. As we have seen, collective reasoning does not differ too much from individual reasoning in terms of observable results; but two inferences rules can be conceptually distinct even if they lead to identical conclusions in most cases.⁵

Finally, some empirical evidence appears to support conceptual irreducibility. A recent paper by Tomasello et al. (2005) suggests that cognitive abilities for shared emotions, shared goals and joint intentions appear in parallel with abilities for their non-shared counterpart, and in particular are not the result of complex formation of beliefs about the beliefs of others—as one could expect if individualistic definitions of joint action were literally true. This is consistent with the view that the joint properties of a situation do not arise from complex sets of individual properties, and consequently supports conceptual irreducibility of the we-mode.

⁴ "If concepts are irreducible [...] then it is reasonable to think in this case that the states and events to which they correctly apply are ontologically irreducible as well." (77f.)

⁵ According to Tuomela, "the we-mode approach here is not reducible to the individualistic, I-mode framework because it necessarily differs from it with respect to two facts: it is based on a group agent as the basic agent and it involves we-reasoning instead of I-reasoning." (82) I agree with the former reason rather than with the latter.

Overall, some of Tuomela's theses are convincing—for instance, it is likely that the way in which humans jointly act is irreducible to individual actions—and others only partially so. But the important point is this: Tuomela shows that literature on joint action is now mature enough to engage decision/game theory and to start having assessable empirical literature. By and large, the jury is still out on the questions of the rationality, efficiency and irreducibility of joint action. But at least these questions can now be asked, and precise answers can be sought.

Bibliography

- Bacharach, M. (1999), Interactive Team Reasoning: A Contribution to the Theory of Co-operation, in: *Research in Economics* 53, 117–147
— (2006), *Beyond Individual Choice*, Princeton
Tomasello et al. (2005), Understanding and Sharing Intentions: The Origins of Cultural Cognition, in: *Behavioral and Brain Sciences* 28, 675–735
Tuomela, R. (2011), Cooperation as Joint Action, *this issue*