

<https://helda.helsinki.fi>

Planning in the We-mode

Hakli, Raul Aleksi

Springer International Publishing
2017

Hakli, R A & Mäkelä, P A 2017, Planning in the We-mode . in G Preyer & G Peter (eds) , Social Ontology and Collective Intentionality : Critical Essays on the Philosophy of Raimo Tuomela with His Responses . , 9 , Studies in the Philosophy of Sociality , vol. 8 , Springer International Publishing , Cham , pp. 117-140 . https://doi.org/10.1007/978-3-319-33236-9_9

<http://hdl.handle.net/10138/311025>

https://doi.org/10.1007/978-3-319-33236-9_9

unspecified

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Planning in the we-mode

Raul Hakli & Pekka Mäkelä

February 16, 2016

Abstract

In philosophical action theory there is a wide agreement that intentions, often understood in terms of plans, play a major role in the deliberation of rational agents. Planning accounts of rational agency challenge game- and decision-theoretical accounts in that they allow for rationality of actions that do not necessarily maximize expected utility but instead aim at satisfying long-term goals. Another challenge for game-theoretical understanding of rational agency has recently been put forth by the theory of team reasoning in which the agents select their actions by doing their parts in the collective action that is best for the group. Both planning and team reasoning can be seen as instances of a similar type of reasoning in which actions are selected on the basis of an evaluation of a larger unit than an individual's momentary act. In recent theories of collective agency, both planning and team reasoning have been defended against orthodox game theory, but, interestingly, by different authors: Michael Bratman has extended his theory of planning to the case of shared agency, but he does not seem to see a role for team reasoning in understanding shared intentional activities. Raimo Tuomela has defended team reasoning in his theory of group agency, but he ignores temporally extended planning in this context. We argue that both accounts suffer from this one-sidedness. We aim to combine the main insights of Tuomela's we-mode approach and Bratman's planning approach into a fruitful synthesis that we think is necessary for understanding the nature of group agency.

1 Introduction

In his new book, *Social Ontology* (2013), Raimo Tuomela develops a theory of group agents and collective intentionality extending his previous theory of I-mode and we-mode sociality (see Tuomela 2002, 2007). Whereas Tuomela has previously been somewhat reluctant to talk about group agents and instead focussed on social action, joint action, and cooperation between individual agents, in the new

book there is a clear focus on group agency. However, his main positions have not changed. He argues that even though groups as interactive social systems can be seen as intentional agents, their intentionality derives from the intentionality of the individual group members (Tuomela 2013, 23): Individuals' mental states are assumed to be intrinsically intentional, with raw feels and qualia, whereas the intentional properties of group agents are based on the individual group members' collective attribution of attitudes, such as beliefs, desires, and intentions, to the group. Thus in Tuomela's terminology, group agents are intentional only in a derived or, alternatively, in an extrinsic sense: Group agents' intentionality is based on the members' collective construction of the group as an intentional entity that they identify with. At the core of this idea is the group members' collective self-attribution, via an irreducible notion of a group, of intentional states to themselves qua a group.

In our understanding Tuomela's position differs from such positions according to which intentionality of an entity or a system is grounded in interpretation or attribution from a perspective external to the entity or system. Tuomela can avoid potential circularity problems that some such accounts may be subject to. For instance, some forms of interpretationism take all intentionality to be based on interpretation of actions as intentional; however, it is not clear how such accounts can explain the intentionality of interpretation and attribution which are normally understood to be intentional as well. Tuomela's account also differs from standard functionalist accounts that aim at finding warranted analogies between individuals and groups that would allow the attribution of intentional attitudes to groups. In Tuomela's view the intentionality of group agents is "anchored" in the intrinsic, realistically understood intentionality of constituent individual group members, and it seems to provide us with a "stronger" view of the intentionality of the group agents than views in which the intentionality is purely a matter of external interpretation, e.g. some forms of instrumentalism.

Obviously Tuomela's position allows for descriptive, predictive and explanatory sense-making of the behaviour of group agents both from internal, group member's perspective, and external non-member's perspective. This sense-making grounds the main argument for positing group agents. From the outside, a group can be viewed as an agent with cognitive and motivational attitudes that can be used to explain and predict its actions. But because the attitudes are grounded in the self-attribution of intrinsically intentional group members there is not only a correctness criterion for external attributions but also an answer to the question how groups can have real causal powers. Namely, as Tuomela argues, when individuals view themselves as a group their behaviour changes. Thus, even if group agents and group attitudes are "fictitious" entities they have real causal effects due to the behavioural changes in the individuals who constitute groups and identify with them.

How do the behavioural changes come about? Viewed from the inside, the attributed cognitive and motivational attitudes help to coordinate the individual group members' actions. If *we* intend to do something then *I* can focus on my part action and rely on others to do their parts. The main mechanism that is responsible for coordinated group action is we-thinking and we-reasoning in which the group members adopt the group agent's perspective: They take the group's beliefs and intentions as premisses for their practical deliberation and select their part actions as if they were mere body parts of a larger intentional agent instead of fully autonomous individual agents.

Indeed, in Tuomela's account there is an important conceptual distinction between thinking and acting in the I-mode and thinking and acting in the we-mode, where the latter means that agents give up some of their "natural authority" and act solely for group reasons, that is, for reasons that are derived from the group agent's attitudes. According to Tuomela (2013, 38), "[a]n ideal we-moder can think and act only for group-centered motives, regardless of whether they conflict with individualistic motives." I-moders can, in contrast, act on mixed motives that are partly personal and partly group-centred. The notion of a we-moder is idealised but it explicates the idea of rational group agency: A group agent can act rationally on the basis of its collectively constructed group attitudes, given that the individual group members—who are ontologically speaking the only real agents—act on its behalf based on its attitudes as if they had no attitudes of their own. This is a sophisticated theory of group agency and it aims to be comprehensive in trying to account both for the construction of group agents and their attitudes and for the way these group agents and attitudes shape the thinking and acting of the individual group members. It is very difficult to find books that try to cover comparable ground in as detailed a manner. There are detailed accounts of either the formation of groups and group attitudes (for instance, List & Pettit 2010) or the effects of groups to individual group members' behaviour and decision-making (for instance, Bacharach 2006), but not both.

However, in spite of the theory's sophistication and coverage, there is no account of planning for group agents in the book, even though planning is arguably a crucial element of rational intentional agency. Historically speaking, rationality of actions has often been understood in terms of selecting the next action on the basis of some desirable features or consequences of that action, for instance, on the basis of how well that particular action serves the agent's goals or desires. This picture has been argued to be inadequate for a long time already by several philosophers (see, e.g., Bratman 1987, Pollock 1995). It is based on an old-fashioned belief-desire (or BD for short) conception of rational agency in which agents select their actions based on how well these actions, according to their beliefs, are expected to satisfy their desires. These accounts have to a large extent been replaced by so-called belief-desire-intention (or BDI) accounts that take into account the cru-

cial role of intentions in rational agency. A particularly influential account that has paved the way for BDI theories is Bratman's (1987) planning account of intention, which emphasizes the long-term commitments in the pursuit of an agent's goals and objectives.

Tuomela has positioned himself in the BDI camp, both individual and group intentions play a role in his theory, and he has previously discussed planning in the context of joint action (Tuomela 2007). However, he does not discuss how group agents plan their temporally extended action sequences but focusses on simple decision-making where the next action is selected, for instance, on the basis of how well that action maximises group utility. In this paper, we aim to argue that planning has an important role to play in rational group agency just as in rational individual agency, and therefore Tuomela's theory still lacks a crucial component.

In addition to this critical point, we aim at providing a constructive extension of Tuomela's theory by sketching a theory of planning for group agents and group members. We hope that such an extension will aid to understanding of how groups coordinate their actions, and it will also establish new connections between Tuomela's theory and literature on planning and deliberation. We argue that his distinction between I-mode and we-mode can be applied to the case of planning in addition to simple decision-making. Moreover, we try to show that we-mode planning is in an important sense different from I-mode planning and that some ideas of Tuomela's we-mode theory can be directly and fruitfully applied to the general theory of planning. We will present our critical argument in more detail in Section 2. We will take the first steps in developing a theory of we-mode planning in Section 3, and we will conclude in Section 4.

2 Importance of planning for rational agency

Still a few decades ago the main position in philosophical action theory was the so-called belief-desire view that intentional action of an agent was to be understood in terms of the agent's beliefs and desires. Typically it was thought that these attitudes both cause the agent's action and rationalise it. An explanation why someone acted (or, taking a normative stance, why they had a reason to act) in the way they did was to be couched in terms of the agent's beliefs and desires. A standard example would be someone opening her umbrella, which could be explained by citing her desire to stay dry and a belief that because it is raining, she would have to open her umbrella in order to stay dry. The belief and desire together serve to causally explain her forming the intention to open her umbrella, and they also give her a reason to open her umbrella. Intentions in this picture figure only as conclusions or outputs of practical reasoning the premises of which concern beliefs and desires. Moreover, intentions were typically not conceived as

distinctive attitudes but reducible to certain combinations of beliefs and desires.

This picture came under heavy criticism and is no longer generally accepted. Some critics argued that the connection from beliefs and desires to intention and action is not causal, whereas others argued that beliefs and desires are not reasons for action. Some argued more generally that folk-psychological concepts like beliefs and desires are not the right categories for understanding intentional action in the first place. Here we will focus on the more specific criticism that endorses the use of folk-psychological concepts in understanding action but argues that more is needed besides beliefs and desires to account for rational agency in the case of agents capable of planning future actions.

Most important criticism of this type came from Michael Bratman (1987). He argued convincingly for the irreducibility of intentions to beliefs and desires and for the important role of intentions, not only as conclusions of practical reasoning but as its premisses as well: Previously adopted intentions must constrain future deliberation and they must therefore be taken into account in practical reasoning in addition to desires and beliefs. This insight together with Bratman's detailed theory, which established a close connection between intentions and plans, paved way to general adoption of BDI accounts. According to Bratman (1987, 8-11), the BD account cannot do justice to the fact that human beings are planning creatures. In addition to intentions in action, we have future-directed intentions that are targeted to actions to be performed later. In order to satisfy our long-term goals and objectives, we need to coordinate several actions and combine them into larger units. Moreover, due to our resource-boundedness we cannot constantly deliberate about all possible courses of action but we need to limit the space of deliberation by fixing some ends or general actions by committing ourselves to them and then deliberate only on the means or on more specific actions. These features of planning cannot adequately be accounted for by the BD model, which focusses on intention in action, on selecting only the next action to be performed. Were we not able to make decisions concerning future actions but always had to go for the satisfaction of the strongest desire, planning would be impractical if not impossible.

Even though these points are now widely recognised in analytical theory of action, the importance of future-directed intentions, plans, and commitments, has not been everywhere acknowledged. For instance, theories of rational choice, that is, decision theory and game theory, stem from old-fashioned belief-desire ideas: In these theories agents are modelled in terms of probabilistic expectations about possible events, and preferences or numerical utility functions about possible outcomes. The agents' expectations are thought to correspond to their beliefs, and their preferences are thought to encode their desires given their beliefs. It is not clear whether these theories have concepts that would correspond to intentions. They do have a notion of the choice of strategy, which could be seen as closely

related to intention in action, but future-directed intentions seem not to be easily defined in decision- or game-theoretical terms, basically because it is not clear how to define commitments to future action.

John L. Pollock (1995) has argued for the inadequacy of the decision-theoretic model of practical reasoning. According to him (1995, 183), the model in which agents choose acts by computing their expected utilities cannot properly deal with situations in which the payoff results from performing sequences of actions. In order to deal with such situations we would need to be able to compare different plans. But the comparison cannot be done by calculating expected utility for alternative acts because one act can be the first step of several plans, and, Pollock argues, there is no way to compute a meaningful expected utility for such an act unless we have already decided which of those plans to follow. For that we would need a way of comparing the relative goodness of different plans. And if we have that, then we should just compare all the relevant alternative plans and perform the first act in the best plan, which means that the expected utility calculations of acts would be redundant. What is really needed is a method of evaluating and selecting between alternative plans, and that is what Pollock tries to develop.

However, Paul Weirich (2004, 197) argues to the contrary that there is no reason to optimize among plans. According to him, if optimization among plans leads to different acts than optimization among momentary acts, it recommends an act that is not the best one. However, it can only do this because the act is part of the best plan and the best plan promises gains that outweigh the gains from the best momentary act. This, according to Weirich, is impossible because the possibility of future gains must be taken into account in the evaluation of momentary acts. Hence, following plans is rational only insofar as every step of the plan maximises expected utility of momentary acts.

There need not be a strong conflict between Pollock's and Weirich's views, because both admit that there is a difference, in Pollock's terms (1995, 199-205), between warrant and reasoning, and in Weirich's terms (2004, 195), between standards of evaluation and decision procedures. Weirich's main interest lies in warrant, or standards of evaluation, and he argues that momentary acts are the basic units of control and should also be the basic units of evaluation. The rationality of plans derives from the rationality of momentary acts, even though cognitive limits might make optimization among momentary acts too high a standard for humans who for that reason are justified in deliberating about plans. Pollock, on the other hand, is mainly interested in practical reasoning. He admits that it may be possible to formulate a theory of practical warrant that does not mention plans, but a theory of practical reasoning for computationally limited agents must mention plans. While it is controversial to claim that planning might have a role to play in evaluation of actions, it seems uncontroversial to claim that planning is important for deliberation and decision-making in the case of resource-limited agents.

Obviously humans are such resource-limited agents, but if this is the case, then groups are resource-limited too. Therefore, if the argument for the necessity of planning is accepted in the case of humans, it should be accepted in the case of human groups as well. After all, groups are composed of human beings and therefore the limitations of humans must pose limits to the resources of groups as well. Even though groups may benefit from combined resources of its human members, there can only be limited number of human members, and thus the resources of groups are limited as well. Group agents may have even greater needs for temporally extended action coordination that plans provide precisely because they are composed of several individual planning agents whose actions must be coordinated. Groups are temporally extended and their goals should provide constraints for the group members just as individuals' intentions provide constraints for their deliberation.

One might think that if human beings are understood as planning agents then groups need not be because whatever goals and intentions groups may form, it is enough if the individual group members treat them as constraints in their own deliberation. There previously adopted group goals could still function as constraints that the individuals should take into account in their deliberation even if they are temporarily not acting as group members but as private persons or as group members of some other groups. However, in general, and in Tuomela's theory in particular (see, e.g. page 26), social groups are understood to be more than just collections of individual people: The identity of a group need not be tied to particular group members. This is clearly the case at least in large organized groups that belong to the domain of Tuomela's theory. Such groups may have roles and positions that are occupied by different individuals at different times, but the goals and intentions that the group adopts should remain in force until they are changed even if the constitution of the group undergoes changes.

Our argument can thus be structured as follows:

1. Tuomela aims at providing a theory of group agency that applies, in addition to small social groups, to large organized groups such as nations, companies, and organizations.
2. A theory of group agency requires an account of how the group's intentions guide intentional actions of group members to satisfy the group's intentions.
3. Tuomela has shown how the group's intentions guide intentional actions of group members only in simple cases of selecting the next action.
4. Groups, especially large organized groups, are temporally extended, and accounting for their agency requires accounting for cases of long-term planning in addition to simple cases of selecting the next action.

5. Therefore, Tuomela has not fully succeeded in accounting for group agency that applies to large organized groups.

Premiss 1 states one of the main aims of the book and, unlike most other theorists in the field, Tuomela explicitly aims to deal with large organized groups (see, e.g., pages x, 5, 90). We think Tuomela would accept premiss 2 because of his sympathy for BDI accounts that emphasise the role of intentions in guiding actions (see, e.g., page 62) and because of his general aim of showing how group members, when they are acting in the we-mode, take the group's intentions and attitudes as reasons for their actions (see, e.g., page 39). Premiss 3 can be verified by noting that Tuomela's focus is on game-theoretic examples and he does not present a detailed account of group planning in his book. He does occasionally mention plans in passing, and in our opinion this supports the assumption that he would accept premiss 4. Consider, for instance, the following passage discussing an example in which a group has formed an intention to paint a house (Tuomela 2013, 107): "The group's intention then involves as its content the goal that the house will be repainted, and directs it to form relevant subplans and to guide its members' conduct in the appropriate manner." Here Tuomela seems to accept the idea that group's intentions have an effect on group member's plans, but nowhere in the book does he say whether group's intentions should be understood as plans nor does he elaborate on how group's intentions direct temporally extended planning. He does not specify whether he endorses theories of planning presented by Bratman and others nor does he develop a planning account of his own.

Although Tuomela (2013, 182) recognises the conceptual poverty of decision theory and game theory, he nevertheless makes a lot of use of game-theoretical machinery in his theorising. He presents game-theoretical arguments to back up his theses about the importance of we-mode thinking and acting in spite of the limitations of game theory. Indeed, he tries to overcome some of its limitations by extending game theory to the case of groups employing theories of team reasoning developed by Sugden (1992) and Bacharach (1999). But the limitations Tuomela is concerned with have more to do with what kinds of agents game theory allows for and not with extending the theory to deal with plans and commitments.

One of Tuomela's main aims in the book is to demonstrate that there are functional differences between we-mode acting and what he calls pro group I-mode acting. He argues that there are cases in which pro group I-mode faces problems and we-mode fares better. However, the arguments and the cases are couched in game-theoretical terms and therefore it becomes hard to see whether they actually apply to accounts that Tuomela classifies as pro group I-mode accounts, most notably Bratman's (2014) account of shared intentions. This is because Bratman's theory steers clear of game-theoretical machinery and resorts to planning concepts instead. We aim here to translate some of Tuomela's ideas concerning we-mode

and pro group I-mode to planning terminology in the hope that this would also enable a fruitful comparison between his theory and theories like Bratman's. In addition, we think that his distinction between I-mode and we-mode can be applied to planning as well: We can see Bratman's theory as an I-mode theory of planning and contrast it with what we will here call the we-mode theory of planning.

Moreover, we would like to point out that there are commonalities in the strategies of Tuomela and Bratman due to analogies between coordinating multi-agent actions and coordinating one's present and future actions discussed, e.g., by Christopher Woodard (2008). In order to see this, consider the criticism to game theory posed by team reasoning. The general criticism is that game theory is committed to a fixed agent type. Agents can be understood either as individuals or groups, but it is not possible to account for a situation in which an agent is part of a larger agent, a group agent. Modelling such situations requires an extension like Bacharach's (1999) theory of unreliable team interactions that incorporates ideas of team reasoning into game theory. Team reasoning, or we-mode we-reasoning in Tuomela's terminology, is basically a method of selecting actions as a group member, as a part of a larger unit. In it, the group members evaluate the possible action combinations from the point of view of the group agent and then perform actions that correspond to their parts in the best group action.

Similarly to the idea of extending decision theory to accommodate intentions or plans, the idea of extending game theory to accommodate group agents consisting of individuals is controversial. Critics of team reasoning, like Paul Weirich (2009), complain that team reasoning, which is based on maximization of group utility, may recommend individuals to perform actions that do not maximise the expected utility of the individuals. Thus, it recommends actions that are irrational for the agents.

The problem seems to be the same in both cases: According to decision theory and game theory, the selection of an individual's momentary action should be based on the evaluation of the individual's momentary action, not on some other entity of which the individual's momentary action is a part. Call this *the atomistic act evaluation principle*. Planning violates the atomistic act evaluation principle by allowing that an individual agent may rationally select her momentary action based on an evaluation of a larger, temporally extended plan, of which the agent's momentary action is a part. Team reasoning, on the other hand, violates the principle by allowing an individual agent to rationally select a momentary action based on an evaluation of a larger, joint action, of which the momentary action is a part. Insofar as the selection of action is based on maximising the value of a larger unit, be it a single-agent sequence of actions or a multi-agent collective action, then, depending on how the value is calculated, the result may be an action that does not maximise the utility of the single-agent momentary action, and is therefore

irrational. Woodard (2008, 64) names a similar principle the Identity Thesis: According to it, the “unit of action” is always identical to the set of actions that an agent could perform now if she chose to do so. Basically this means that an agent should base her deliberation only on the properties of those actions that are immediately subject to choice by the agent, thus leaving out of consideration collective actions that depend on others’ choices and temporally extended plans that depend on one’s future choices.

Both Bratman and Tuomela seem to violate the principle. Bratman does it by allowing previously adopted plans to serve as input for deliberation and to constrain the choice of future actions. Tuomela does it by allowing the group’s intentions to serve as input for deliberation and constrain the choice of actions of individuals. In both cases, it is possible that an agent has to discard the utility maximising action because it violates the agent’s commitment, either to a temporally extended plan or to a group intention.

But is this too quick? Recall the distinction between warrant and reasoning (in Pollock’s terminology) or between standards of evaluation and decision procedures (in Weirich’s terminology). Maybe Bratman and Tuomela are only concerned with reasoning and decision procedures but not about warrant and evaluation. If they were only concerned with heuristics concerning how resource limited agents can make practical decisions under non-ideal conditions but refrain from attempts to justify those decisions independently of expected utility calculations they might evade the criticism of violating the atomistic evaluation principle.

This is not the case, however. Both Bratman and Tuomela allow that the evaluation of the larger unit of action affects the evaluation of the single-agent momentary action. They both allow that prior commitments provide reasons for actions (albeit of special type of reasons) and thereby have an effect on what is rational to choose, an effect that seems to be independent of expected utility calculations based on beliefs and desires. In the case of Bratman (1987, 32-35), prior plans constrain future actions providing what he calls a filter of admissibility. Actions that are inconsistent with adopted plans are excluded from consideration. This is not merely a heuristic device for making deliberation tractable, but a reason-giving mechanism: Prior plans provide what he calls framework reasons for future actions. These reasons do not directly compete with ordinary reasons for actions but they structure the process of weighing those reasons. They provide coherence and consistency considerations that deem some options inadmissible or irrelevant and thus to be excluded from the process of deliberation. Similarly, Tuomela (2013,11) talks about group reasons, which are reasons for individuals to participate in group activities. The group reasons are based on the group’s intentions and attitudes, and they may override the agent’s private, possibly conflicting reasons. It seems then that these reasons for action provide evaluation considerations that are independent of the agent’s expected utility calculations. Both Bratman’s

framework reasons and Tuomela's group reasons seem to fall within the category that Christopher Woodard (2008, 2011) calls pattern-based reasons: They are reasons for or against an action because that action is part of some larger pattern of action which has certain evaluable qualities. It is the goodness of the pattern that provides reasons to perform its parts, not the goodness of the part itself.

Both Bratman's theory of planning agency and Tuomela's theory of group agency therefore seem to reject the atomistic act evaluation principle. It is particularly interesting to note that both endorse one way of violating it and give very little attention to the other. As we have noted, Tuomela (2013) has a lot to say about group agency and team reasoning in his new book but he ignores temporally extended planning almost completely. Similarly, in his new book, Bratman (2014) is exclusively focussing on temporally extended planning but makes only a few remarks concerning team reasoning, remarks that seem to indicate that he does not see a role for it in understanding shared intentional activities. That understanding he takes already to be provided by his planning theory extended to the multi-agent case in which individuals form shared intentions and shared plans.

It seems to us that once you have already violated the atomistic act evaluation principle either by allowing planning or team reasoning, you can as well allow both. This is because of the deep relationships between planning and team reasoning. They both seem to be instances of the same type of decision-making: They both combine several single actions into complex actions, comparatively evaluate these complex actions, and then select (and rationalise) part actions of the best combined action. Team reasoning is coordination of actions between individuals, planning is coordination of actions between one's past and future selves. One could therefore argue that team reasoning is momentary planning across individuals, and planning is individual team reasoning over points of time. Given these interconnections why not combine them, why not consider planning group agents?

In the end, this is what we would like to do: To combine Bratman's general idea of the importance of planning and the roles of plans in deliberation with Tuomela's idea of group agency and we-mode practical reasoning applied to the selection of plans. The combination will lead to a model of practical reasoning in which the primary agent is the group and plans are the central objects of selection: The group selects the best plan, and the individuals extract their sub-plans from the best group plan. We think that a theory of group agency is incomplete without a theory of planning group agency, even though we can only adumbrate such a theory in this paper: We focus on the main features and postpone details like applying Tuomela's we-mode criteria to planning to further work. Moreover, we think that a general theory of planning, like that of Bratman's, might also benefit from some ideas from theories of group agency, in particular, the idea of team reasoning, and we try to argue to that effect in the rest of the paper.

3 Planning for group agents

Bratman has presented a theory of planning for rational agents in which future-directed intentions are a central element. The major roles of intentions are, according to Bratman (1987, 16–17) the following: (1) Intentions are conduct-controlling attitudes. This means that if a prior intention manages to survive until the time to take the action, then it will determine the course of action: The agent will at least try to execute the intention (Bratman 1987, 108). Unlike desires, which are merely potential influencers of action, intentions will not only influence action, but they will control it due to the commitment that is involved in intention. (2) Intentions have stability or inertia. This means that they resist reconsideration and constrain other intentions (even though they are not irrevocable): In normal circumstances in which there is no special reason to reconsider prior intentions, they persist and constrain the agent’s future deliberation because the agent considers the course of action settled. (3) Intentions pose problems for further deliberation and lead to formation of new intentions. For instance, an intended end requires that the agent plans for how to achieve it and adopts an intention concerning the means. Similarly, a general intention to do something requires specifying further details leading to the formation of more specific intentions. Bratman calls the first role the volitional dimension of commitment and the second and third the reasoning-centred dimension of commitment.

Plans, according to Bratman, (1987, 29) are “intentions writ large”: They share the three main roles of intentions. In addition, they are typically partial, meaning that they are not fully specified but allow for details to be filled in in due course, and they typically have a hierarchical structure, meaning that plans concerning ends embed plans concerning means, and more general plans embed more specific ones. These features make sense for agents with bounded resources, because they enable us to coordinate our future actions both intra- and interpersonally in broad outlines without requiring all the details to be specified in advance, which could easily lead to waste of resources due to unanticipated changes in the environment.

Bratman (2014) has extended his theory of single-agent planning to the multi-agent case. In the case of intentional group activities, Bratman talks about shared agency and shared intentions. What is important for Bratman is that shared intentions are individuals’ intentions concerning shared activities. They are ordinary intentions attributed to individuals, not group intentions attributed to a group (as in Tuomela’s theory) or to a plural subject (as in Gilbert’s theory), nor is there need to attribute specific we-intentions to the individuals (as in Tuomela’s and Searle’s theories). Bratman advocates what he calls the continuity thesis, according to which the move from the single-agent case to the multi-agent case is conceptually conservative (Bratman 2014, 4): Understanding sociality and shared agency

does not require radically new conceptual, metaphysical, or normative machinery beyond what is needed to account for individual planning agency. With respect to the intentional attitudes, the basic building blocks of individual beliefs, desires, and intentions, are thus sufficient to account for collective intentionality as well. Note that the strategy is clearly different from the mainstream approach of understanding collective agency which is based on finding warranted analogies between individual agents and group agents that allow the attribution of intentional attitudes to groups. This approach can be seen in Tuomela's account, in Gilbert's plural subject theory, in List and Pettit's functionalism and in Tollefsen's interpretationism.

Let us see how Bratman characterises shared intentions. A central idea for him is that individuals can have intentions towards collective actions: I can intend that we do something together. For instance, I can intend that we dance tango and you can intend that we dance tango. Given that we have such intentions and certain other conditions hold under common knowledge, we have a shared intention to dance tango. And if our subintentions and actions are mutually responsive to these intentions, our shared intention leads to our dancing tango. This, according to Bratman, is a central case of sociality and acting together, and it is individualistic in the sense that it does not require more than attitudes attributed to individuals. However, there is a controversial bit, which comes from the idea of intending that. According to many philosophers, *intentions that* are derivative from *intentions to*, and intentions to have a built in restriction that we can only intend our own doings. In the literature this is known as the own action condition. The own action condition is violated by the idea of intending that we do something because it is very difficult to understand what it means for me to intend to do something that involves your doings as well.

However, because plans are understood as intentions in Bratman's theory, we can describe the difference in planning terms, and this may help us to understand what Bratman is after. Recall first, that having a plan should be not be understood merely as having a recipe for doing something but being committed to doing something, thus as having an intention. In Bratman's account, individuals can have plans that specify not only their own actions but also other agents' actions. When an agent considers alternative plans, these plans may be multi-agent plans in the sense that they specify what each agent is to do. It is not difficult to understand what it means for an agent to commit herself to a plan. It entails performing the actions specified for her in the plan and possibly also keeping an eye on whether the other agents can handle their parts and maybe having a disposition to help them if needed. In terms of plans this appears commonsensical and not problematic at all, but the problem appears when the situation is described in terms of intentions because the usual understanding of intentions involves the idea that one can only intend one's own actions.

Bratman (2014, 103) gives in a compressed form his sufficient conditions for shared intention to J as follows:

- A.** Intention condition: We each have intentions that we J; and we each intend that we J by way of each of our intentions that we J (so there is interlocking and reflexivity) and by way of relevant mutual responsiveness in sub-plan and action, and so by way of sub-plans that mesh.
- B.** Belief condition: We each believe that if the intentions of each in favor of our J-ing persist, we will J by way of those intentions and relevant mutual responsiveness in sub-plan and action; and we each believe that there is interdependence in persistence of those intentions of each in favor of our J-ing.
- C.** Interdependence condition: There is interdependence in persistence of the intentions of each in favor of our J-ing.
- D.** Common knowledge condition: It is common knowledge that A–D.

What is needed for shared intentional activity and modest sociality is that this shared intention to J leads to our J-ing in accordance with the following connection condition:

- E.** Mutual responsiveness condition: our shared intention to J leads to our J-ing by way of public mutual responsiveness in sub-intention and action that tracks the end intended by each of the joint activity by way of the intentions of each in favor of that joint activity.

Bratman (2014) does not talk about group agents, but if we follow Tuomela (2013) and accept that human groups can be agents, it seems clear that, like individual humans, they too are resource-bounded planning agents. Groups have an even stronger need for coordination of actions than individuals because they consist of several individuals who will have to be able to act together. Thus in order to account for agency of groups, we need to give an account of planning for group agents. The main elements of Bratman's (1987) planning theory can be applied to group agency as well, but now there are two levels to consider: the level of the group agent and the level of individual group members. The volitional dimension concerns mainly the level of individuals. Because groups cannot act without their members, all actual tryings and physical movements must be initiated by individual group members.

On the level of the group agent, the reasoning-centred roles are central: Once the group makes a decision concerning a course of action, it is committed to it, and rationality demands it to exclude from consideration other plans that are inconsistent with it. This applies to the individual agents as well: Due to the collective

commitment to the collectively accepted intention, the group members are normally committed to abstain from making plans that are inconsistent with or make it difficult for them to contribute to the execution of the adopted plan. If circumstances change during the execution of the plan, adjustments must be made. Small changes can usually be made at the individual level, but in some cases a need to reconsider the adopted plan may arise and, due to the collective commitment, such reconsideration must take place at the group level. Pollock (1995, 201-203) notes that reconsideration may be rational if the agent discovers an alternative, possibly preferable, plan that has not previously been given consideration. In such cases reconsideration will take place when the situation allows it. For instance, if the discovery has been made by an individual group member but the discovered plan affects the parts of the others, she will have to inform them and the reconsideration will be performed only when the group has a chance to evaluate the alternative plan.

The case of the other reasoning-centred role is similar: Once the group makes a decision concerning an end, it has to consider means to the end. The adopted intention poses problems for further deliberation. To an extent this can be done on the level of the group. The preliminary steps and subtasks have to be identified, their dependencies and resource demands have to be analysed in order to come up with a partial order of tasks, and some kind of understanding of how the tasks will be distributed to group members must be reached. Once the tasks are allocated to group members, refinement of the sub-plans can be carried out at the level of individuals. Of course, the individuals need to take care that the mesh between the sub-plans is maintained in the process.

Here the evaluation of plans plays a crucial role. Assuming that the group has agreed on a method of evaluating plans, for instance, the kind of a utility function that John Pollock envisages but applied from the group's point of view, the individual group members need not negotiate about which plan to adopt unless there are ties between alternative plans. If there is just one plan that is optimal from the group's point of view, the group members can find it out using we-mode we-reasoning (or team reasoning applied to plans) and extract their own part-plans out of it. What we get is a we-mode theory of planning which is top-down in contrast to bottom-up theories: In the we-mode theory, planning starts from group intention which is first specified to the level of detail that provides the roles for the members which are then able to infer and extract their sub-plans from the general group plan. In bottom-up theories, in contrast, the agents start from their individual intentions and then try to combine their sub-plans in a meshing way.

Utility functions measure the utility of outcomes, that is, possible states of the world, and the utility of plans derives from the expected outcomes that follow their execution. Note the difference between ordinary decision theory and decision-making based on planning: In the former, the agent selects between single actions.

The best action is the one that is expected to lead to the best outcome. In the latter, the agent selects between plans consisting of multiple actions. The best plan is the one that is expected to lead to the best outcome where the expected costs are taken into account. (In the case of resource-bounded agents like humans and groups of humans, the deliberation costs play a role: There must be a limit to the exploration in the infinite space of all possible plans.) The action to be performed is the first action of the best plan found and it may lead to a worse outcome than the best action, but its rationality derives from the rationality of the best plan.

Joint goals and other attitudes typically have a major effect on the evaluation of plans. If there is only one joint goal, it should filter out those plans that do not satisfy it. Then there might be other shared values or the agents may agree upon shared values, as in Bratman's case. For instance, the efficiency of the plan might be an important consideration favouring plans without redundant or otherwise inefficient steps. The plan with the highest utility value is then selected. Assume for the moment that there is only one plan with the highest utility value and it is fully specified, including the part actions of each of the agents. Applying the idea of team reasoning, the agents can then extract their part plans from the group plan. In principle, there is no need to negotiate or bargain about their individual parts, nor is there need to worry about meshing or potential conflicts because the plan is already fully specified in the group level. The individuals can then simply perform their parts.

Of course, this is a highly idealised situation. In reality, (i) the agents' part plans may be only partially specified, (ii) only agent roles may be specified but not which agent takes which role, (iii) there may be contingencies requiring re-planning during the execution, and (iv) there may be ties and no unique best plan. In case there are ties, those need to be broken. Oftentimes, direct communication is the best way to do it: Each agent may either suggest one plan over the others or ask the others which plan to adopt. In cases of disagreement, negotiation and persuasion may take place. Using language is not always necessary, because communication can be in terms of non-linguistic actions that are perceived by others: For example, if the plans can be distinguished by the first action, one of the agents may simply communicate the selection by initiating the first action in the plan. The others will then realise that the other plans are no longer compatible with the actions taken so far and can discard them. If the plans cannot be distinguished by the first action, the selection may be postponed to a later moment in which either the draw may have been resolved or one of the agents may then take initiative and communicate the selection of the plan.

The main difference to I-mode planning in our view is that I-mode planning (at least typically) starts from individuals' plans (or sub-plans), which are then combined together to form a group plan. For instance, Bratman's (2014) theory of shared planning seems to allow this kind of an idea and should probably be

classified as an I-mode planning theory. In our view it is I-mode because it seems to start with individuals' intentions that we do something (recall that these are to be understood as plans in Bratman's theory), and then proceeds, by negotiation, persuasion, bargaining, etc., to make these sub-plans mesh.

We do not know whether Bratman (2014) would insist that shared planning must always proceed in such a bottom-up order but clearly he at least allows that it can so proceed. This is clear from his discussions (e.g. Bratman 2014, 53-54) concerning some of his examples of going to New York City: He introduces the idea of meshing sub-plans in order to exclude cases in which you and I both intend to go to NYC but I intend to trick you into Amtrak train in spite of your intention to take the New Jersey local train. He introduces conditions to ensure that in addition to both of us intending to do something together we must intend to do it by way of sub-plans that mesh. It seems that Bratman's idea is rooted in bottom-up planning in which individuals first have their own plans concerning their activities and then negotiate with each other in order to make these plans mesh. He does not introduce a condition that would require that the sub-plans derive from a shared group plan, a condition that we think is essential for we-mode planning.

Let us clarify the difference further. At this point, one may ask how the group plan in the we-mode planning theory comes about if not by individuals' plans concerning joint activities. As we have said, a group cannot do anything without its members' doing something, so it may not be clear what we mean exactly by a group making a plan if it is not by individuals making plans and then trying to combine them. And if this is what is meant, the difference between I-mode planning and we-mode planning seems to vanish. Let us explain.

The idea of formation of group plans that we have in mind resembles Tuomela's idea of how other group attitudes or group properties come to be collectively accepted (see Tuomela 2013, 105): Individual group members make suggestions as to what they think the group should do and then the group members together make the decision following the group's decision-making procedures. So it might very well be that individuals have various ideas for group plans and then negotiate with others and maybe try to persuade them to adopt their plan just as in the I-mode case, but the crucial difference can be spelled out by recalling from Bratman (1987, 28-29) the distinction between two different notions of what a plan is. According to Bratman, plans can be thought of as recipes or they can be thought of as mental states. As Bratman says, if plans are understood as recipes, one can have a plan to roast a lamb without actually intending to roast a lamb. So having a plan means merely having a recipe or a procedure for doing something. Bratman prefers to think of plans as mental states involving commitments: In Bratman's I-mode planning theory, the individuals' plans concerning joint activities are to be understood to be intentions of the form "I intend that we J". To illustrate the we-mode planning theory, before the best plan is found and selected, the candidate

plans that are discussed should rather be thought of as recipes: The individuals suggest various multi-agent plans or procedures without yet being committed to them. To be sure, they have intentions but these are ordinary intentions, for instance, intentions like “I intend to present my suggestion for a plan that we J to the other group members and persuade them to adopt it”.

It is only when the group members collectively accept a plan and commit themselves to it that an intention to carry out a plan is created. Only then can we see the plans as mental states, or perhaps something akin in the case of groups. In Tuomela’s terminology, that intention is then a group intention (Tuomela 2013, 87) of the form “Our group will J” (“Our group intends to J”) or a joint intention (Tuomela 2013, 73) of the form “We will J” (“We intend to J”). From an individual’s point of view, these collective level intentions materialise as we-intentions (Tuomela 2013, 78) of the form “I we-intend to J with the others” and participation intentions of the form “I intend to do my part in our J-ing”. Thus, in the we-mode planning theory, there are no intentions of the form “I intend that we J” (which have been found problematic by some people). Instead there are intentions of the form “I we-intend to J with the others” (which have also been found problematic by some people, not necessarily the same people as the ones having a problem with “I intend that we J”).

As compared to I-mode planning, we-mode planning has certain anticipated benefits and also some potential downsides. First, it should be mentioned that it makes strong assumptions and is therefore not generally applicable. The main assumption is that it assumes a cooperative situation in which the agents are functioning in the we-mode. Therefore, they act on the basis of shared goals putting aside their private goals. This requires agreement on a shared evaluation function, but once such a function is established, there is reduced need for communication and bargaining in the we-mode approach because selections between alternative plans are based on the shared function and not on personal preferences.

In principle, we-mode guarantees the optimality of plans by definition because the best plan is always selected. However, this is not feasible in practice due to time constraints and possible changes during execution: It is not generally possible to evaluate all possible plans, because there are infinitely many of them, so some constraints are needed to restrict the set of plans. These problems of course affect I-mode reasoning as well, but in we-mode reasoning it would be more important that all the agents are aware of the constraints: If the agents are looking for the best plans in different search spaces, their results may be different and coordination may fail. In general, finding the group plan may be more complex than finding individual plans. This may be alleviated somewhat by doing it in stages as suggested above. Finally, shared evaluation function guides coordination in case of new decisions. What is crucial to notice is that even though also I-mode planning can make use of a shared evaluation function, it may still be subject to

the problems illustrated by the Hi-Lo case which is one of the game-theoretical examples discussed by Tuomela (2013, 11):

Hi-Lo	H	L
H	3,3	0,0
L	0,0	1,1

The Hi-Lo problem is a coordination problem in which all the agents share the same utility function, and there are several equilibria only one of which is Pareto-optimal. Such situations are problematic for traditional game theory equipped with a solution concept like the Nash equilibrium because it cannot rule out the suboptimal equilibria. This is known as the equilibrium selection problem. Team reasoning has no such problem in the case of Hi-Lo because it is based on the idea of group agent selecting between various outcomes based on its preferences: A group consisting of all players can choose the best outcome because the group can specify all of its members' strategies, which together determine the outcome. And what is best for the group, or preferred by the group, is in this situation trivial because everyone shares the same preferences. Team reasoning is basically a method of coordinating multi-agent actions in situations in which there is one unique optimal outcome and this is common knowledge among the agents. Individual players, on the other hand, can only choose their strategies. Even if they all know that they are trying to maximize the same utility function, the choice of a maximizing strategy is dependent on the choices of other agents. Because everyone is in the same position they are trapped in an indeterminate situation.

Let us consider whether the Hi-Lo might pose a problem for Bratman's (2014) theory. Will Bratmanian planning agents engaged in shared intentional activity be able to find their way out of the indeterminacy? At least initially it seems that they should. After all, they are not dependent on the individuals' intentions concerning only their own activities, but they, too, can in a sense take a group's point of view by employing the "I intend that we J" structure. What they are intending is not merely their own strategy but the whole group's. Thus they can avoid the problem illustrated by Tuomela (2013, 187, see also Hakli et al. 2010) in which two agents are trying to maximize group utility. In terms of group intentions, the reasoning goes as follows (all the premises are assumed to be common knowledge in the reasoning patterns):

1. We intend to maximize group utility.
2. Our choosing HiHi uniquely maximizes group utility.
3. Therefore, I will perform my component in HiHi, that is, Hi.

In terms of individuals' intentions, practical reasoning does not succeed because there is no unique maximizing strategy for individuals:

1. You intend to maximize group utility
2. I intend to maximize group utility.
3. If you choose Hi, my choosing Hi maximizes group utility.
4. If you choose Lo, my choosing Lo maximizes group utility.
5. Therefore, I will perform what?

In the case of Bratman (2014), a corresponding piece of practical reasoning should reach the correct conclusion:

1. You intend that we maximize group utility
2. I intend that we maximize group utility
3. Our choosing HiHi uniquely maximizes group utility.
4. Therefore, I will perform my component of HiHi, that is Hi.

This is essentially team reasoning, or at least it closely resembles team reasoning. But does Bratman endorse it? The problem is that we do not know. In his book (Bratman 2014) he does not give patterns of practical reasoning that would show how planning agents actually select their individual actions in cases like this one. Bratman does occasionally comment on team reasoning though, but in a somewhat critical tone (182, n. 19), sometimes contrasting it with his idea of shared deliberation (189, n. 2). He does not see a role for team reasoning in his planning theory but thinks that appealing to intentions that we J and shared deliberation can do the job of getting the individuals to realize the activities needed to satisfy the shared intention. However, without presenting any reasoning patterns it is hard to see how they do it but in trivial cases in which every action is specified in minute detail or everyone can continuously observe what others are doing and adjust their actions correspondingly. However, in real life cases of small scale modest sociality that Bratman is interested in, people often face non-trivial cases in which they will have to be able to select actions independently of each other on the basis of shared values and intentions.

This is precisely the purpose for which team reasoning is designed. Bratman's shared policies for weights and acceptances cannot do the job because they only provide the common ground for deliberation corresponding to shared beliefs and values (or preferences). As Bacharach's (1999) result shows, people may share

all their beliefs and values and still face problematic situations in which they will be unable to coordinate their actions even in simple cases like Hi-Lo in which intuitively there is an easy solution. Shared deliberation does not really help here either because, as Bratman (2014, 189, n. 2) says, shared deliberation is something that the agents do together as a shared intentional activity. Consider the following example.

Assume that we have a shared intention to paint a house tomorrow. Assume further that we have previously adopted a shared policy of giving more weight in our shared deliberation to painting in blue than to red and to using oil paint to latex paint. We do some shared deliberation and decide to paint the house red tomorrow. We agree that everyone buys a can of red paint and brings it to the house to be painted at 9:30 in the morning. For one reason or another we fail to make a decision concerning whether to use oil paint or latex paint. For instance, we may forget. We only realize this at 9:00 in the morning when each of us is at their favourite hardware store to buy a can of red paint. For one reason or another we cannot reach the others and each of us faces the decision whether to buy oil paint or latex paint. We have the shared policy of favouring oil paint to latex paint in our shared deliberation and we may assume that this is common knowledge, but it is not obvious whether that helps in this situation. Recall that we are not doing shared deliberation now, but each of us faces the decision alone. Clearly, this is a case that resembles Hi-Lo: There are better and worse ways of coordinating our actions and there is only one that is Pareto-optimal, namely the case in which both of us buy the oil paint. We cannot mix different types of paint and if we decide not to buy and go to the house we lose valuable painting time.

For team reasoners the solution is obvious: The group action in which everyone buys oil paint is best for the group, so everyone buys oil paint and brings it to the house at 9:30. For everything Bratman says in the book, it seems that Bratmanian planning agents cannot reach the same conclusion. They do have meshing subplans even though these plans are not fully specified, and all the conditions A-D in his compressed basic thesis (Bratman 2014, 103) seem to be satisfied. Clearly, they should be able to use the information that they share a policy of favouring oil paint over latex paint, but according to Bratman, these policies concern shared deliberation, and he explicitly says that shared deliberation differs from team reasoning in that it is something that we do together (Bratman 2014, 189, n. 2). Even if “doing together” is understood as including doing things alone but part of shared intentional activity, Bratman’s theory lacks a reasoning mechanism that the agents can use in situations like this. Thus, there is no guarantee that the agents will reach the conclusion that each should buy the red oil paint and they may not be able to satisfy condition E and fulfill their shared plan. Instead, they are forced to go to the house empty-handed and do some more shared deliberation.

The point is that Bratman talks only about shared deliberation but not about

individual deliberation in a group context. He does not seem to present a reasoning mechanism corresponding to team reasoning that can be used also in cases in which the agents have to make decisions alone based on shared values or policies. It seems that Bratmanian groups will have to agree on all the details of the plans together and then execute them precisely as agreed. Individual agents can refine their sub-plans only to the extent that their choices are independent of the choices of the other agents. As soon as there is something that has not been specified but which requires coordination of action, the agents must communicate in order to coordinate their choices even if it were obvious what the best choice is in light of the shared values and policies of the group. This gives a somewhat unrealistic picture of what joint action and shared agency is because people who are cooperating with others often face situations in which they will have to make decisions by themselves on the basis of shared values using common sense. Team reasoning is one way of specifying what that common sense is. The we-mode theory can easily incorporate team reasoning, and it would be good for an I-mode theory to adopt this bit from the we-mode theory or develop an alternative method that can deliver similar results.

If Bratman cannot say how agents come to reason from A-D to E, this seems to put his main claim under threat. In more detail, the argument is as follows:

1. Bratman claims in his continuity thesis that his theory provides an account of basic forms of sociality based on an underlying model of individual planning agency without the need to introduce fundamentally new elements like we-intentions or joint commitments.
2. To account for basic forms of sociality requires (among other things) showing how the intention targeted to realizing a social activity leads to, or at least tends to lead to, the realization of the social activity in question and thus to the satisfaction of the intention. In Bratman's terms, this amounts to showing how our shared intention to J leads to (or at least tends to lead to) our J:ing. (In terms of Bratman's compressed thesis, this amounts to showing how condition E tends to become satisfied whenever conditions A–D are satisfied.)
3. Bratman has shown how a shared intention to J leads to (or at least tends to lead to) our J:ing only in trivial cases where everyone can constantly communicate with each other and monitor each others' activities.
4. Basic forms of sociality include non-trivial cases in which it is not constantly possible to communicate with each other and monitor each other's activities.

5. Therefore, Bratman's theory has not accounted for basic forms of sociality and thus not demonstrated the plausibility of his continuity thesis.

Some justifications for the premisses might be in order: Premiss 1 comes directly from Bratman's book (see, e.g., page 4). Premiss 2 is crucial and we think that Bratman would accept it: In the case of individuals, Bratman finds it important to demonstrate what is the relation between intentions and intentional actions. Consider, for instance, the opening sentences in a previous article (Bratman 1984):

We do things intentionally, and we intend to do things. Our commonsense psychology uses the notion of intention to characterize both our actions and our mental states: I might intentionally start my car, and I might intend to start it. My intention to start it clearly does not guarantee that I do. But beyond that it is not obvious how these two phenomena are related. A central problem for a theory of intention is to provide a plausible account of this relation.

It seems plausible to assume that he would agree that it is equally important to show how shared intentions relate to shared agency. Just as individuals' intentions constrain the individuals' choices and coordinate her actions, also shared intentions constrain the group's choices and coordinate the actions of the group members. According to Bratman (2014, 91), "[s]hared intentions coordinate planning and action and frame relevant bargaining, all this in ways that tend to conform to characteristic norms of social rationality. And the claim is that this social functioning and social rationality will emerge from the individualistic structures described by the basic thesis". Also, Bratman (2014, 153) writes: "Within such modest sociality shared intention plays a characteristic explanatory role: relevant intentions of each work their way through, as intended, to shared action by way of relevant mutual responsiveness."

Premiss 3 is based on the fact that Bratman only says that shared intentions coordinate the planning of individual agents but does not tell how they do it. He does not provide schemas of practical inference and does not discuss problematic coordination cases like Hi-Lo in sufficient detail. Perhaps he thinks that these cases are unproblematic: As long as the subplans mesh and everyone aims at satisfying the shared goals (the conditions A-D are satisfied), then the agents will somehow manage to carry out the intention. As Bacharach's (1999) result shows, there is no obvious solution to such coordination problems. At least orthodox game theory has no good answer to such deceptively simple coordination cases as Hi-Lo in which one outcome is better than the rest. Bratman's ideas concerning shared deliberation and shared policies about weights do not seem to provide an answer, whereas the we-mode theory that relies on we-intentions and team reasoning does.

Premiss 4 is not based on Bratman's theory, but it seems plausible. Even in such simple cases as painting a house, agents may easily face choices in which the result will be better if they manage to coordinate their actions but they will have to make the decisions independently of each other. Such cases are ubiquitous in real life situations and therefore any theory that tries to account for basic forms of sociality should say something about them.

Bratman could extend his theory by allowing team reasoning but it is not clear whether it would be compatible with his continuity thesis or whether team reasoning should count as a fundamentally new element. Bratman does not explicitly endorse team reasoning but has a somewhat critical tone toward it. We think that it would not be impossible for him to incorporate team reasoning to his theory despite its relatively collectivistic background presuppositions: Even though the standard way of understanding team reasoning presupposes a notion of collective preferences (which Bratman might be reluctant to incorporate in his theory), there are alternatives that are based on a more individualistic notion of mutual advantage (see Sugden 2011). Moreover, Bratman already accepts the idea of shared valuing and shared policies, and therefore it might not be a big step to talk about shared preferences either. Indeed, it seems to us that Bratman comes very close to concepts like group preferences and group beliefs as understood by Raimo Tuomela (1992, 2007) and Margaret Gilbert (1989, 2001) in his treatment of shared commitments to weights (Bratman 2014, 134) and shared policies of acceptance (Bratman 2014, 147) even though he still seems reluctant to attribute such properties or attitudes as preferences or beliefs to groups.

Possibly Bratman's critical tone toward team reasoning stems from the fact that he is commenting on authors (Gold & Sugden 2007) who see the role of team reasoning very differently from the way Tuomela sees it and which we think is more in line with Bratman's thinking. Basically, Gold and Sugden see team reasoning as a mechanism by which individuals create group intentions, whereas Tuomela sees it as a mechanism by which individuals select their actions on the basis of existing group intentions (for elaboration, see Hakli forthcoming). The latter reading, we think, would be compatible with Bratman's account as well. Our suggestion is that just as Tuomela's we-mode theory would benefit from adding some elements of Bratman's planning theory, Bratman would do well to incorporate some elements from we-mode theory, in particular, team reasoning as understood by Tuomela into his planning theory. As we have agreed earlier in this paper, planning and team reasoning are but two instances of reasoning mechanisms that coordinate individual actions into more complex actions.

4 Conclusions

We have argued that Tuomela's theory of group agency is still incomplete because it lacks an important element, namely an account of planning. We have tried to sketch how to extend his theory by incorporating Bratman's idea of planning to the case of group agents. We have shown that even though Tuomela does not consider planning himself, his distinction between I-mode and we-mode provides a conceptual framework for distinguishing between different methods of planning as well. In particular, it has provided us means for characterising we-mode planning in which the planning proceeds in a top-down fashion from the group plan to individual sub-plans. We have contrasted we-mode planning to such cases that we call I-mode planning which start in a bottom-up fashion from individual plans that are then combined into a group plan by using negotiation and persuasion.

We have also argued that Bratman's account of shared agency seems to lack an important reasoning mechanism that enables the agents to coordinate their actions in pursuit of group's objectives. Such a mechanism is provided by Tuomela's account of we-mode we-reasoning, or team reasoning as it is more commonly known. Extending Bratman's theory by such a mechanism would yield an account of modest sociality that is somewhat less modest in that it would reduce the need for shared deliberation and allow coordination of actions in more realistic cases in which opportunities for communication are limited. We have argued that planning and team reasoning are closely related decision-making mechanisms that both share the general idea that an individual act is selected on the basis of it being a part of the best possible complex act. In the case of planning the complex act is a temporally extended sequence of acts and in the case of team reasoning the complex act is a collective act.

We think that it is possible to combine the main insights of Tuomela's we-mode approach and Bratman's planning approach into a fruitful synthesis that is necessary for understanding the nature of group agency, and we have taken the first steps to that direction by sketching the main features of an account of planning group agency.

References

Bacharach, M. (1999) Interactive team reasoning: A contribution to the theory of co-operation. *Research in Economics* 53, 117–147.

Bratman, Michael (1984) Two faces of intention. *The Philosophical Review* 93, 3, 375–405.

Bratman, Michael E. (1987) *Intention, Plans, and Practical Reason*. Harvard University Press.

- Bratman, Michael E. (2014) *Shared Agency: A Planning Theory of Acting Together*. Oxford University Press.
- Gilbert, Margaret (2001) Collective Preferences, Obligations, and Rational Choice, *Economics and Philosophy* 17, 109–119.
- Gold, N. and Sugden, R. (2007). Collective intentions and team agency. *Journal of Philosophy* 104, 109–137.
- Hakli, Raul (forthcoming) Team reasoning and collective intentionality. Paper presented at The Social Mind Workshop, Oslo, Norway, August 30, 2014.
- Hakli R., Miller, K., Tuomela, R. (2010) Two kinds of we-reasoning. *Economics and Philosophy* 26, 3, 291–320.
- List, Christian & Philip Pettit (2011) *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press.
- Pollock, John L. (1995) *Cognitive Carpentry: A Blueprint to How to Build a Person*. The MIT Press.
- Sugden, Robert (1993) Thinking as a team: Towards an explanation of non-selfish behavior. *Social Philosophy and Policy* 10, 69–89.
- Sugden, R. (2011). Mutual advantage, conventions and team reasoning. *International Review of Economics* 58, 9–20.
- Tuomela, Raimo (1992) Group beliefs, *Synthese* 91, 285–318.
- Tuomela, Raimo (2002) *The Philosophy of Social Practices: A Collective Acceptance View*. Cambridge University Press.
- Tuomela, Raimo (2007) *The Philosophy of Sociality: The Shared Point of View*. Oxford University Press.
- Tuomela, Raimo (2013) *Social Ontology: Collective Intentionality and Group Agents*. Oxford University Press.
- Weirich, Paul (2004) *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances*. Oxford University Press.
- Weirich, Paul (2009) *Collective Rationality: Equilibrium in Cooperative Games*. Oxford University Press.
- Woodard, Christopher (2008) *Reasons, Patterns, and Cooperation*, Routledge Studies in Ethics and Moral Theory, Routledge.
- Woodard, Christopher (2011) Rationality and the Unit of Action. *Review of Philosophy and Psychology* 2, 2, 261–277.