TEAM REASONING AND GROUP AGENCY

Raul Hakli
University of Helsinki
April 5, 2016

Introduction

My interest here is in philosophical theories of joint action, not in experiments or empirical findings.

But, empirical findings relevant! Theories should be consistent with them. (compare: whether Bratman's common knowledge requirement cognitively too demanding or not)

Also we can compare theories on the basis of how well they can explain the empirical findings (e.g. concerning how people coordinate).

Major debate between individualists and collectivists.

What are the collective building blocks of joint action?

Is it enough to use notions that reduce to individuals' intentional states or whether it is necessary to have some irreducibly collective intentional states.

Philosophically interesting as such, but may also be of interest to roboticists:

If we can decide on a ”correct” theory of joint action, then that may aid us in implementing robots that engage in joint action. (Probably not motor action but at least conceptual framework and maybe intentional-level reasoning.)

Based on yesterday's discussions: One important thing in joint action is coordination.

Today I will be talking about one particular kind of coordination.

Not so much on coordination of physical movements (like handing over items) when the partner's actions can be perceived, and there can be communication

Theoretical focus on cases where the actions are made independently of each other and communication is not possible. (like the meeting in Paris)

These cases have been studied in game theory, and famously Thomas Schelling suggested that we converge to focal points.

Theoretically focal points are problematic because there is no rational justification for them and they seem to be highly context-dependent (e.g. where to meet in Toulouse?)

This also means that it would be very difficult to implement a robot that would be able to coordinate based on focal points.

My focus here is in a class of coordination problems that seems much easier than those. In the coordination problems mentioned above, it is usually assumed that we have no preferences concerning where we meet: every place is equally good, as long as we both select the same one.

	A	B
A	$1, 1$	$0, 0$
B	$0, 0$	$1, 1$

Now I am assuming that one combination of actions is strictly better than the others! (Like in this Hi-Lo game.)

	H	L
H	$3, 3$	$0, 0$
L	$0, 0$	$1, 1$

And I am assuming that they agents have a shared goal so it is actually strictly better for all of them.

And I am assuming that everyone knows this, in fact I am assuming that this is all common knowledge.

For instance, we may have agreed to meet for lunch tomorrow but we didn't agree on the place but we both prefer the same one.

For human beings, it is very easy to coordinate our actions on the better option.

However, our best theoretical approach to rational multi-agent action and coordination, namely game theory, cannot recommend one over the other.

This is a big problem on a theoretical level.

Given that game theory is increasingly used in AI, MAS, and robotics, it may become a practical problem too: Robots built on the basis of this theory alone will not be able to coordinate their actions in a way that would be natural for humans. (consider e.g. handover tasks)

Human beings almost invariably manage to coordinate on the HH outcome.

However, there is no theoretical explanation for this.

Philosophers and economists working on game theory have been interested in theoretical models that could explain why H is selected and why it would be rational to select H instead of L.

Moreover, they are interested in how to model the reasoning that would lead to the selection of that action.

A theory that has had lots of interest recently is the theory of *team reasoning* (Bacharach, 1999; Sugden, 1993).

We know that people's behaviour in coordination cases matches well with the predictions of team reasoning.

We don't know whether this behaviour is really produced by team reasoning, or by any kind of explicit reasoning whatsoever.

Be that as it may, it may still be useful in implementing human-like behaviour in robots, especially if it can be formalised as a BDI-type of reasoning.

There are other theories that make similar predictions but they all have some problems.

The controversial point of team reasoning is that it is not individualistic. Seems to require group intentions.

However, some individualists argue that team reasoning can be captured by individuals' intentions.

I argue this is not the case: They can only capture some special limiting cases of team reasoning but not team reasoning in its full generality.

Then I will consider whether it can be done in Bratman's conceptual framework, using individuals' intentions concerning group's actions.

I argue this fails too, and conclude that we need something stronger (such as approaches of Tuomela, Searle, or Gilbert).

Human-Robot Joint Action

Human-Robot joint action requires coordination of action to reach common goals.

People seem to be quite good at coordinating their actions.

Theoretical work on coordination often done in game-theoretic terms.

How to model the practical reasoning leading to action coordination in robots?

How to do it in the standard BDI-framework of rational agency?

Bratman on joint action

Michael Bratman (1987) argued against BD theories of intentional action. His theory influential in the adoption of BDI. Bratman (2014) extends the theory to multi-agent case.

Shared intentions are individuals' intentions concerning shared activities, analysed in terms of individuals' attitudes

Individualism: continunity thesis: 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.

Basic building block: "I intend that we J". However, the idea of "intending that" controversial!

I-mode vs. we-mode

Raimo Tuomela (2007): Distinction between I-mode and we-mode action

The intuitive idea in we-mode:

- the group is seen as an agent that has attitudes of its own and can select between joint actions
- individuals do their parts as if they were the limbs of the larger agent

$\Rightarrow$ the idea of *group agents!*

Roughly, I-mode reasoning is based on individuals' attitudes, we-mode reasoning is based on the group's attitudes.

Also, in I-mode the agents are committed to themselves, but in we-mode there is collective commitment.

Close relation to team reasoning (Hakli et al., 2010).

Aim of practical reasoning

How to get from beliefs, desires, and future-directed intentions to intentions concerning actions?

In game-theoretic terms: How to get from preferences over outcomes to preferences over strategies?

Need to establish the agent's available choices and a preference ordering among them (if more than one choice available)

Team Reasoning

Team reasoning proposed as an alternative to game-theoretic models of decision-making.

- *Traditional game theory*: individuals select actions that lead to outcomes they prefer (given their expectations of the other agents' actions)
- *Team reasoning*: individuals select their part actions in the profiles of actions leading to outcomes preferred by the group (assuming that others do the same).

Team Reasoning

Different ways to frame a decision problem: ”group identification” (Bacharach, 2006) involves two steps:

- *Preference transformation*: from private preferences to group-directed preferences
- *Agency transformation*: not ”What should I do?”, but ”What should we do?” (involves transformation of reasoning: from best-reply reasoning to team reasoning)

Practical reasoning

From Bacharach (2006, p. 161) (also in Gold and Sugden 2007):

(1) I am a member of *S*.

(2) It is common knowledge in *S* that each member of *S* identifies with *S*.

(3) It is common knowledge in *S* that each member of *S* wants the value of *U* to be maximised.

(4) It is common knowledge in *S* that *A* uniquely maximises *U*.

I should choose my component of *A*.

Practical reasoning

The case of Hi-Lo (here we can assume that group utility = individual utility):

	H	L
H	$3, 3$	$0, 0$
L	$0, 0$	$1, 1$

From Hakli et al. (2010):

(1) We intend to maximize group utility

(2) Outcome *HH* uniquely maximizes group utility

Therefore, I will perform my component in *HH*, viz. *H*

Practical reasoning

The case of Hi-Lo:

	H	L
H	$3, 3$	$0, 0$
L	$0, 0$	$1, 1$

From Hakli et al. (2010):

(1) You and I intend to maximize group utility

(2) If you choose *H*, my choosing *H* maximizes group utility

(3) If you choose *L*, my choosing *L* maximizes group utility

---

Therefore, I will perform ?

Formulations like these have opened the door for individualistic critics, like Ludwig (forthcoming 2016), who correctly notes that nothing in these syllogisms requires there to be a group agent.

He argues that premiss (2) of the first argument conflicts with premisses (2) and (3) of the second.

In his view, the Hi-Lo can be solved using this syllogism:

(1) You and I intend to maximize group utility
(2) Outcome *HH* uniquely maximizes group utility

Therefore, I will perform my component in *HH*, viz. *H*

Ludwig is right that the formulations of the premisses are misleading: It should be made clear that maximization is a property of an agent's act: Premiss (2) should read "Our choosing *HH* uniquely maximizes group utility".

But then we will see that you cannot get to the conclusion without first deriving the intermediate conclusion "Therefore, we will choose *HH*." But this expresses a group intention!

Maybe we could replace that with another premiss "You and I maximize group utility only if you choose *H* and I choose *H*."

What about Bratman's theory? Can we do team resoning with intentions that we J?

(1) You intend that we maximize group utility

(2) I intend that we maximize group utility

(3) Our choosing *HH* uniquely maximizes group utility

Therefore, I will perform my component of *HH*, that is *H*.

What about Bratman's theory? Can we do team resoning with intentions that we J?

(1) You intend that we maximize group utility

(2) I intend that we maximize group utility

(3) Our choosing *HH* uniquely maximizes group utility

Therefore, I will perform my component of *HH*, that is *H*.

This piece of reasoning is fine, but it is not team reasoning, rather it is ordinary means-end reasoning!

Premiss 3 states a necessary condition: We maximize group utility *only if* you perform your component of *HH* and I perform mine).

Analysis of the problem

The problem with all of the formulations above is that they mix intentional attitudes and game-theoretic or decision-theoretic concepts. Agents do not *intend* to maximize utility functions! Utility-maximisation is a theoretician's way of modelling decision-making.

Agents intend to satisfy their goals. The idea of maximization is included in our concept of rational agency. When an agent can choose between better or worse ways to satisfy its intentions, rationality demands it to choose the better way.

The syllogisms have to be modified accordingly. Once we do that, we will see that team reasoning requires agency at the group level.

Maximization and agency

Quote from Aristotle (*Nichomachean Ethics*, book 3, 1112b11):

> *We deliberate not about ends but about means. For a doctor does not deliberate whether he shall heal, nor an orator whether he shall persuade, nor a statesman whether he shall produce law and order, nor does any one else deliberate about his end. They assume the end and consider how and by what means it is to be attained; and if it seems to be produced by several means they consider by which it is most easily and best produced [...]*

Back to Hi-Lo

	H	L
H	$3, 3$	$0, 0$
L	$0, 0$	$1, 1$

A more correct formulation of team reasoning in the Hi-Lo case is the following:

(1) We intend to J (group intention)

(2) We J just in case we select *HH* or *LL*

(3) We prefer *HH* over *LL*

Therefore, we will select *HH* (group intention)

Therefore, I will perform my component in *HH*, viz. *H* (we-intention)

Back to Hi-Lo

	H	L
H	$3, 3$	$0, 0$
L	$0, 0$	$1, 1$

A more correct formulation of team reasoning in the Hi-Lo case is the following: Options for group:

(1) We intend to J {HH,HL,LH,LL}

(2) We J just in case we select *HH* or *LL* {HH,LL}

(3) We prefer *HH* over *LL* {HH}

Therefore, we will select *HH*

Therefore, I will perform my component in *HH*, viz. *H*

A new attempt at Hi-Lo

	H	L
H	$3, 3$	$0, 0$
L	$0, 0$	$1, 1$

Could this be done with Bratman's (2014) shared intentions?

(1) You and I intend that we J {HH,HL,LH,LL}

(2) We J just in case we select *HH* or *LL*

(3) We prefer *HH* over *LL*

Therefore, ?

A new attempt at Hi-Lo

	H	L
H	$3, 3$	$0, 0$
L	$0, 0$	$1, 1$

Could this be done with Bratman's (2014) shared intentions?

(1) You and I intend that we J {HH,HL,LH,LL}

(2) We J just in case we select *HH* or *LL*

(3) We prefer *HH* over *LL*

Therefore, ?

We cannot select *HH* because there is no "we" that intends and can select between options. The only agents ("intenders") are you and I.

Back to Hi-Lo

	H	L
H	$3, 3$	$0, 0$
L	$0, 0$	$1, 1$

So the formulation should rather be this:

(1) You and I intend that we J

(2) We J just in case (you select *H* and I select *H*) or (I select *L* and you select *L*)

(3) We prefer that (you select *H* and I select *H*) over that (I select *L* and you select *L*)

Therefore, ?

Back to Hi-Lo

	H	L
H	$3, 3$	$0, 0$
L	$0, 0$	$1, 1$

So the formulation should rather be this:

(1) You and I intend that we J
(2) We J just in case (you select *H* and I select *H*) or (I select *L* and you select *L*)
(3) We prefer that (you select *H* and I select *H*) over that (I select *L* and you select *L*)

Therefore, ?

Neither of us can conclude selection of *H* because there is no ”we” that selects anything, only you and I, but each of us can only select between strategies, not between outcomes.

There does not seem to be a way to restrict the options of the agents to only one option…

…nor a way to establish a preference between the options available to the agents…

without taking the idea of maximization into the scope of the intention.

But then the reasoning is no longer team reasoning, but instead reasoning about a necessary means to an end.

This is a special case and not nearly as useful form of reasoning because usually there are more than one way to satisfy an intention.

Conclusions

In general, understanding rational agency requires both Bratman-type planning and Tuomela-type we-mode joint action involving team reasoning (Hakli and Mäkelä, 2016).

Team reasoning involves the idea of a group agent with (irreducible) group intentions.

The idea of maximization should not be understood as being in the content of an intention, rather it is in our concept of rational agent that selects between better or worse ways to satisfy its intentions.

Hence, intentional-level conceptualisation of team reasoning requires group intentions and we-intentions (Tuomela) instead of mere individuals' intentions (Ludwig) or intentions that we J (Bratman).

Implications for Robotics

Group agency relevant also for robotics, however, it does not mean that we will have to implement new agents in addition to individual agents: We only need to recognise that individuals may conceptualise and reason in terms of group agents and group attitudes.

Hence, if we want to implement robots that do explicit reasoning about intentional states, we will have to include also collective intentional states.

This is required in order for robots to be able to do practical reasoning as group members and also in order for them to be able to do theoretical reasoning about what human beings do when they are acting as group members.

References I

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

Bacharach, M. (2006). *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton University Press.

Bratman, M. (2014). *Shared Agency: A Planning Theory of Acting Together*. Oxford University Press.

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

Gold, N. and Sugden, R. (2007). Collective intentions and team agency. *Journal of Philosophy*, 104:109–137.

References II

Hakli, R. and Mäkelä, P. (2016). Planning in the we-mode. In Preyer, G. and Peter, G., editors, *Critical Essays on the Philosophy of Raimo Tuomela with His Responses*, Forthcoming in Studies in the Philosophy of Sociality. Springer.

Hakli, R., Miller, K., and Tuomela, R. (2010). Two kinds of we-reasoning. *Economics and Philosophy*, 26:291–320.

Ludwig, K. (2016). Methodological individualism, the we-mode, and team reasoning. In Preyer, G. and Peter, G., editors, *Critical Essays on the Philosophy of Raimo Tuomela with His Responses*, Forthcoming in Studies in the Philosophy of Sociality. Springer.

References III

Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfish behavior. *Social Philosophy and Policy*, 10:69–89.

Tuomela, R. (2007). *The Philosophy of Sociality*. Oxford University Press, USA.

Tuomela, R. (2013). *Social Ontology: Collective Intentionality and Group Agents*. Oxford University Press.