

Collective Reasoning and Commitment

Abstract

Proposed for Collective Intentionality Conference VI

Introduction

The paper is divided into three sections. First, I distinguish between two types of collective practical reasoning: collective reasoning *prior* to the formation of the we-intention, and collective reasoning *after* the formation of the we-intention. In the second section I discuss the latter case and argue that it involves social and collective *commitments*. In this section I also study the roles of these commitments in collective practical reasoning, and I apply the analysis of commitment to the problems of “de-conditionalisation” and infinite regress, e.g. in coordination problems. In the final section I return to the distinction of the first section, and study the pre-we-intention case in terms of practical reasoning models.

1. Distinction between two types of collective reasoning

I differentiate collective practical reasoning *prior* to the we-intention from *post-we-intention* collective reasoning. In my paper I compare and analyze these two types of “team reasoning”. Briefly, the former aims at a decision *what* we will do, and the latter is about *how* we will do it. The former aims at the formation of a we-intention, while the latter concerns the satisfaction of the we-intention. To account for the former is to account for the history of the we-intention, while the latter is concerned with its future. For example, when encountered with collective action dilemmas like PD, Hi-Lo or Chicken, the parties can collectively reason how to choose a collectively acceptable outcome. This is pre-we-intention reasoning. In the post-we-intention case the we-intention is presupposed, and the parties are reasoning what it entails for them jointly and for each of them individually. The post-we-intention cases have been widely studied in distributed artificial intelligence, often inspired by M. Bratman’s work, whereas e.g. N. Gold and R. Sugden¹ are mainly concerned with pre-we-intention cases.

2. Post-we-intention collective reasoning and commitment

Basically, I will account for collective commitment in terms of a notion of *joint persistent goal*. Intention implies persistence in attempts to act.² Intending is being committed to acting, but the intention to act does not imply that the intention itself cannot be freely abandoned, for whatever reason or for no reason at all.

¹ “Collective Intentions and Team Agency”, *The Journal of Philosophy*, 2007, CIV, 109-137

However, if the agent has entered a *collective* commitment, he is not free to drop his intention as he pleases. I argue for an account according to which, firstly: there is no collective commitment unless the parties agree on what they are committed to. Secondly: if there is a proper agreement backing the commitment, it involves obligations and rights for the parties. Thirdly: a collective commitment to the content of the agreement emerges if and only if the parties also intend to satisfy these obligations, including obligations not to violate the rights of the other parties. Fourthly: a collective commitment has proper parts, and it entails various commitments for the individual members. These commitments form an ordered structure, and the relations of dependence between the sub-commitments can be very intricate. For example: we intend to cook dinner together. This we-intention *commits* us to acting, it has “gaps” to be filled later, typically represented by conditional sub-intentions. Our post-we-intention reasoning “tracks” these sub-commitments within conditions of mutual awareness (mutual knowledge or belief), and aims at de-conditionalising them.

3. Practical reasoning aiming at the formation of we-intentions

Pre-we-intention collective reasoning can involve various commitments: commitment to seek *some* solution, commitment to follow a decision procedure (e.g. a majority vote), and commitment to accept what will turn out to be the result of the procedure. Or, in less regimented cases, it is just concerned with whether to do this or that together, or whether to do anything at all together (e.g. whether to cook dinner together, to go to a restaurant together, or to go our own ways). Sometimes mere desires “spontaneously” lead to we-intentions in circumstances of mutual awareness.

In this final section of the paper I apply the above distinctions to some practical reasoning schemas, but I will begin with a discussion of an alleged problem that some accounts of collective intentions face: “a general problem how to differentiate collective intentions from the mutually-consistent individual intentions that lie behind Nash equilibrium behaviour in games” (Gold and Sugden 2007, p.109).

They argue that the “the core analyses provided by Tuomela and Miller, Searle, and Bratman seem to imply that

(1) *all Nash equilibrium situations are instances of collective intentions*” (p. 110),

where “instances of collective intentions” means behaviour that results from collective intentions.

As it stands, (1) says that ‘whenever there is Nash equilibrium behaviour, it has resulted from collective intentions’, i.e. it says that ‘there is no Nash equilibrium behaviour without the behaviour being collectively intended.’ This so obviously false that if indeed the analyses of collective intentions would “imply” it, they would be equally false, as well. – The converse of (1), viz.

(2) *all instances of collective intentions are Nash equilibrium situations,*

² This kind of feature has been emphasized by M. Bratman, cf. e.g. 1999, *Faces of Intention. Selected Essays on Intention and Agency*, and Cohen and Levesque (in many papers since 1990).

would be a more interesting thesis, and not so obviously false, and in the paper I discuss some additional constraints under which it could be true. I will also argue that one of the above accounts³ does not “imply” either of the theses (1) or (2), and therefore, be it mistaken or not, it is not mistaken on those grounds.

The account by Gold and Sugden provides a novel and fruitful approach to team reasoning and it shows how we-intentions can come about. I discuss three simple practical reasoning schemas in the first person singular. The conclusion of the first schema says what I as a group member *have reason to do*, the conclusion of the second what I as a group member *ought to do*, and the third what I in fact *intend to do*. Briefly, I argue that in so far as the last, viz. the strongest schema is to yield my we-intention as its conclusion, then that we-intention has to be contained already in the premises, and therefore this schema cannot serve as an account of the process that has a we-intention as a result. It is not a pre-we-intention schema. The two weaker and genuine *pre-we-intention* schemas *can* lead to the formation of the we-intention, but they need not.

³ Tuomela and Miller, “We-intentions,” *Philosophical Studies*, LIII (1988)