I investigate the apparent tension between the idea that collective intentions are the result of team reasoning and the idea that there can be spontaneous collective intentions. This raises a more general question about the relationship between reasoning and spontaneous intentions, including in the individual case. I show that the tension need not arise in accounts that separate intentions from intentionality, as they can deny spontaneous intentions while retaining spontaneous intentionality in both the individual and the collective case. However, individual reasoning is a special case in the team reasoning model and spontaneous individual intentions are plausible, so it would be advantageous for team reasoning theorists to be able to account for spontaneous intentions in the collective case. In order to do this, we need to show how spontaneous intentions are compatible with reasoning. I consider how reasoning is understood in philosophy, economics, and cognitive science, and I show how spontaneous collective intentions can be reconciled on at least some accounts of what it is to do reasoning, which are compatible with the way "reasoning" is used in the team reasoning literature. I argue that we should think of team reasoning as a “computational-level model”, as used in cognitive science. I draw on research from philosophy of computation, and show how, on some theories, the view of reasoning as computation has sympathetic implications for theories of group agency, as it would allow that groups can be reasoners.

Acknowledgments

Work on this paper was supported by funding from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 283849.
intentions et intentionnalité, car il devient possible de rejeter l’existence d’intentions spontanées, tout en conservant l’existence d’une intentionnalité spontanée, et ce dans les cas individuels et collectifs. Néanmoins, dans la mesure où le raisonnement individuel est un cas particulier du modèle de raisonnement en équipe, et que l’existence d’intentions individuelles spontanées est plausible, il semble souhaitable que les théoriciens du raisonnement en équipe soit également en mesure d’expliquer les intentions spontanées dans le cas collectif. Ceci nécessite de montrer que les intentions spontanées peuvent être compatibles avec le raisonnement. J’examine les différentes interprétations de l’idée de « raisonnement » en philosophie, économie, et science cognitive, et montre comment les intentions collectives spontanées peuvent être réconciliées avec certaines de ces interprétations, interprétations qui sont compatibles avec la manière dont le raisonnement est appréhendé dans la littérature sur le raisonnement en équipe. Je propose de comprendre le raisonnement en équipe comme un « modèle computationnel », tel qu’utilisé en science cognitive, et montre à partir de certaines théories en philosophie de la computation que cette approche présente des implications favorables aux théories d’agentivité collective, car elle permettrait que les groupes puissent être des raisonneurs.


1. Introduction

When people act together, we say that they have joint agency and collective intentions. For instance, when two people play a duet, one person plays the cello part and the other plays the piano part, but they jointly play the piece. We say that they have a collective intention to play the duet as well as individual intentions about their participation. So there is a sense in which the cellist may say “we intend to play a duet”, as well as a sense in which she may say “I intend to play my cello part”. Analyses of collective intentions aim to identify the features that distinguish the joint action case from the case in which the cellist and the pianist each individually practise their parts in next door rooms and just happen to play them synchronously.

One difference between the joint and the individual case is that a joint action involves cooperation. In game theory, team reasoning has been proposed as a mode of reasoning that people use when they are cooperating or coordinating. In standard game theory, each individual asks separately “what do I want to achieve and what should I do to achieve it?” Team reasoning extends the syntax of game theory, to allow players to ask “what do we want to achieve and what should I do to play my part in achieving it?”. This may lead to cooperative actions. Gold and Sugden [2007a] have argued that collective intentions are the intentions that result from team reasoning and, therefore, that the standard analyses which start with the intentions themselves have already missed what it is that makes collective intentions special.

*REP* 128 (3) mai-juin 2018
In response to this argument, Tuomela [2009] claims that the idea that team reasoning leads to collective intentions is incompatible with the existence of spontaneous collective intentions (see also Schweikard & Schmid [2013]). Call this the “spontaneous collective intention critique”. As Tuomela ([2009], p. 299) puts it, “there seem to be spontaneously formed intentions... in which no actual reasoning takes place” (Tuomela, [2009], p. 299). For Tuomela, “reasoning” refers to a conscious process and he imputes this definition to team reasoning theorists. However, it is not at all clear that this definition is shared by team reasoning theorists, especially considering the theory’s origin in economics. In order to judge whether team reasoning can accommodate spontaneous collective intentions, it would be charitable to consider how team reasoning theorists understand reasoning – at least as a first pass – and then to consider whether there is any other conception of reasoning that will fit the bill. That is the approach I will take in this paper.

I will start by explaining how team reasoning is supposed to be connected to collective intentions. Then I consider how important it is for a theory of collective intentionality to be able to accommodate spontaneous collective intentions. On some accounts, one might be able to deny that spontaneous collective intentions exist. However, I will suggest that this response is not available to team reasoning theorists. Therefore I explore the conception of reasoning implicit in team reasoning and show the conditions under which it is compatible with spontaneous collective intentions.

2. Team reasoning and collective intentions

The theory of team reasoning was developed separately by Sugden [1993, 2003] and Bacharach [1999, 2006]. Its development was motivated by games that are puzzling for orthodox game theory because they have an arguably rational solution, which a substantial number of people play in real life, whose play game theory cannot explain or predict.

One of these games is Hi-Lo, a version of which is shown in Figure 1. There are two pure-strategy Nash equilibria, \((high, high)\) and \((low, low)\), and \((high, high)\) is strictly better than \((low, low)\) for both players. It seems clear that the two players should each play \(high\). However standard game theory cannot recommend that. A Nash equilibrium involves a player maximizing her payoff given what the other player is doing. Thus it can recommend that, if a player expects the other player to play \(high\), then it is rational for her to play \(high\). However, if she expects the other player to play \(low\), then it is
rational for her to play low. What it is rational for Player 1 to do is conditional on what Player 2 does, and standard game theory gives her no reason to expect Player 2 to play high rather than low.¹

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Player 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>2,2</td>
</tr>
<tr>
<td>low</td>
<td>0,0</td>
</tr>
</tbody>
</table>

Figure 1: Hi-Lo.

In standard game theoretic reasoning, an individual player asks “what do I want to achieve and what should I do to achieve it (given my beliefs about what other players will do)?” The answer is a complete strategy, roughly speaking a contingency plan, which she then carries out. Team reasoning extends the syntax of game theory, to allow players to ask “what do we want to achieve and what should I do to play my part in achieving it?”, thereby choosing to do their parts in the best profile of actions for the team. Then each individual can reach the conclusion that she should choose her component of that profile. This is analogous to individual action over time, where a self at a time makes a plan of action to be carried out by her current and future selves, and then the current self plays her part (Gold [2013]; Gold, in press [2018]).

We say that an individual “identifies” with a group if she conceives of that group as a unit of agency, acting as a single entity in pursuit of some single objective and considers herself to be a part of that group. In the Hi-Lo game it is clear that (high, high) would be the unique profile of actions that maximizes the team payoff function so, in the simplest case, if there is common knowledge that each member group identifies and common knowledge that each member aims to maximize the team payoff function, then each can reason that she should play high, as a part of playing (high, high). This is similar to individual reasoning, except that the subparts of the team agent are dispersed in space, rather than in time. Hence team reasoning can provide an explanation of rational coordination and cooperation (Bacharach [2006]; Gold and Sugden [2007b]).

Gold and Sugden [2007a] have also argued that team reasoning is involved in collective intentions. We can identify three types of analysis of collective intentions. First, they can be understood as a mental state which guides behaviour, the distinctive attitude of an individual participant in the joint action. This is the stance taken in the work of Searle [1990a] or Tuomela and Miller [1988], who refer to them as “we-intentions” or “joint intentions”. Alternatively, following Bratman [1992], a collective, or “shared” intention

¹ For fuller statements of this argument, see Hodgson [1967], Hurley [1989], Sugden [1993], or Bacharach [2006].
can be considered a state of affairs, which consists primarily of the attitudes of each individual participant and their interrelations. Finally, following Gilbert [1989, 2008, 2009], collective intentions can be understood as a type of “joint commitment” held by a group, or “plural subject”, which is committed to intend as a body, and therefore the collective intention is separate from the individuals’ contributory personal intentions. All three types of analysis start with the collective intention and put it in the context of the surrounding individual intentions and knowledge conditions; in particular the accounts of joint intentions and shared intentions define the collective intention in terms of the individual intentions and the knowledge states of the participating individuals, among other criteria.

Gold and Sugden [2007a] argue that collective or joint intentions, understood as mental states held by individuals, are distinguished from individual intentions by reference to the unit of agency in the reasoning that led to the formation of the intention. If joint intentions are the intentions that result from team reasoning, then analyses of joint intentions that start with the intentions themselves have already missed what it is about them that makes them special, as that was upstream of the intention. Gold and Sugden also suggest that a Bratman-style analysis, of shared intentions as states of affairs, could be considered an analysis of our disposition to reason and act as a member of a group, in relation to the objective of executing some broadly-defined joint activity. So they identify Bratman’s “shared intentions” with high-level strategic intentions, which set the framework within which subsequent tactical reasoning takes place. These strategic intentions guide the practical reasoning that leads to low-level tactical intentions, which are the individual’s mental states. Gilbert’s vision, of a plural subject making joint commitments, can be understood as corresponding to the reasoning schema that Gold and Sugden [2007a, 2007b] call team reasoning “from a group viewpoint”, in combination with the idea that group identification should be interpreted as an act of commitment.

3. The spontaneous collective intention critique

One worry is whether Gold and Sugden’s [2007a] analysis of collective intentions can account for spontaneous collective intentions (Tuomela [2009]; Schweikard & Schmid [2013]). As Tuomela [2009, p. 299] puts it, “there seem to be spontaneously formed intentions... in which no actual reasoning takes place” (Tuomela [2009], p. 299). It seems clear that Tuomela’s critique refers to spontaneous mental states, as per his own theory of collective intentions. For example, in the same way that it may spontaneously come to mind that “I will stand up and applaud” after the concert, it might spontaneously come to mind that “We will give a standing ovation” after a stupendously good concert.

Spontaneous collective intentions are less plausible than spontaneous individual intentions. On some accounts there may be no such thing as a
spontaneous collective intention, either because there are no spontaneous collective intentions, or because there are no spontaneous intentions full stop. In either of these cases, the spontaneous collective intention critique will have no bite.

The discussion of collective intentionality often involves examples where there would have been prior discussion of the cooperative activity, such as going on a trip together or painting a house, so we would not expect the collective intentions that occur in these cases to arise spontaneously. In Gilbert's [1989] account, if joint commitments involve explicit communication between the parties, then there would not be spontaneous collective intentions. Common knowledge requirements, of the intentions and beliefs of the other parties, may also make it unlikely that collective intentions could exist without prior discussion.

But these requirements jibe with some examples where cooperative activity emerges organically, without discussion – for example strangers joining together to push a car, or someone spontaneously joining in with the harmony of a song or piece of music – and where there does seem to be scope for spontaneous collective intentions. In addition, there are cases where a group collaborates in improvised performance – jazz, dance, and theatre are obvious examples, but even conversations arguably have improvised elements – and spontaneous collective intentions are potentially an issue here (Preston [2013]). We might spontaneously intend to perform a particular dance movement as a part of our improvisation. Even Gilbert [2008, 2009] is clear that her account also includes tacit commitments that arise without communication, which leaves open the door to spontaneous collective intentions.

On some accounts, there will be no such thing as a spontaneous intention, in which case it will be no problem if there cannot be spontaneous collective intentions. For instance, take Bratman's [1987] account of intentionality. For him, an intention is a planning state. In a later work, he elaborates on this, saying that a planning agent who “can settle in advance on partial plans of action, fill them in as time goes by and as need be, and follow through in the normal course of events, is thereby in a better position to satisfy needs for coordination given basic cognitive limits” (Bratman [2000]). This does not seem to leave a role for spontaneous intentions. Indeed, Stoutland [2002] has complained that Bratman's account makes spontaneous action “both secondary and obscure”. Bratman's account has also been accused of being unable to explain the intentions that we form immediately before acting, which therefore will not set the stage for further planning, such as when someone holds out a plate of cookies to me and, at that point, I form the intention to take one and immediately act on it (Velleman [2007]).

The answer, for Bratman, is that spontaneous actions simply do not involve an intention. We can identify two reasons for this. First, he says that many spontaneous actions are “automatic and unreflective” [1987, p. 126]. In other words, they are not intentional actions. Second, his theory allows that spontaneous actions can be intentional but not intended, so long as they are backed by a standing intention or fall within the “motivational potential” of an intention (Bratman [1984]). These actions would still be purposive actions, as they are voluntary, under the agent's control, and
involve the pursuit of goals in the light of representations of the world. However, the use of “intention” would be saved for a particular planning state. Thus Bratman does not need to grant the existence of spontaneous individual intentions.

Given that Bratman does not need to grant spontaneous individual intentions, there is no need for him to grant spontaneous collective intentions either. Bratman’s own account of shared intentions is compatible with this idea. They are a part of shared cooperative activity, which Bratman envisions as an activity that takes place over time (Bratman [1992]). The shared intention is a standing intention or planning state that covers the period of activity and that includes a commitment to the activity, a commitment to supporting the other player and helping her play her role, and a mutual responsiveness to the intentions and actions of the other. He illustrates mutual responsiveness with the example of a duet, where each singer allows elements of her singing – for instance when and how to come in – to be guided by the singing of the other. In the same way that Bratman allows that individual actions can be intentional but not intended, he could allow there are spontaneous collective actions that involve collective intentionality, but not spontaneous shared intentions. We might think of improvisation and spontaneous collective activity as being kinds of intentional action that are backed up by standing collective intentions, but not as involving the formation of spontaneous collective intentions (Preston [2013]).

Team reasoning theorists could accept Bratman’s position that there are no spontaneous high-level strategic collective intentions, which set the framework within which subsequent tactical reasoning takes place (which is not to imply that they would necessarily want to do this). Nevertheless, they might wish to accommodate spontaneous low-level tactical collective intentions. It seems plausible that there are spontaneous individual intentions, in the sense of mental states that come to mind without conscious reasoning. If we grant this, then the relationship between reasoning and spontaneous intentions will be a problem in the individual case, which needs to be solved. Gold and Sugden’s [2007a] argument strategy is that the same concept of intention and agency that we use in the individual case extends naturally to cover the collective intentions that occur during cooperative activity. (Other users of this strategy, who come up with different answers, include Pettit [2003], Tollefsen [2002], and Velleman [1997]). Individual reasoning can be thought of as a special case of team reasoning when a team only has one member (Gold & Sugden [2007a]; Gold & Sugden [2007b]). We can also model the individual as a team consisting of the self at different times (Gold [2013]; Gold, in press [2018]). Since there can be seemingly spontaneous individual intentions, if collective intentions are the same type of thing as individual ones, then the account of collective intentions must leave room for spontaneous collective intentions.

The spontaneous collective intention critique raises the general question of what it is to be a mode of reasoning and what, if any, accounts of reasoning can reconcile the apparent tension between the view that intentions are the result of reasoning and the possibility that there can be spontaneous intentions. The reason that Bratman’s position is compatible with there being conscious reasoning before all intentions is that he is willing to...
divorce intentions from intentionality, allowing that outcomes may be brought about intentionally without being intended. Other philosophers have disagreed, arguing that all intentional action involves a proximate intention. Amongst those who identify acting intentionally with acting on an intention, there is precedent for the idea that intentions do not require conscious reasoning (Goldman [1970]; Brand [1984]; Mele and Moser [1994]). However, they do not give an account of how spontaneous intentions are connected to reasoning. Providing such an account will be one contribution of this paper. I will consider how team reasoning is related to collective intentions, but the solution will also explain how individual reasoning relates to individual intentions.

4. Game theoretic reasoning is not conscious deliberation

Philosophers often characterize intentions as the conclusions of, or downstream from the conclusions of, practical reasoning (Harman [1976]; Ross [2009]). For philosophers working on practical reasoning, the norm is to regard reasoning as involving conscious deliberation, either weighing reasons or making inferences from beliefs and desires and, in both cases, then drawing conclusions about what to do. (Wallace [2014], gives an overview.) For example, Pettit [2007] characterises reasoning as done at the intentional, personal level. He draws on the personal-subpersonal distinction of Daniel Dennett [1969, p. 93], who contrasts personal explanation, “the explanatory level of people and their sensations and activities”, with “the sub-personal level of brains and events in the nervous system”. For Pettit, reasoning involves “a form of control in which I intentionally pursue the satisfaction of rational desiderata, rather than merely relying on my non-intentional processing” [2007, p. 503].

This conception of reasoning as conscious deliberation seems to be driving Tuomela’s [2009] critique. That interpretation is supported by Tuomela’s more general criticisms of team reasoning, namely that some cases of joint action “basically are not cases for the exercise of “Machiavellian intelligence” and strategic thinking” and that “the common-sense goals that people collectively intentionally adopt and purport to achieve are typically more specific goals than maximizing expected utility” (Tuomela [2009], p. 298). Hence Tuomela seems to attribute the conscious maximisation of expected utility to the team reasoning agent.

However, none of these problems would be entailed by the standard positions of game theorists. Team reasoning was developed as an extension of game theory. Bacharach [1999, 2006] presented it using the framework of expected utility theory, Sugden [2003] presented it using formal logic, and Gold and Sugden [2007a, 2007b] represented it using schemata of practical reasoning, where conclusions about what actions should be taken are inferred from premises about the decision environment and what agents are
seeking to achieve. It is possible to use these as alternative formalizations of the same account because, under certain conditions, it is possible to represent expected utility theory as modus ponens reasoning.

Expected utility theory is an instrumental theory of rationality and behaviour, modelling agents who choose so as to achieve their goals. The basic building block of expected utility theory is a preference ordering. The remarkable representation theorem of von Neumann and Morgenstern [1947] shows that, given some arguably unobjectionable constraints on a preference ordering, it is possible to represent that ordering as a cardinal utility function. Numbers can be assigned such that, if outcome $x$ is ranked higher than outcome $y$ in the preference ordering, then $u(x)$ is greater than $u(y)$. Then the decisions of an instrumentally rational agent can be represented as a maximisation problem. But this is the theorist’s framework for modelling the agent, it is not a process that is attributed to the agent. What the agents of game theory are seeking to achieve is simply their most preferred outcomes. There are no presumptions about what these outcomes are, why they are preferred or how the agents describe the outcomes to themselves. Indeed one criticism of the theory is that it is insensitive to the way that the agent describes the options (Schick [1997]).

The terminology of “maximizing expected utility” is that of the modeller. There is no implication that the agent describes herself in this way and she most likely expresses her goals to herself in a much more specific manner. When Tuomela claims that maximizing collective utility “is a highly general goal that seems not to be applicable to all contexts where people intend collectively and cooperatively – think of walking together or singing a song together”, he is right that it is highly general at the level of the theorist, but the theorist and the agent need not have a shared terminology, hence it is wrong to think that this implies that the agent’s description of what she is doing is also highly general.

Most game theorists make no commitments about actual processes of reasoning, they model agents who act as if they follow the specified decision procedure. In fact, agents may choose according to expected utility theory because they have played the game before and “learnt” the correct solution, where learning can include imitation, or basic trial and error learning which does not require a great degree of cognitive sophistication. But this does not necessarily imply a retreat to a purely descriptive model: the model is one of rational behaviour, not of correct reasoning. Game theorists also make no commitments about what is being ordered, or how to interpret the preferences and probabilities in their theories. One standard line in economics is that preferences simply describe behavioural dispositions, another is that preferences are intimately related to the welfare of the agent; both of

---

2. Although Von Neumann and Morgenstern proved their theorem in the context of game theory and are therefore most strongly associated with expected utility theory, as used by game theorists, it should be noted that they were not the only, or even the first, theorists to provide representation theorems. For instance, an earlier representation theorem was proved by Frank Ramsey [1926] and an influential later theorem is that of Leonard Savage [1954]; both of these derive probabilities as well as utilities.

3. For an excellent and wide-ranging discussion of “rationality” in economics, see Hammond [1997].
these can be disputed (see Hausman [2008]; Hausman [2012]; Dietrich & List [2016]). These questions about the primitives of the theory are orthogonal to the argument of this paper, which is about the commitment (or lack of) to processes.

We can distinguish between an external standard of rationality, whether or not an agent achieves her goals regardless of the process by which she gets there, and an internal standard, which places further conditions on the decision-making process by which the agent achieves those goals (Gold [2018]). Expected utility theory uses an external standard. It may seem somewhat counter-intuitive that a theory of rational behaviour makes no reference to processes of deliberation and, on some interpretations, to mental states. However, when modeling rational behaviour, an external standard of rationality may be preferable to an internal one. Presumably much of our everyday behaviour is instrumentally rational. But most of that behaviour does not involve conscious reasoning, which may be the exception rather than the rule.

To allow the possibility that everyday behaviour is instrumentally rational, we must allow that there can be rationality without conscious reasoning. Pettit [2007], who would reserve “reasoning” for conscious inference, contrasts reasoning with the sort of rational judgments that “bubble up” and the spontaneous rationality that results from subpersonal processing, hence he recognizes that there can be rationality without reasoning, in the sense of conscious inference (p. 514).

In the literature on team reasoning, Hurley [2005a, 2005b] has explicitly endorsed an external standard of rationality. She investigates local procedures and heuristics from which collective units of agency can emerge. Hurley argues that the processes in an agent that actually generate his or her rational behaviour need not be isomorphic with the theoretical account of why the behaviour counts as rational. Consistent with this position, she thinks that collective rationality can be had by animals, including very basic single celled creatures that are clearly not capable of conscious reasoning. Bacharach had an open mind about whether animals could team reason, from which it seems likely that he did not think that conscious deliberation was necessary for team reasoning (Gold [2005]). He also explicitly denies conscious access to reasoning processes, saying, “We should not expect people to be able to identify the reasoning principles that govern their conclusions even when these principles are sound” (Bacharach [2006:45]).

Expected utility theory makes no assumptions about reasoning processes so, to the extent that team reasoning strictly adheres to expected utility theory, it should not be a problem for the account that there seem to be spontaneously formed intentions in which no actual reasoning takes place.
5. Reasoning as computation
and team reasoning
as a computational-level model

Seen purely as an extension of game theory, and hence expected utility
theory, team reasoning is clearly compatible with spontaneous collective
intentions. However, Gold and Sugden [2007a, p. 111] imply that they intend
a thicker interpretation than that of expected utility theory, identifying team
reasoning with the mental process by which intentions are formed and
saying that “an intention is interposed between reasoning and action”. Indeed,
part of the intuitive appeal of team reasoning is that, as well as
showing that coordination and cooperation can be considered rational by
game theory’s own lights, we can all recognize team reasoning as a mode of
reasoning that we sometimes use. If we identify team reasoning with a
mental process, then the question of how the theory can handle spontane-
ous collective intentions still arises.

Philosophers working on practical reasoning often identify reasoning with
conscious deliberation. However, some philosophers who work on reason-
ing leave its mechanisms very open. For instance Broome [2002] says that
reasoning is a process that takes you from some of your existing mental
states to a new one, without specifying anything about the nature of the
process. Other philosophers have denied that reasoning involves either
deliberation or consciousness, which indicates that a similar approach may
be successful for team reasoning.

David Velleman [2014] denies that reasoning involves deliberation. His
“supervisory” conception of reasoning even allows reasoning and inten-
tion formation to be post hoc. Since that clearly is not how people in the
team reasoning literature conceptualise team reasoning, I will not explore
Velleman’s position any further. (But see Gold, unpublished manuscript, for
discussion.) Instead, I will focus on the possibility that reasoning is not
always a conscious activity.

Some philosophers have argued that reasoning is not always a conscious
process (Harman [1986], Velleman [1989]). David Velleman [1989] says in his
early work that reasoning shouldn’t be confused with thinking. He implies
that reasoning does not refer to an actual conscious process: “Of course, if
we wish to discuss a person’s reasoning, we must state the contents of his
premises and conclusions, and we must state what logical relations he
believed to obtain among them. But we shouldn’t be deceived by this neces-
sity into assuming that the reasoning under discussion consisted in similar
statements rehearsed in the reasoner’s head.” (Velleman [1989], p. 106). Har-
man [1986] is explicit that reasoning may not be a conscious activity. He
distinguishes reflection from making revisions to ones views. Revision may
occur without reflection. It is implied that reflection is conscious thinking,
but Harman identifies reasoning with making revisions. He says that, “One
can reason without knowing what the relevant principles of revision are and

© Dalloz | Téléchargé le 14/10/2023 sur www.cairn.info (IP: 35.160.27.221)
it may well be that reasoning is a relatively automatic process whose outcome is not under one's control" [1986, p. 2].

The idea that reasoning may not be a conscious activity is consistent with the way that Bacharach conceives of team reasoning. He regarded team reasoning as a “mental activity”, which is a causal process that determines, or partially determines, choice [2006, p. 121f.]. However, he explicitly did not assume that people have full conscious access to their reasoning:

“We should not expect people to be able to identify the reasoning principles that govern their conclusions even when these principles are sound: for example, most people easily and reliably reason in accordance with modus ponens, but almost no-one can tell you that it is modus ponens that sanctions their conclusions” (Bacharach [2006], p. 45).

In the pilot experimental test of his theory, Bacharach [2006] solely observed behaviour, as have other tests of team reasoning (see Karpus & Gold [2016] for an overview); they do not usually use protocol analysis, which elicits verbal reports of the thinking process.

When Bacharach [2006] refers to agents as “team reasoning”, he does not mean to contrast team reasoners with other agents who do not deliberate or who decide unconsciously. Rather he contrasts it with other “team mechanisms”. Bacharach defines a choice mechanism as a causal process that determines the choice of agents ([2006], Ch. 4). Choice mechanisms are diverse. As well as ratiocination, they include acting on advice and randomization. A team mechanism is any choice mechanism that ensures that the team’s common goal is achieved. Team reasoning contrasts with other team mechanisms, for example where one agent (the director) does all the computations, identifies the individual agent in control of each component, and instructs each of them in their parts. Team mechanisms differ in the way in which computation is distributed and the pattern of message sending; all mechanisms apart from team reasoning require communication.

Therefore Bacharach’s use of “team reasoning” contrasts with that of Ross [2014], who makes a similar point to the one I made in Section 4, but with different terminology. Ross retains the term “reasoning” for deliberation and argues that, because most economic responses only sometimes involve deliberate reflection, it would be better to use the term “team agency” rather than “team reasoning”, and to think of team reasoning as one special mechanism that supports team agency. Whereas Ross contrasts “reasoning” as a deliberative process with other non-deliberative processes that can support team agency, for Bacharach “team reasoning” picks out a particular team mechanism, in contrast with other team mechanisms (and, in his framework, it might make more sense to say that collective intentions are the result of a team mechanism than the result of team reasoning, but I leave that issue to one side).

The status of team reasoning in Sugden’s theory is rather different from Bacharach’s. Sugden [1990] presents team reasoning as a model that cap-
tures salient features of reasoning. His central concept is “reason to believe”, where to say that an agent has reason to believe a proposition is to say that the proposition can be inferred from propositions that she accepts as true, using rules of inference that she accepts as valid (Sugden [2003]). Sugden says that a person does not necessarily believe everything that she has reason to believe: the logic of reason to believe can be counter-factual. His characterization of decision theoretic rationality is as providing “formal representations of modes of reasoning that people take to be valid, and that they choose to act on.” (Sugden [2003], p. 169) There is nothing in Sugden’s presentation to suggest that taking a mode of reasoning to be valid and acting on it necessarily involves consciously thinking it through. Equally, there is nothing to preclude the idea that people sometimes consciously team reason, or have awareness that team reasoning was the basis for their conclusions. When Gold and Sugden refer to what they are doing as presenting core features of team reasoning using “explicit” schemata [2007a p. 137, see also p. 121; 1007b p. 280], we should understand that as leaving open the possibility that team reasoning could be in some sense implicit.

These ideas about reasoning are consistent with an account where people reason in accordance with a rule without being able to specify the rule that they are following, which Polanyi [1962] called tacit knowing. Sometimes people draw the conclusions that would be mandated by the rules of logic, not only without knowing those rules, but also whilst they are doing something different at a conscious level. For instance, we know from research on the Wason selection task that people are not very good at making abstract inferences using the material conditional (Wason [1966]; Wason & Shapiro [1971]). However there is plenty of evidence that people draw the conclusions that it mandates when the inferences are presented within a more concrete context (Cosmides & Tooby [1992]; Griggs and Cox [1982]; Johnson-Laird, Legrenzi and Legrenzi [1972]; Wason and Green [1984]). Although subjects’ answers are in accordance with the material conditional, since they don’t use the material conditional in the abstract context, it is unlikely that they consciously use it in concrete contexts – or even know what rule it is that sanctions their conclusions.

The idea that reasoning is like tacit knowing is compatible with the way reasoning is studied in cognitive science, artificial intelligence, and machine learning. Following Marr [1982], cognitive scientists distinguish between three different levels at which one can describe a system, sometimes called an information processing system: the computational level, where the goal of the system and the logic behind its output are specified; the algorithmic level, which specifies the representation for the input and output, and the algorithm by which inputs are transformed into outputs; and the level of implementation, or how the algorithm is physically realised in the brain or whatever the hardware is. Cognitive science has largely abandoned the idea that people use classical logic to implement reasoning at the algorithmic level, for a reason similar to the spontaneous collective intention critique: logical inference mechanisms are too slow to model the “automatic” information
processing that is antecedent to decision. It is important for cognitive scientists to have a conception of reasoning that is compatible with seemingly spontaneous actions. However, logic or the rules of probability are still used to provide a functional model at the computational level, of the task that our mental processes are designed to perform. For an example using logic, see Stenning and van Lambalgen [2008]; for Bayesian probability calculus, see Jaynes [1988] or Oaksford and Chater [2007]. Further, if one assumes that a system is designed to perform a certain class of tasks at the computational level, then one can learn about the unobservable algorithms that it uses by observing which of those tasks it performs badly (Simon [1969]).

From the perspective of cognitive science, team reasoning should be understood as a computational-level model. It models the goal of the system and the logic behind the output. The goal is to realise the optimal group behaviour, “what should we do?”, as computed by individuals according to the logic of team reasoning. However, the algorithmic and implementational levels may look very different from this. In particular, they need not be conscious processes. People might sometimes consciously use the procedure, but reasoning processes can also occur at a sub-agential level and an agent need not have conscious access to her reasoning. Therefore there could be collective intentions that result from team reasoning but which are not the consequence of conscious deliberation. Individual reasoning is a special case of team reasoning (Gold & Sugden [2007a]; Gold & Sugden [2007b]; Gold, in press [2018]); the idea that team reasoning is a computational-level model applies inter alia to the case of expected utility theory and individual reasoning, and the possibility of spontaneous individual intentions.

Note that, identifying team reasoning with a computational-level model, which is divorced from actual processing, it is compatible with dual-process theories. According to dual-process theories, there are two different processing systems: System 1, which is fast and intuitive, and System 2, which is slow and deliberative (Evans & Stanovich [2013]). However, just because a process is fast and automatic does not mean that it cannot be modelled by logic. Although logical reasoning is sometimes considered to belong to System 2 (Evans [2014]), there is no reason that System 1 cannot be modelled as a computational system (Stenning & van Lambalgen [2008]). Therefore we do not need to take a stand on whether the reasoning leading to collective intentions is a System 1 or a System 2 process, or even whether the System 1/ System 2 distinction is a good one (see Osman [2004], Osman [2013] for a critique).

The idea that reasoning is computation and team reasoning is a computational-level model captures Bacharach’s [2006] insight that team reasoners are computing and Sugden’s [1990] insight that team reasoning is a model. Since the algorithmic and implementation levels are not specified, the computational-level model of reasoning can accommodate both the idea that team reasoning results in collective intentions and the idea that there

---

4. Stenning & van Lambalgen [2008] are a notable exception to this rule, arguing that logic can provide a better model at the algorithmic level than is generally thought.
can be spontaneous collective intentions, which are not the result of conscious deliberation.

6. Putting bounds on what counts as a computational system: excluding simple creatures from being reasoners

One problem with my argument in the previous section, that the algorithms and implementation of reasoning can happen at a sub-agential level, is that it is too generous about what is allowed to count as reasoning. If reasoning can be done by any sub-agential process, then we might worry that Hurley’s [2005a, 2005b] single celled creatures are not only vindicated as rational, but also as reasoners. This seems like a step too far. We should look for a principled way of constraining “reasoning” that will exclude such simple creatures, even if we allow that they exhibit rational behaviour according to an external standard.

One possibility whereby we could exclude slime moulds from being reasoners is to invoke principles of parsimony. List [2016] takes this approach when considering whether simple mechanical devices could be agents. On his theory, anything can be an agent if it can be attributed belief-like and desire-like states, which lead it to intervene in the environment (List and Pettit [2011]). The theory implies that, in principle, even a simple mechanical object like a thermostat might qualify as an agent, since we can interpret it using Dennett’s [1987] intentional stance. However, List says that nothing would be gained in seeing a thermostat as an agent. Further, since a more straightforward mechanistic explanation is possible, principles of parsimony imply that we should not attribute agency to thermostats.

We could apply List’s [2016] argument strategy to the question of what counts as reasoning. Even if simple biological organisms and mechanical items can be characterized as having sub-agential mechanisms and therefore as reasoning, simpler explanations of their behaviour are possible, so a principle of parsimony suggests that we do not need to view them as reasoning.

However, following on from my definition of team reasoning as a computational-level model and the idea that reasoning is computation, we can make a stronger case than this. We can exclude slime moulds from being reasoners by giving criteria for what it means to do computations that exclude simple biological organisms and mechanical objects. There is a literature in philosophy of cognitive science that aims to put bounds on what counts as doing a computation and which offers a number of solutions. From this literature, we can sketch out some possible routes for defining computation that would exclude slime moulds from being reasoners.
In philosophy of cognitive science, researchers have worried that we can describe any physical system as implementing computations. Theories of computation are multiply realizable, what it is to be a computation is independent of the hardware on which the computation is done. So a computing machine could be, for instance, a system of cogs and levers, a hydraulic system, an electrical system, a brain, or even a set of suitably trained pigeons (Pylyshyn [1984]). Another way of glossing this feature, starting with the idea that rationality is a property of information processing systems (Oaksford & Chater [2009]), is that any of these set-ups could in theory be information processing systems. They all manipulate strings of states, transforming inputs into outputs. However, it is surprisingly philosophically tricky to differentiate brains and computers from stones and walls. For instance, Searle [1990b] argues that a wall implements any computer program since, for a large enough wall, we will be able to find patterns of molecule movements that are isomorphic to the formal structure of any program.

There are at least three routes that we can take to constrain what counts as a computation. (Rescorla [2015] and Piccinini [2015a] both provide excellent overviews; the taxonomy presented here is mine.) We could place non-triviality constraints on computations that would exclude walls from carrying them out, we could put constraints on how the computation is carried out at the algorithmic level, or we could put constraints on the implementation mechanisms. These are not three separate routes; the distinction between them is not robust because a computation involves a relation between a computational-level model, algorithms for carrying it out, and an implementation system. So, for example, non-trivial computations often require a more complex system at the implementation level.

There is disagreement about the best approach to constraining what counts as computation, in order to make the right exclusions, but there is broad agreement that it can be done. Even Searle [1990b] recognises that his example of a calculating wall would be excluded by a more demanding definition of a computation. One can exclude walls in several ways: via counter-factual conditions that put constraints on what the system does when the formal structure is not isomorphic to a program; via representations and semantics, insisting on “no computation without representation” (Fodor [1981], p. 180); or in terms of the mechanistic properties of the system. Any of these might do, but I will go into a little more detail about the mechanistic account because it is intuitively appealing and has some sympathetic consequences for what will count as a reasoner.

Piccinini [2007, 2015a, 2015b] distinguishes what it is for a physical system to be modelled computationally from what it is for the system to actually perform computations. He argues that, in order to be said to perform computations, it must be possible to give a mechanistic explanation of a system. By this he means a description of the system in terms of its spatiotemporal components, their functions, and their organization, to the effect that the system possesses its capacities because of how its components and their functions are organized. He identifies a computation as involving an input string, a rule for transforming it given by the mechanism’s function, and an output string. It may be possible to decompose computing mechanisms into sub-mechanisms that also compute. However, the most primitive compo-
nents of computation do not themselves compute. Piccinini’s definition includes all paradigmatic examples of computing mechanisms but excludes some systems whose behaviour we might merely model using computations, such as weather systems. It also excludes simple rocks and walls and, at the very least, makes it hard to argue that sub-personal biological systems, such as the digestive system, compute. What is interesting for my purposes about this definition of a computing system is that it leaves open the possibility that group agents, in the sense of List and Pettit [2011], could be computing systems. So long as they have suitable functional mechanisms for turning inputs into outputs, then we could consider them as computing and therefore as doing reasoning. I go into more detail about this possibility elsewhere (Gold, unpublished). Putting mechanistic constraints on computation, and hence reasoning, excludes slime moulds but not necessarily groups from being reasoners.

7. Conclusion

Tuomela argues that team reasoning will have trouble explaining spontaneous collective intentions. I showed that not all theories of collective intentionality need to accommodate spontaneous collective intentions. However, Gold and Sugden’s [2007a] argument strategy is that the same concept of intention and agency that we use in the individual case extends naturally to cover the collective intentions that occur during cooperative activity, so there is also a need to explain the relationship between reasoning and intentions in the individual case. Showing how team reasoning connects to spontaneous collective intentions also implies showing how individual reasoning connects to spontaneous individual intentions.

The notion of reasoning used in game theory is very thin, it says nothing about the process by which decisions are achieved, so it can easily accommodate spontaneous intentions. However, I argued that this solution is unsatisfactory for team reasoning, as an account of the formation of intentions. I argued that team reasoning should be understood as a computational-level model of reasoning, in the sense of cognitive science, which is compatible with the results of reasoning being spontaneous mental states, including spontaneous collective intentions. Individual reasoning is a special case in the team reasoning model, so the same solution applies to the connection between individual reasoning and individual spontaneous intentions.

Computational-level models of reasoning themselves place few restrictions on what counts as doing reasoning. I suggested that Piccinini’s [2007, 2015a, 2015b] mechanistic interpretation of a computational system is promising. It provides bounds on what counts as a computational system, which would exclude simple creatures from being reasoners, while allowing that computational systems can be made of smaller computational systems, and therefore leaving open the possibility that groups can be reasoners. This seems to me to be exactly the right result for theories of team reasoning and of group agency.
References


GOLD N. [unpublished paper], "Collective Intentions, Team Reasoning, and Rationality".


REP 128 (3) mai-juin 2018


REP 128 (3) mai-juin 2018


