**The Judicial Community and Team Reasoning**

Natalie Gold

Forthcoming in T. Marquez & C. Valentini (eds.) *Collective Action and the Law*, Routledge

Abstract. Sociality is key to the legal system, which is a constellation of different institutions, where individuals make decisions or take actions as members of groups. Some legal positivists even think that the law per se is a social practice, characterized by 'massively shared agency'. Shapiro has adapted Bratman's theory of shared cooperative activity to account for the fact that massively shared agency involves authority relations and participants who are not committed to a shared objective. However, his account still involves common knowledge, which seems implausible in large groups. Further, it fits the definition of what Bratman calls 'pre-packaged cooperation', which is not shared cooperative activity in the sense Bratman studies and is explicitly out of scope. There is another theory of cooperation and coordination that is better placed to explain massively shared agency: Team reasoning is an extension of standard game theory, which allows teams of individuals to count as agents and solve problems of cooperation and coordination. I show how team reasoning can encompass instances with authority relations and group members who are not committed to the activity, and that it does not require common knowledge of the identities of the participants. Further, solving some questions in the theory of team reasoning—about what reasoning consists of—shows that team still has something to offer even to legal scholars who have criticized the idea that the judicial community is an instance of massively shared agency. A court is the classic example of a 'group agent', a suitably organized collective that can be an intentional agent in its own right. Team reasoning is premised on the idea that the team can be an agent. By comparing group agency with team is an agency, I show that group agents like courts can be considered to be doing reasoning and even team reasoning. These two approaches correspond to two ways we could attribute collective responsibility: at the micro level, by understanding the reasoning and intentions of the individuals who participate in collective actions, or at the macro level, by focusing on the group agent and attributing collective responsibility on the basis of group membership.

## 1. Introduction

The legal system is a constellation of different institutions, including many where individuals make decisions or take actions as members of groups. There are courts, which have juries, and law firms; in the UK there are barristers' chambers and professional

bodies, such as the UK's Law Society and Bar Association. For some legal philosophers, there is also a deeper form of sociality that underpins the legal system. Following Hart (1961), legal positivists have argued that the foundation of the law and the source of its validity is social convention, involving widely shared patterns of behaviour, attitudes and beliefs. Therefore, we should not be surprised that philosophers of law have drawn on developments in social ontology to explain the validity and practice of legal systems, and the groups within them. In this chapter, I show how developments in theories of rational decision-making can help legal philosophers who are grappling with these issues, applying them to reasoning in groups and reasoning of groups.

Standard theories of rational decision-making assume that agency is vested in individuals and start with the premise that individuals act on their preferences and beliefs to achieve their goals. However, the idea that teams can be agents and that there are distinctive modes of reasoning that are used by members of groups has been around for a long time (Hodgson 1967; Regan, 1980; Gilbert, 1989; Hurley, 1989; Sugden, 1993, 2003; Hollis, 1998; Bacharach, 1999; 2006). Bacharach (2006) used the tools of decision theory to formalise team reasoning, to show how individuals could rationally act on a group's goals and preferences (or, for those of a less collectivist cast, their 'team-directed preferences', as per Sugden, 1993). Gold & Sugden (2007a, 2007b) show how team reasoning is a mode of practical reasoning that can account for collective intentions and shared cooperative activity, thus contributing to debates in social ontology.

There are two somewhat distinct contributions that team reasoning can offer to the philosophy of law and the study of legal systems. The first is an explanation of social practice that could inform legal positivism (reasoning in groups). The second is a theory of decision-making in groups that can inform our understanding of entities in the legal system (reasoning of groups). In each case, team reasoning offers the potential to refine or improve on existing suggestions. I take these two contributions in turn in each of Part I and Part II.

In Part I, I start with Shapiro's (2002, 2011, 2014) idea that the law involves 'massively shared agency'. I will present team reasoning (Section 2) and show how it fulfills the desiderata for a theory of massively shared agency (Section 3). Some authors are skeptical about whether the law really is an instance of massively shared agency (Smith, 2006; Kutz, 2001), I will not take a stand on that issue. However, there is a benefit of knowing about team reasoning, even for them. In Part 2, I examine the implications that team reasoning has for the sort of structured 'group agents' that exist within the legal system, the paradigm example of which is a court of judges deciding a case (List and Pettit, 2011). By resolving an issue for the theory of team reasoning, namely how we should understand 'reasoning' (Section 4), I contrast group agents with team agency and

show how my account can explain the deliberations of group agents (Section 5). A group agent, like a court or law firm, can be considered to be doing reasoning and even team reasoning.

## Part I

According to legal positivism, the law is socially constructed: its existence, content, and authority are based on conventions and norms (Hart, 1961). Therefore, it may not be surprising that contemporary legal theorists are drawing on social ontology, the philosophical study of the nature of social entities and social phenomena. Shapiro (2002, 2011, 2014) has argued that we should think of the law as a social practice, adopting Bratman's (1992, 1993) theory of 'shared cooperative activity'. For Bratman (1992), shared cooperative activity is characterised by three features: (i) *mutual responsiveness*, each participating agent attempts to be responsive to the actions and intentions of others, (ii) *commitment to the joint activity*, each participant has an appropriate commitment to the joint activity and their mutual responsiveness is in support of that commitment, and (iii) *commitment to mutual support*, each agent is committed to supporting the efforts of the other to do her part in the joint activity. Bratman then analyses the interlocking web of intentions that he takes to be sufficient for the existence of shared cooperative activity.

However, many theorists have argued that the commitments involved in Bratman's account are too strong to correctly describe the judicial community (Smith, 2006; Kutz, 2001). Even Shapiro acknowledges that Bratman's (1992, 1993) account needs to be modified. Bratman seeks to understand 'modest sociality', by which he means small-scale activities carried out by non-hierarchical groups, such as singing a duet or having a conversation together (Bratman, 2009). However, Shapiro wants to explain large-scale activities, which harness the agency of multitudes. He calls this 'massively shared agency' and as examples he gives 'business corporations, consumer cooperatives, trade unions, research universities, philanthropic organizations, professional associations, standing armies, political parties, organized religions, governments and legal systems, not to mention the collaborative ventures made possible by the internet, such as Wikipedia, MMOG's (massively multi-member on-line games), open-access software and the World Wide Web itself' (Shapiro 2014, p. 258). Shapiro identifies two key points of difference between modest sociality and massively shared agency. Firstly, massively shared agency often involves imbalances of power, which may be formalised in relations of authority. Secondly, it cannot be assumed that all individuals taking part have the same degree of shared commitment as in small-scale activities; indeed some of the individuals taking part may simply not be committed to the activity.

In order to accomodate authority relations, Shapiro discards the mutual support and mutual responsiveness conditions (Shapiro, 2014). However, for Bratman, mutual support and mutual responsiveness are central to an activity being a shared cooperative activity. When it is possible for each participant to do their part in the activity alone, then Bratman says that we have 'prepackaged cooperation', not shared cooperative activity (Bratman, 1992, p.239). It follows that Shapiro is using Bratman's theory to analyse a case of pre-packaged cooperation, which—according to Bratman—ought to be out of scope. Further, in order to allow for the possibility that some of the participants who are engaged in the activity are not fully committed to it, Shapiro modifies the joint commitment condition. For Bratman, the appropriate commitments of the participants involve a 'shared intention', which is a central theme in his work. However, in Shapiro's version, massively shared agency no longer involves a shared intention, he weakens the conditions: there is a shared plan to *J*, each member of the group intentionally follows her part, members of the group resolve their conflicts about *J*-ing in a peaceful and open manner; these conditions are common knowledge; and *J* takes place in virtue of them. One might wonder how much of Bratman's account is really left!

Having said that, there is one component of Bratman's theory that remains in Shapiro's solution which looks distinctly implausible in the case of massively shared agency, namely common knowledge. The idea that every participant knows that each other participant knows that each participant is intentionally doing her part, and so on ad infitum (Lewis, 1969; Aumann, 1976) seems implausible in some of Shapiro's examples of structured large-scale activity and downright impossible in the unstructured examples, such as Wikipedia and open source software. In these cases, participants may not even know the identity of others who are engaging in the activity.

Luckily, there is an alternative theory of cooperative activity that explains pre-packaged cooperation, which has variants that can cover authority relations, cases where group members are not committed to the activity and cases where participants do not know each other's identity. Bratman's archetype of pre-packaged cooperation is the sort of game theoretic model where individuals make their decisions and play their parts without communication. This suggests that accounts of cooperation from game theory might be of some interest. Classical game theory, where each individual asks separately 'What should *I* do?', has trouble explaining cooperation and coordination. Team reasoning is an extension of the standard theory, which allows teams of individuals to count as agents and for players to ask the question 'What should *we* do?' (Bacharach, 2006; Sugden, 1993). This may lead to cooperative actions: team reasoning has been used to explain cooperation and coordination (Gold & Sugden, 2007a) and it retains the connection to Bratman's shared intentions (Gold & Sugden, 2007b).

**2. Team reasoning: a theory of cooperation and coordination**

The theory of team reasoning was developed separately by Sugden (1993, 2003) and Bacharach (1999, 2006), and was connected to the literature on shared cooperative activity and collective intention by Gold and Sugden (2007a, 2007b). Its development was motivated by games of cooperation and coordination 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.

| | | Player 2 | |
|---|---|---|---|
| | | *stag* | *hare* |
| Player 1 | *stag* | 9, 9 | 0, 8 |
| | *hare* | 8, 0 | 7, 7 |

Figure 1: Stag Hunt

One of these games is the Stag Hunt, a version of which is shown in Figure 1. It is based on a story told by Rousseau, of two hunters who have to decide whether to hunt stag or hare. A stag has more meat but takes two to catch, whereas a hare can be caught by a single hunter acting alone. So the best outcome for each player is where they both hunt stag; the worst is to choose to hunt stag, only to find that the other player has hunted hare, so the lone stag-hunter goes without food for dinner. There are two pure-strategy Nash equilibria, (*stag, stag*) and *(hare, hare*), and (*stag*, *stag*) is strictly better than (*hare*, *hare*) for both players. It seems clear that the two players should each play *stag*. 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 only recommend that, if a player expects the other player to play *stag*, then it is rational for her to play *stag*. However, if she expects the other player to play *hare*, then it is rational for her to play *hare*. 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 *stag* rather than *hare*.

Game theory also has a 'refinement programme', which aims to introduce criteria that can select between Nash equilibria, but this does not help. Harsanyi and Selten (1988) introduce two possible criteria, but they indicate different solutions. According to 'payoff dominance', rational players should choose the Nash equilibrium where all players receive a higher payoff than other outcomes, if one exists. This recommends each player to choose *stag*. According to 'risk dominance', the players should play *hare* because it is less risky—the players would sacrifice the extra payoff from catching a stag but they would each be

certain of having something for dinner no matter what the other player does. So we have two conflicting refinements. The payoff-dominance criterion solves the problem by fiat, begging the question of why it is rational to choose the strategies that lead to the payoff dominant equilibrium. Indeed, Harsanyi (1995) abandoned the payoff-dominance principle—for more details on the argument against it that he found so convincing see Gold & Colman (2018).

Team reasoning can explain why rational players should choose *stag* (Gold, 2012; Gold & Colman, 2018). In standard game theoretic reasoning, an individual player asks 'what should *I* do?', and 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 should *we* do?', and to select the best profile of actions for the team. Then each individual can reach the conclusion that she should choose her component of that profile. There are different modes of reasoning that players may use: they may either use standard individual reasoning or team reasoning. A player team reasons when she 'identifies' with a group, which is to say that she conceives of that group as a unit of agency, acting as a single entity in pursuit of some single objective, and considers herself a part of it.

In the Stag Hunt game, it is clear that (*stag, stag*) would be the profile of actions that is preferred by the team 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 (which we should just think of as a technical way of saying that they aim to achieve the team's objective, see Gold 2018), then each can reason that she should play *stag*, as a part of playing (*stag*, *stag*) (Gold & Colman, 2018). In a game with a payoff dominant equilibrium, we expect the payoff dominant profile to be preferred by the team, so team reasoning can explain why it is rational to choose the strategies that lead to the payoff dominant equilibrium. In other games of cooperation, such as the prisoner's dilemma, team reasoning can recommend co-operation if the cooperative solution is known to be the team's most preferred option. (Contra Harp (2009), there does not need to be agreement on the team utility function 'all the way down'.)

It is simplest to explain how team reasoning leads to coordination and cooperation in situations with common knowledge. However, team reasoning does not rely on common knowledge assumptions. In particular, if we relax common knowledge of group identification, then a team reasoner can take into account the risk that the other player will play *hare* (or *defect* in a prisoner's dilemma), but she does so from the perspective of the team, as I will explain in Section 3.

Team reasoning was first proposed as a theory of rational cooperation and coordination, but Gold and Sugden (2007b) have argued that team reasoning leads to

*collective intentions*, where these are understood as mental states held by individuals, which involve a plurality of participants in a joint activity, for instance, when two singers say, 'we intend to sing a duet'. They argue that collective intentions are distinguished from individual intentions by reference to the unit of agency in the reasoning that led to the formation of the intention. A team reasoner in the stag hunt plays *stag* because the team prefers (*stag, stag*). Gold and Sugden show how participatory intentions of the form 'I intend to do my part' can be derived from intentions of the form 'we intend', fulfilling a desideratum of Searle (1990).

Gold and Sugden (2007b) also show how a Bratman-style analysis, of a shared intention as a state of affairs that obtains when people participate in a joint activity, could relate to team reasoning. Bratman (1992, p.339) differentiates shared cooperative activity from 'pre-packaged cooperation', where agents play their roles in the joint action with no further interaction. Team reasoning looks like pre-packaged cooperation. Gold and Sugden suggest that Bratman is analysing our disposition to reason and act as a member of a group, in relation to the objective of executing some broadly-defined joint activity. They identify Bratman's 'shared intentions' with high-level strategic intentions, which set the framework within which subsequent tactical (team) reasoning takes place. These strategic intentions guide the practical reasoning that leads to low-level tactical intentions, which are the individual's mental states.

**3. Team reasoning and massively shared agency**

To account for massively shared agency, we need to be able to explain situations where there is an authority figure directing the activity, where some members of the group may not be committed to the activity, and where participants do not have common knowledge. There are variants of team reasoning that can explain each of these.

(i) Authority (simple direction)

Massively shared agency will often require centralized control over participants' behaviour, with an authority that guides individuals' conduct. Shapiro (2014) notes that, in these cases, participants may look towards the authority to coordinate their behaviour rather than taking their cues from each other.

In order to understand how the theory of team reasoning can account for such authority, we can draw on Bacharach's (2006) idea of a *choice mechanism*, a causal process that determines what people do. A *team mechanism* is one is that ensures the team's common goal is achieved (at least in the usual scheme of things, since the reasoning is always ex ante). A team mechanism consists of three steps: computing the best profile of actions for the team, identifying the component of the profile that falls to any particular

individual, and telling that individual what her component is. Then the individual agents all need to follow their instructions.

Team reasoning, as presented above, is a team mechanism which has each individual team member doing the computation, identifying her own component, and following her own instructions. However, we can imagine another team mechanism, which Bacharach (2006) called *simple direction*, where the computing, identifying, and instructing is done by one individual—the *director*—and all that the other team members need do is to follow the instruction that the director communicates to them. Provided that the team members have reason to think that the director's judgment has some validity, then each of them only needs to play their part. For instance, in the Stag Hunt, if an authority figure is working out the plan and has told each player their part is to play *stag*, and they both know that the other player knows his or her part, then that is enough to make it rational to play *stag* in order to achieve the best team outcome.

In some ways, simple direction could be considered more basic than the kind of full-blown team reasoning considered above. Bacharach (2006) says that team reasoning is basically 'do-it-yourself direction'. The important core of team reasoning as a theory of collective intention and shared cooperative activity is preserved, namely the idea of doing something because it is a part of the best plan for the team. The conditions for the simple team reasoning presented in Section 2 (common knowledge that each member group identifies and common knowledge that each member aims to maximize the team payoff function) can be fulfilled in situations where the members do not do any of the computing themselves. Therefore, the mechanism of simple direction is well placed to model authority relationships.

(ii) Commitment (restricted team reasoning)

In general, Shapiro (2014) thinks that accounts of shared agency require too much in the way of commitment from participants to account for massively shared agency. He argues that, with massively shared agency, it is implausible that every individual is committed to the same goal, calling popular accounts 'hyper-committed'—even Kutz's (2000, 2007) account of participatory intentions, which is a minimalist account of shared cooperative activity that is supposed to be suitable for large groups. Shapiro gives examples of agents who are 'alienated' from a joint activity, participating without being committed. For instance, they may fulfill their part because they are being paid and not because they are committed to realising the shared goal.

Here, we can draw on Bacharach's (2006) variant of *restricted team reasoning*. This models the case where there is a group with an objective, but some members of the group do not group identify and will not team reason (the 'remainder'). What Shapiro calls

alienation, Bacharach refers to as a failure to function. In this case, the active members of the team are a strict subset of the group. A team mechanism will determine the choices of those individuals who identify with the team, but will have no causal influence on the choices of the remainder. Members who team reason should do the actions that best achieve the aims of the group, given the choices of those who fail to function.

It is important for Bacharach (2006) that a team reasoner decides to do her part in the team action *because* it is her part in the best possible combination for the team. There are no restrictions on what actions the remainder take, they may turn out to be a help or a hindrance. However, even if they are helpful, the actions of the remainder are not produced by team reasoning. Therefore, according to the team reasoning account, they are not a part of any collective intention or shared cooperative activity, even if their actions appear to be directed at the same ends.

(iii) Common knowledge (circumspect team reasoning)

Restricted team reasoning retains the feature that team reasoners know which group members will not function and what those group members will do instead. Bacharach saw these as serious limitations, which he addressed in a variant called *circumspect team reasoning*, where he makes membership of the remainder random (Bacharach, 2006). He introduces a probability $\omega$ that any individual identifies with the group and functions, which is realised independently for each individual. The team is now the set of individuals who turn out to function. Team reasoners must work out the best team plan, given the possibility that others in the group may fail to function as team members.

Bacharach (2006) illustrates circumspect team reasoning using the prisoner's dilemma. It provides an easy case study because it is clear that the individual reasoners in the remainder should choose defect. What a team reasoner should do will depend on the relative payoffs (Gold and Sugden, 2007a, 2007b). If the group ranks 'off-diagonal' outcomes (where some members cooperate and some members defect) above the outcome where all defect, then team reasoning will tell team members to cooperate regardless. If the group ranks the outcome where all defect above the off-diagonal outcomes, then whether a team reasoner should cooperate depends on the value of $\omega$: they cooperate if $\omega$ is high and defect if it is low. It would also be possible to apply circumspect team reasoning to Stag Hunt, but you would need to say more about a potential individual reasoner's strategic reasoning in order to solve the game.

In circumspect team reasoning, there is no common knowledge of who is functioning. This variant relaxes the assumption of common knowledge of group identity and the assumption of common knowledge about what the team is trying to do. Bacharach

(2006) formalised team reasoning using game theory, which assumes that the properties of the game are common knowledge among the players; Gold and Sugden (2007) translated it into reasoning schema, which make the role of common knowledge more explicit. In circumspect team reasoning, common knowledge conditions are replaced by what they call 'T-conditional common knowledge', whereby group members know that *if* another group member identifies, then they know the content of the proposition. They illustrate T-conditional common knowledge with an example. Imagine a member of an organisation that operates via underground cells, so members' identities are unknown to each other. If all new members are inducted by oath, which they are told is common to the whole organisation, then they know that *if* someone is a member of the organisation, then she knows the content of the oath. If *T* is the set of members, then the content of the oath is *T*-conditional common knowledge. But there is no common knowledge about who is a member of *T*. So in circumspect team reasoning there is *T*-conditional common knowledge that each member aims to maximize the team payoff function: those who identify know that, if other players group identify and turn out to be team members, then they will maximise the team payoff function. But they do not know who will group identify.

Circumspect team reasoners, who do their part because it is their part in the best combination, are engaged in a cooperative activity with other circumspect team reasoners even though they do not know who the other team reasoners are. Therefore, circumspect team reasoning can explain massively shared agency in the absence of common knowledge of group membership.

In many ways, we have ended up close to Shapiro's own solution: there is a group, at least some members of which (the team) have an objective or goal (the best outcome from the point of view of the group) and a plan to achieve it (the optimal team profile). The team members may work out their part in the plan, but they may also take instructions from an authority (simple direction). There may be group members who are not participating in the plan (in which case the team members engage in restricted team reasoning). However, unlike Shapiro's account, there does not need to be common knowledge of group membership (the team members can engage in circumspect team reasoning), which seems essential for analysing some of his examples.

There is a second interesting way in which the team reasoning account diverges from Shapiro's account of massively shared agency. For Shapiro, team members who are not committed to the group still count as a part of the shared activity. But it follows from the team reasoning account that only the members of the group who are committed to the objective (the team) are a part of the web of shared intentionality; members of the group who are not committed and do not plan to contribute are not.

This will have implications for the collective responsibility of groups and how it relates to the responsibility of their individual members. When arguing that groups have collective responsibility, a common move is that groups have collective responsibility in virtue of having shared intentions. In that case, team reasoning implies that only those who are committed to the activity bear responsibility (i.e. only those who are team members, which may not be everyone whose actions contribute to the outcome or every member of the wider group). Members who contribute to the outcome, but not for the reason that they are realising the group plan, will not be a part of the collective responsibility. However, they may still turn out to bear some individual responsibility, either because they enable the group or because they are an accessory to the outcome. In other words, they bear responsibility because of their causal role in the outcome.

An alternative move would be to ascribe collective responsibility in virtue of being a member of the group that caused the outcome, rather than in virtue of being a part of the shared intention. Such an argument might start by appealing to the individual's responsibility for choosing to become a part of the group or for not exiting the group, or to the benefits that the individual derives from being a member of the group. Then, to the extent that individuals have the possibility of making those entry and exit choices or to derive benefits, the group members might be collectively responsible—even those who were not committed to the shared activity. If we go the route of ascribing collective responsibility on the basis of group membership (and not causal contribution), then we would need to say more about how we attribute group membership, since a group may not have a well-defined boundary.

**Part II**

In the above, I have shown how team reasoning can be useful for a theory of massively shared agency. I have not addressed the question of whether the legal system actually is an instance of massively shared agency. One might wonder whether there can really be said to be a shared objective or a shared plan. In the remainder of this chapter, I show how team reasoning is of relevance when thinking about group agents, such as courts, so it has something to offer even to legal theorists who do not believe that the legal system involves massively shared agency.

Shapiro's examples of massively shared agency encompass both structured groups that have a formal organisation and procedures—such as corporations, trade unions, and universities; and unstructured groups, which are looser and informal—such as many of the ventures that people engage in over the internet, such as open source software. (NB. I do not mean to claim that this is a binary distinction, there may be a spectrum of

structuredness.) Many of the entities in the legal system are structured groups, such as courts or law firms. A structured group might also be a *group agent* in List and Pettit's (2011) sense: a suitably organized collective that can be an intentional agent in its own right, over and above individual members. A three-member court is the paradigm example of a group agent. Now we might doubly wonder what the usefulness of team reasoning is to philosophers of law, when they already have a theory of group agency to draw on.

So far, I have presented team reasoning without defining what I mean by reasoning, relying on the reader's intuitive idea. However, it turns out that philosophers have vastly different intuitions—and theories—about what reasoning consists in. Therefore, in what follows I reflect on how we should understand the reasoning in the theory (Section 4). Then, in Section 5, I show how this has relevance to group agents, such as law courts. I compare group agency and team reasoning, and connect the two.

**4. How should we understand the "reasoning" in team reasoning?**

Velleman (2014, Ch.13) criticizes philosophers for their implicit conception of practical reasoning as a time-consuming process that must occur before action. Indeed, some philosophers seem to have interpreted team reasoning as conscious reasoning, leading them to argue that team reasoning requires too much reasoning to explain collective intentions (Schweikard & Schmid, 2013; Tuomela, 2009). Since many of our everyday actions occur without conscious reasoning, if team reasoning must be done consciously, then the theory would have a restricted range of application. I have explained elsewhere why team reasoning is not necessarily conscious reasoning (Gold, 2018). I won't rehash those arguments here but, as well as those arguments, we can simply note that the basic capacity for cooperation identified above, which forms the basis for collective intentions, is to carry out one's part in a team mechanism for the reason that it is one's part in reaching the team's goal.

Nevertheless, there is still a question of how we should understand the 'reasoning' in team reasoning. It is helpful to look in more detail at Velleman's critique and his proposed solution, before proposing a solution of my own.

Velleman (1989, p.106) has also argued 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 necessity into assuming that the reasoning under discussion consisted in similar statements rehearsed in the reasoner's head.' He thinks that we can engage in some of the activities that we associate with reflective reasoning, such as making observations and inferences 'without

consciously thinking them through' (Velleman, 1985, p.52). Therefore, for Velleman, even when an agent entertains no conscious reflective thoughts, she can still be engaged in practical reflection.

Velleman (2014) presents his own account of practical reasoning. He denies that reasoning is a mental procedure of deliberation that precedes action, which leads him to make the following two moves.

Velleman (2014, Ch. 13) denies that reasoning consists of deliberation: 'deliberation is not a mode of practical reasoning: it's a procedure ancillary to practical reasoning'. Instead, he proposes a *supervisory* conception of practical reason. In his picture, most behaviour is caused by sub-agential mechanisms. It is as though we are all aircraft flying on auto-pilot, which steers us according to our desires and beliefs, and corrects course in response to feedback about progress toward the desired outcome. These functions are sub-agential automatic processes. Agential-level practical reasoning is like the pilot, who perceives and supervises these automatic processes, but only rarely wrests control. However, Velleman is explicit that this does not imply that we should see our behaviour as automatic or rote. Perception and supervision are types of thinking—Velleman even characterises them as being reflective thinking. So, for Velleman, practical reasoning involves thinking, it just does not necessarily involve deliberation.

Velleman (2014) also denies that reasoning must be antecedent to action: 'practical reasoning does not necessarily precede action, much less take time in advance.' He thinks that the standard conception of practical reason—as making a plan and carrying it out—is the exception, rather than the rule. Velleman (2007) notes that it is a problem for Bratman's (1987) planning conception of agency that we sometimes form intentions just before acting: if intentions are plans, then there would be no point in doing this. Instead of assisting with planning, Velleman claims that one purpose of practical reasoning is to help us come to an understanding of ourselves: he gives an epistemic account of intention, whereby the way that the agent describes her action conveys information that explains the behaviour to herself. For Velleman (2014), this is another function of practical reason: In the course of supervising our stream of behaviour, practical reason 'places it under action concepts'. By this, Velleman seems to mean that reason describes behaviour, in the Davidsonian sense of placing our actions under an intentional description. Practical reasoning can come up with these action descriptions antecedent to, simultaneous with, or retrospective to behaviour. This is not supposed to be a deliberative process even though, contra Velleman, one might think that deliberation is sometimes involved in processes of self-description and self-understanding.

Thus, Velleman (2014) distinguishes practical reasoning from deliberation antecedent to action in two ways: (a) practical reasoning consists of the supervision of a

process that is already ongoing, and (b) practical reasoning consists of placing our actions under an intentional description, allowing the agent to make sense of herself. These are both agent-level activities.

Velleman's (2007, 2014) idea that practical reasoning allows us to make sense of ourselves is potentially compatible with the team reasoning account of collective intentions. A team reasoner cooperates for the reason that it is her part in reaching the team's goal. In a Velleman-type picture, team reasoning would be a particular way of rationalizing behaviour; collective intentions would be a way that the agent understands her behaviour—that she did an action because it was a part of the best team plan. In addition, it is worth noting that Velleman (1997) has his own account of what it is to share an intention, which is consistent with Sugden's (2003) account of mutually assured team reasoning and which draws on Gilbert's (1989) idea of a conditional commitment, whereby all the agents commit to participating in the joint action if the others do too. We can explain this using the metaphor of a country signing an international treaty, which includes among its conditions that it will come into effect only if and when it has been ratified by some number of other nations; once this condition is met, the treaty is binding on all the nations that have ratified it (Gold and Sugden, 2007a, 2007b). Similarly for a shared intention, which only comes into effect if others actually participate.

However, in the team reasoning literature, team reasoning is clearly supposed to precede action. Gold and Sugden (2007b) explicitly say that an intention is interpolated between a decision and an action—precisely the view that Velleman rejects! Bacharach (2006, p.121f.) regarded team reasoning as a 'mental activity', which is a causal process that determines, or partially determines, choice. We can square this circle by dropping a condition that Velleman implicitly places on reasoning, namely that it is an agential-level activity.

The idea that reasoning is a sub-agential activity 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: 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 abandoned the idea that people use classical logic to implement reasoning at the algorithmic level, since logical inference mechanisms are too slow to model the 'automatic' information processing that is antecedent to decision. 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 (e.g. Stenning and van Lambalgen, 2008; Jaynes, 1988; Oaksford and Chater 2007). These rules can predict what a well-operating system will do in a specified class of tasks. Then, by observing which of those tasks it performs badly, one can learn about the unobservable algorithms that the system uses (Simon, 1969).

Team reasoning—whether represented as the probability calculus of Bacharach (1999, 2006) or the manipulation of propositions of Gold and Sugden (2007a, 2007b) or the formal logic of Sugden (2003)— should be understood as a computational level model. It models the goal of the system and the logic behind the output. The goal of the individual (the 'system') is to realise the optimal group behaviour, 'what should *we* do?', as computed according to the logic of team reasoning. Therefore, the claim that team reasoning leads to collective intentions should be understood as the idea that the team goal and the logic of team reasoning underpin the behaviour associated with collective intentions. When we go to Barcelona together, then we both have the goal that we go to Barcelona together and our individual actions can be described by a specific logic, the logic of team reasoning. These are a different goal and a different logic from individual reasoning.

However, this picture carries no implication that the formal model represents anything that is available in conscious thought, or even that they provide a correct model of algorithmic processing. When Gold and Sugden (2007a, 2007b) specify team reasoning as the manipulation of conscious propositions, we should take them to be specifying a rule of inference. This might sometimes be used consciously, but it need not be. Inference still counts as reasoning if the processing is done purely at the sub-agential level. Bacharach 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).

We know that people may 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 much reasoning may occur at a sub-agential level is consistent with a picture 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'.

If reasoning can be done by any sub-agential process, then we might worry that any goal-directed system will end up being categorised as a reasoner. That could include very simple creatures or machines. Therefore we should regard the computational-level model as a necessary but not a sufficient condition for the attribution of reasoning and look for a principled way of constraining what counts as 'reasoning'. There is no space to go into that question here (instead, see Gold, 2018), but note that there are a number of ways that we could do this, by putting constraints on any of Marr's (1982) three levels. We could insist that reasoners execute non-trivial computations. We could constrain the algorithm, building on the idea that rationality is a property of information processing systems (Chater & Oaksford, 2009), and insisting on a demanding notion of processing. Or we could constrain the implementation system that physically realises the computation.

Of particular interest is Piccinini's (2007, 2015a, 2015b) mechanistic account of the implementation system. He argues that, in order to say a system performs computations, it must be possible to describe 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. The computation itself involves an input string, a rule for transforming it given by the mechanism's function, and an output string. This account seems to place the boundaries of what counts as a computing system in the right place, excluding many systems that we might model using computations without wanting to say that they are performing computations.

Team reasoning is a computational model, but that is not a sufficient condition. To actually count as doing reasoning a system needs to be performing computations; one way we can understand that is that there needs to be a suitable mechanistic account of the system. I take it that that can be provided at the sub-agential level for an individual and therefore that we can understand reasoning as a sub-agential process that occurs before action. This account is also compatible with there sometimes being conscious processes, either at the time of action or after action, when we understand ourselves by reconstructing our reasons.

**5. Group agents as team reasoners**

Courts and law firms are examples of structured groups, which have a formal organisation and procedures. A structured group might also be a group agent. List and Pettit's (2011) theory of group agency was inspired by the *doctrinal paradox*, where the members of a three-member court each hold positions that are individually consistent, but the aggregate of those positions is collectively inconsistent (Kornhauser & Sager, 1986; Kornhauser & Sager, 1993). The classic example is a civil suit for breach of contract, where the judges have to find both that there is a contract and that the defendant acted in breach of it if they are to decide that damages are due to the plaintiff. It is possible for three judges to have the following consistent positions on each of these three propositions, which are individually consistent but collectively inconsistent (see Figure 2). Groups can hold positions that no individual within them supports.

| | **Is there a contract?** | **Is the defendant in breach?** | **Does the defendant owe the plaintiff damages?** |
|---|---|---|---|
| Judge 1 | Y | N | N |
| Judge 2 | N | Y | N |
| Judge 3 | Y | Y | Y |
| Majority opinion | Y | Y | N |

Figure 2: The doctrinal paradox

The fact that we can ascribe judgments to the group without them being easily reducible to the attitudes of individuals led List and Pettit (2011) to introduce the concept of a group agent. They define an agent as an entity that exhibits the following three features: (i) representational states that depict how things are in the environment, (ii) motivational states that specify how it requires things to be in the environment, and (iii) the capacity to process its representational and motivational states, leading it to intervene in the environment whenever that environment fails to match a motivating specification. In other words, the group agent can be attributed belief and desire-like states, which underpin its actions. A group agent is defined by properties that are attributed to the group, not the individual members. In addition, a group agent is the sort of institution that exists over time and being an agent also involves an impetus to *rational unity*, taking steps to ensure that its intentional attitudes are consistent with each other (Pettit, 2003). Therefore, a court is a group agent because it has appropriate mechanisms for aggregating judgments in order to come to a consistent point of view.

In contrast, team reasoning allows that groups can be agents, but all the action is at the level of the individual group members. The individual must *entify* the group, conceiving of it as a unit of agency, acting as a single entity in pursuit of some single objective, and then 'identify' as a part of it. So there is a shared goal, represented by Bacharach (2006) as a group utility function, but referred to by Sugden (1993) as the individuals' 'team-directed preferences'. All beliefs are held at the level of the team members. Team reasoning is a type of reasoning that is valid for members of a team agent to use and which is similar in structure to modes of individual practical reasoning. It is done by individuals who have the individual concept *I* and a notion of grouphood (Gold & Harbour, 2012). We can use team reasoning to explain how we can reason about what 'we want to achieve' and about what 'we should choose', given that mental states are located in individual minds.

It is consistent with this that, unlike a group agent, a team agent may be ephemeral. Team reasoning can explain one-off incidents of cooperation between groups of people who may never meet again. The existence of a team agent need not imply any metaphysical claim about the status of agents in our social ontology and team agency need not be a very thick sort of agency (Gold & Sugden, 2007b). So the two theories have different targets.

The two theories may sometimes overlap, since individuals within a group agent may team reason, if they take it to be a unit of agency and ask 'What should we do?'. However, List and Pettit's (2011) theory of group agency is neutral about the motivations or modes of reasoning used by the individuals within group agents. The whole point of the theory is that one can attribute group agency based on the structure and processes of the group, rather than on the basis of claims about the individuals within it. However, it seems unlikely that individuals within courts are team reasoning, since it seems more likely that they ask themselves the question 'What does the law require?'. (Indeed, one issue with accounts of the law as an instance of massively shared agency it that it seems to be stretching it to say that there is a shared objective, except maybe in the vaguest of terms like 'achieving justice'. Is that precise enough to get the shared practice story off the ground?)

However, drawing on the discussion in Section 4, we might think that a group agent itself is the sort of thing that can reason. As Pettit (2001, p.210) puts it, a group agent must 'collectivise reason'. There will be internal processes that maintain the consistency of its intentional attitudes. If these are functional mechanisms for turning inputs into outputs, then we could consider group as computing and therefore as doing reasoning. We can think of the internal processes as reasoning processes. Therefore, one implication is that a

group agent itself may be a locus of reasoning (Gold, 2018). The reasoning of a court is the result of sub-agential processes (namely the individual judges in the court).

It follows from this that a group agent could also be a locus of team reasoning. A group could ask itself 'what should *we* do?', where the *we* refers to multiple group agents. For example, a number of firms in an industry could team reason, to set prices or achieve other gains from cooperation. Or a number of chambers could act together. In this case, the firm or chambers would be a group agent, which can team reason, and the collection of firms or chambers a team agent.

The contrast between team agency and group agents brings us back to the options for attributing collective responsibility, set out at the end of Part I. The theory of team reasoning has implications for collective responsibility, if we go the route of ascribing responsibility based on the contribution of the individuals involved. However, the theory of group agency abstracts from individual processes and could provide a basis for attributing collective responsibility on the basis of group membership (on the grounds that the group itself is reasoning), irrespective of the motivations of actions of individuals.

## 6. Conclusion

Team reasoning should be of interest to legal scholars who take a social practice view of the law and, even for those who do not, it has implications for our understanding of groups and organisations that are part of the constellation of the legal system.

For scholars who believe that the law is a social practice, team reasoning provides an account of social cooperation that does not rely on mutual responsiveness and support, and there are versions that relax common knowledge conditions. All of these are necessary if we are to account for massively shared agency in large unstructured groups. More broadly, we can apply the idea that reasoning is a sub-agential process to understand how structured group agents, such as courts, law firms and barristers' chambers, can be doing reasoning in virtue of the actions of their individual members. Team reasoning *in* groups and reasoning *by* groups correspond to two ways we could attribute collective responsibility: at the micro level, by understanding the reasoning and intentions of the individuals who participate in collective actions, or at the macro level, by focusing on the group agent and attributing collective responsibility on the basis of group membership.

## Acknowledgments

Work on this chapter 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.

**References**


Aumann, R. J. (1976). Agreeing to disagree. *The annals of statistics*, pp.1236-1239.

Bacharach, M. (2006). *Beyond Individual Choice.* Princeton: Princeton University Press.

Bacharach, M. (1999). Interactive team reasoning: a contribution to the theory of cooperation. *Research in Economics*, 53, pp.117-147.

Bratman, M. E. (2009). Modest sociality and the distinctiveness of intention. *Philosophical studies*, *144*(1), 149-165.

Bratman, M. (1993). Shared Intention. *Ethics,* 104, pp.97–113.

Bratman, M. E. (1992). Shared cooperative activity. *The Philosophical Review*, 101(2), pp.327-341.

Bratman, M. (1987). *Intention, plans, and practical reason.* Cambridge Ma., Harvard University Press.

Cosmides, L., and Tooby, J. (1992). Cognitive adaptations for social exchange. In: J. Barkow, L. Cosmides, and J. Tooby, ed., *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford: Oxford University Press, pp.163-228.

Gilbert, M. (1989). *On social facts*. London: Routledge.

Gold, N. (2018). Team Reasoning and Spontaneous Collective Intentions. *Revue d'économie politique*, 128(3), pp.333-353.

Gold, N. (2012). Team reasoning and cooperation. In: S. Okasha and K. Binmore, eds, *Evolution and Rationality: Decisions, Cooperation and Strategic Behaviour.* Cambridge: Cambridge University Press

Gold, N. (2005). Teamwork: Multi-disciplinary perspectives. In: N. Gold, ed., *Teamwork: Multi-Disciplinary Perspectives*, Palgrave MacMillan: Basingstoke, pp.1-21.

Gold, N., & Colman, A. M. (2018). Team Reasoning and the Rational Choice of Payoff-Dominant Outcomes in Games. *Topoi*, pp.1-12.

Gold, N. and D. Harbour. (2012). Cognitive Primitives of Collective Intentions: Linguistic Evidence of our Mental Ontology. *Mind and Language*, 27, pp.111-136.

Gold, N. and R. Sugden (2007a). Theories of Team Agency. In: F. Peter and S. Schmidt, eds, *Rationality and Commitment.* Oxford: Oxford University Press, pp.280-312.

Gold, N. & R. Sugden. (2007b). Collective Intentions and Team Agency, *Journal of Philosophy,* 104, pp.109–137.

Griggs, R. A., and Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, 73(3), pp.407-420.

Harp, R. (2009). Collective Goals, Collective Reasoning, Collective Action. (Unpublished doctoral dissertation). Stanford University, California.

Harsanyi, J. C., and Selten, R. (1988). A general theory of equilibrium selection in games. Cambridge, Ma.: MIT Press.

Hart, H. L. A. (1961). *The concept of law*. Oxford: Clarendon law series.

Hollis, M. (1998). *Trust within reason*. Cambridge: Cambridge University Press.

Hurley, S. L. (1989). *Natural reasons: Personality and polity*. Oxford: Oxford University Press.

Hurley, S. (2005a). Rational Agency, Cooperation and Mind-Reading. In: N. Gold, ed., *Teamwork: Multi-Disciplinary Perspectives*. Palgrave MacMillan: Basingstoke, pp.200-15.

Hurley, S. (2005b). Social Heuristics that Make Us Smarter. *Philosophical Psychology*, 18, pp.585-612.

Jaynes, E. T. (1988). How does the brain do plausible reasoning? In: G. J. Erickson and C. R. Smith, eds, *Maximum-Entropy and Bayesian Methods in Science and Engineering*. Springer: Dordrecht, pp.1-24.

Johnson-Laird, P. N., Legrenzi, P., and Legrenzi, M. S. (1972). Reasoning and a sense of reality. *British Journal of Psychology*, 63(3), pp.395-400.

Karpus, J. and Gold, N. (2016) Team Reasoning: Theory and Experiment. In: J. Kiverstein, ed., *Handbook of the Social Mind.* New York: Routledge, pp.400-17.

Kornhauser, L. A., & Sager, L. G. (1993). The one and the many: Adjudication in collegial courts. *Cal. L. Rev.*, 81, pp.1-59.

Kornhauser, L. A., & Sager, L. G. (1986). Unpacking the court. *Yale LJ*, 96, pp.82-117.

Kutz, C. (2007). *Complicity: Ethics and law for a collective age*. Cambridge: Cambridge University Press.

Kutz, C. (2001). The Judicial Community. *Noûs*, 35(s1), pp.442-469.

Kutz, C. (2000). Acting together. *Philosophical and Phenomenological Research*, pp.1-31.

Lewis,, D. (1969). *Convention: a philosophical study.* Cambridge: Cambridge University Press.

List, C. and Pettit, P. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press.

Marr, D. (1982). *Vision: A computational approach.* New York: W.H. Freeman and Company.

Oaksford, M., and Chater, N. (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences*, 32(01), pp.69-84.

Oaksford, M., and Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning.* Oxford: Oxford University Press.

Pettit, P. (2007). Rationality, Reasoning and Group Agency, The Dialectica Lecture 2006, *Dialectica*, 61, pp.495-519.

Pettit, P. (2003). Groups with Minds of their Own. In: F. Schmitt, ed., *Socializing Metaphysics*. New York: Rowan and Littlefield, pp.167-93.

Pettit, P. (2001). *A theory of freedom: from the psychology to the politics of agency*. Oxford: Oxford University Press.

Piccinini, G. (2007). Computing mechanisms. *Philosophy of Science*, 74(4), pp.501-526.

Piccinini, G. (2015a). Computation in Physical Systems. In: Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), [online] Available at: <http://plato.stanford.edu/archives/sum2015/entries/computation-physicalsystems/>.

Piccinini, G. (2015b). *Physical Computation: A Mechanistic Account*. Oxford: Oxford University Press.

Polanyi, M. (1962). Tacit knowing: Its bearing on some problems of philosophy. *Reviews of Modern Physics*, 34(4), pp.601-615.

Regan, D. H. (1980). Utilitarianism and Co-operation. Oxford: Clarendon Press.

Schweikard, D. P. and Schmid, H. B., (2013). Collective Intentionality. In: Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy* (Summer 2013 Edition), [online] Available at: <http://plato.stanford.edu/archives/sum2013/entries/collective-intentionality/>

Searle, J. (1990). Collective intentions and actions. In: P. Cohen, J. Morgan, and M. Pollack, eds, *Intentions in Communication*. Cambridge: MIT Press, pp.401–415.

Shapiro, S. (2014). Massively shared agency. In: M. Vargas and G. Yaffe, eds, *Rational and Social Agency: The Philosophy of Michael Bratman*. Oxford University Press: USA, pp.257-293.

Shapiro, S. (2011). *Legality*. Cambridge, Ma.: Harvard University Press.

Shapiro, S. J. (2002). Law, plans, and practical reason. *Legal Theory*, 8(4), pp.387-441.

Simon, H. A. (1969). *The sciences of the artificial.* Cambridge, Ma.: MIT Press.

Smith, M. N. (2006). The law as a social practice: Are shared activities at the foundations of law?. *Legal Theory*, 12(03), pp.265-292.

Sugden, R. (2003). The logic of team reasoning. *Philosophical Explorations,* 6, pp.165-181.

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

Stenning, K., & van Lambalgen, M. (2008). *Human reasoning and cognitive science*. Cambridge Ma.: MIT Press.

Tuomela, R. (2009). Collective intentions and game theory. *Journal of Philosophy*, 106, pp.292-300.,

Velleman, J. D. (1997). How to share an intention. *Philosophy and Phenomenological Research: A Quarterly Journal*, pp.29-50.

Velleman, J. D. (2014). *The Possibility of Practical Reason.* 2nd edition. Ann Arbor, MI: Michigan Publishing, University of Michigan Library, [online] Avaliable at: http://dx.doi.org/10.3998/maize.13240734.0001.001

Velleman, J. D. (2007). 'What good is a will?' In: A. Leist, ed., *Action in context*. Berlin/ New York: Walter de Gruyter, p193-215. Reprinted as Ch.12 in Velleman (2014).

Velleman, J. D. (1989). *Practical Reflection.* Princeton: Princeton University Press.

Velleman, J. D. (1985). Practical reflection. *The Philosophical Review*, 94(1), pp. 33-61.

Wason, P. C. (1966). Reasoning. In: B. M. Foss, ed., *New horizons in psychology*. Harmondsworth: Penguin, pp.135-51.

Wason, P. C., and Green, D. W. (1984). Reasoning and mental representation. *The Quarterly Journal of Experimental Psychology*, 36(4), pp.597-610.

Wason, P. C., and Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology,* 23, pp.63–71.