Team Reasoning: Theory and Evidence

Jurgis Karpus and Natalie Gold

for *Routledge Handbook on Social Cognition*, ed. Julian Kiverstein

February 15th, 2016

Introduction

Orthodox game theory identifies rational solutions to interpersonal and strategically interdependent decision problems, games, using the notion of individualistic best-response reasoning. When each player's chosen strategy in a game is a best response to the strategies chosen by other players, they are said to be in a Nash equilibrium—a point at which no player can benefit by unilaterally changing his or her strategy. Consider the Hi-Lo and the Prisoner's Dilemma two-player games illustrated in Figures 1 and 2. The strategies available to one of the two players are identified by rows and those available to the other by columns. The numbers in each cell represent payoffs to the row and the column players respectively in each of the four possible outcomes in these games.

	Hi	*Lo*
Hi	2, 2	0, 0
Lo	0, 0	1, 1

Figure 1: The Hi-Lo game

	C	*D*
C	2, 2	0, 3
D	3, 0	1, 1

Figure 2: The Prisoner's Dilemma game

There are two Nash equilibria in the Hi-Lo game, (*Hi*, *Hi*) and (*Lo*, *Lo*), since, for either player, the strategy *Hi* is the best response to the other player's choice of *Hi* and the strategy *Lo* is the best response to the other's choice of *Lo*.[1] As such, individualistic best-response reasoning identifies two rational solutions of this game but does not resolve it definitively for the interacting players. For many people, however, (*Lo*, *Lo*) does not appear to be a rational solution and it seems that the outcome (*Hi*, *Hi*) is a clear definitive resolution of this game. In the case of the Prisoner's Dilemma game, there is only one Nash equilibrium, (*D*, *D*), since, for either player, the strategy *D* is the best response to whatever the other player is going to do. As such, individualistic best-response reasoning resolves this game definitively. However, due to the inefficiency of the outcome (*D*, *D*) compared to the outcome (*C*, *C*)—both players are better off in the latter than they are in the former—for some the outcome (*C*, *C*) is not obviously irrational and there is a division of opinion (at least outside the circle of professional game theorists) about what a rational player ought to do in this game.

Orthodox game theory's inability to resolve these games satisfactorily, in particular its inability to definitively resolve the Hi-Lo game, motivated the development of the theory of team reasoning.[2] According

1 These are Nash equilibria in pure strategies. There is a third Nash equilibrium in mixed strategies, in which both players randomize between the two available strategies by playing *Hi* with probability 1/3 and *Lo* with probability 2/3.

2 For some of the early and later theoretical developments see Bacharach (1999, 2006), Sugden (1993, 2000, 2003, 2011, 2015), Gold and Sugden (2007a, 2007b), and Gold (2012).

to the theory of team reasoning, people may not always be employing individualistic best-response reasoning in games. The theory allows that people may, instead, identify rational solutions from the perspective of a team, a group of individuals acting together in the attainment of the best outcome(s) for that group. This, in turn, enables team reasoning to show how the Hi-Lo game can be rationally resolved definitively and how the outcome (*C*, *C*) in the Prisoner's Dilemma game can be rationalized.

The theory of team reasoning gives a new account of why coordination and cooperation can be rational by introducing the possibility of multiple levels of agency into classical game theory. But it is also supposed to tell us something about how people reason. It is a model of decision-making, which abstracts and simplifies, but "it captures salient features of real human reasoning" (Sugden, 2000, p. 178). We might think of team reasoning as operating at Marr's (1982) computational level, specifying the goal of the system and the logic behind the output, but leaving open how the computation is implemented and how it is realised in the brain (Gold, in press).

A number of different versions of the theory of team reasoning have been proposed and developed. These differ with respect to what triggers decision-makers' adoption of the team mode of reasoning and what team-reasoning individuals try to achieve. We review these developments in the first part of this chapter (Sections 1 to 3). There is also a nascent but growing body of experiments that attempt to test the theory. We review some of these studies in the second part (Sections 4 to 6). Finally, with Section 7 we conclude and present a suggestion for further experimental work in this field.

I. Theory

1. What is Team Reasoning?

The individualistic best-response reasoning of orthodox game theory is based on the question of which of the available strategies in a game a particular player should take, given his or her individual preferences and his or her beliefs about what the other players are going to do. Each player's personal motivations in games are represented by the payoff numbers they associate with the available outcomes, and the optimal strategy is that which gives the player in question the highest expected payoff. In this light, the best strategy for an individualistically reasoning player in the Hi-Lo game (see Figure 1 above) is conditional on that player's belief about what the other player is going to do: play *Hi* or play *Lo*. In the Prisoner's Dilemma game (see Figure 2) the best strategy is unconditionally to play *D*.

Team reasoning, on the other hand, is based on the question of what is optimal for the group of players acting together as a team. A team reasoner first identifies an outcome of a game that best promotes the interests of the team and then chooses the strategy that is his or her part of attaining that outcome. If the outcome (*Hi*, *Hi*) is identified as uniquely optimal for the team, then team reasoning resolves the Hi-Lo game definitively. Similarly, if any of the outcomes associated with the play of *C* in the Prisoner's Dilemma game, e.g., the outcome (*C*, *C*), are ranked at the top from the point of view of the team, the strategy *C* can be rationalized.

It is important to note that reasoning as a member of a team is not a mere transformation of players' personal payoff numbers associated with the available outcomes in games. To see this, consider again the Hi-Lo game. Suppose that, from the point of view of the team, the outcome (*Hi*, *Hi*) is deemed to be the

best, the outcome (*Lo*, *Lo*) is deemed to be the second-best and the outcomes (*Hi*, *Lo*) and (*Lo*, *Hi*) the worst. Replacing the two players' original payoff numbers with numbers that correspond to the team's ranking of the four outcomes in the game does not change the payoff structure of the original game in any way, since the players' individual payoffs are already in line with the valuation of outcomes from the team's perspective. The key difference here is that individualistic reasoning is based on evaluating and choosing a particular strategy based on the associated expected personal payoff, whereas team reasoning is based on evaluating the outcomes of the game from the perspective of the team, and then choosing a strategy that is associated with the optimal outcome for the team.

There are two important questions that the theory of team reasoning needs to address: "when do people reason as members of a team?" and "what do people try to achieve when they reason as members of a team?". In other words, is it possible to identify circumstances or types of games in which the interacting players are likely to adopt team reasoning and the mechanism by which they adopt it, and, once they team reason, is it possible to specify a functional representation of what they take the goals of the team to be? We turn to reviewing the various proposals for answering these questions in the following two sections.

2. What Triggers Team Reasoning?

Different versions of the theory of team reasoning have different answers to the question of when people team reason. One answer, mainly associated with Bacharach (2006), is that the mode of reasoning an individual uses is a matter of that decision-maker's psychological make-up, which in turn may depend on certain features of the context in which decisions are made, but otherwise lies outside of the individual's conscious control. A second answer, proffered by Sugden (2003), is that an individual may choose to endorse a particular mode of reasoning based on considerations about the potential benefits of one or another possible mode of reasoning and his or her beliefs about the modes of reasoning endorsed by other players, but this choice is outside of rational evaluation. A third possibility, proposed by Hurley (2005a, 2005b), is that individual decision-makers come to choose the team mode of reasoning as a result of rational deliberation itself.

The first position, the idea that the adoption of team reasoning is outside of an individual's control, can be found in the version of the theory of team reasoning presented by Bacharach (2006) and Smerilli (2012). Here the mode of reasoning that an individual adopts is a matter of a psychological frame through which he or she sees a decision problem. The idea is similar to that of Tversky and Kahneman (1981, p.453), who define a frame as, "the decision-maker's conception of the acts, outcomes, and contingencies associated with a particular choice". In Tversky and Kahneman's Prospect Theory, framing a decision in terms of losses or gains affects the part of the value function that decision-makers apply, thus affecting their choices. In Bacharach's theory of team reasoning, framing a decision in terms of "we" or "I" affects the goals that decision-makers aim to achieve, which one might think of as using the individual or the group value function, and the mode of reasoning that they apply.

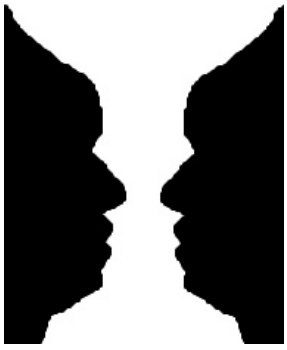
Figure 3: The goblet illusion

If the individual frames the decision problem as a problem for him or her individually, i.e., in terms of individualistic best-response reasoning, then he or she identifies solutions offered by that mode of reasoning alone. If, on the other hand, the individual frames the decision problem as a

problem for a group of players acting together as a team, i.e., in terms of team reasoning, then he or she identifies solutions offered by team reasoning and not by individualistic reasoning. This idea can be likened to that of seeing either a goblet or two faces in the goblet illusion picture illustrated in Figure 3 (also known as the Rubin's vase). Looking at this picture it is possible to see either a goblet or two faces opposite from each other, but only one of these images at a time and not both of them simultaneously. In the same way, a decision-maker is said to frame a decision problem either from the point of view of individualistic best-responding or from the point of view of reasoning as a member of a team, but not in terms of both these perspectives at the same time.[3]

The psychological frame through which an individual analyzes a particular decision problem may depend on factors that lie outside of the description of a game, but it can be influenced by the payoff structure of the game itself: Bacharach mentions as possible triggers the strong interdependence and double-crossing features. Roughly speaking, strong interdependence occurs when there is a Nash equilibrium that is worse than some other outcome in the game from every player's individual point of view. Both the Hi-Lo and the Prisoner's Dilemma games have this feature: in Hi-Lo, the Nash equilibrium (*Lo*, *Lo*) is worse for both players than the outcome (*Hi*, *Hi*); in the Prisoner's Dilemma, the outcome (*D*, *D*) is worse for both than the outcome (*C*, *C*). This means that the outcomes (*Lo*, *Lo*) and (*D*, *D*) are not Pareto efficient. (An outcome of a game is said to be Pareto efficient if there is no other outcome available that would make some player better-off without at the same time making any other player worse-off.) According to Bacharach, strong interdependence increases the likelihood that an individual would frame a decision-problem as a problem for a team.

The double-crossing feature is the possibility of an individual personally benefiting from a unilateral deviation from the team reasoning solution. It is the incentive to act on individual reasoning when one believes that the other player is acting on team reasoning. This feature is present in the Prisoner's Dilemma but not the Hi-Lo game. In the Prisoner's Dilemma, each individual would personally benefit from a unilateral deviation from the cooperative play of (*C*, *C*). There is an incentive to double-cross the other player, playing *D* if the other player is expected to play *C*. According to Bacharach, the possibility of double crossing decreases the likelihood of a particular decision-maker framing a decision problem as a problem for a team. Smerilli (2012) formalizes this intuition, providing a model where the double-crossing feature causes players to vacillate between frames.

Another possibility, suggested by Bardsley (2000, Ch. 5, Section 6), is that payoff differences within cells introduce an inter-individual aspect to game situations and Pareto superior outcomes a collective one, which respectively inhibit or promote team reasoning. Zizzo and Tan (2007) introduce the notion of “game harmony”, a generic game property describing how conflictual or non-conflictual the players' interests are, and suggest some ways of measuring it, the simplest one being just the correlation between the players' payoffs across outcomes. They show that game harmony measures can predict cooperation in some 2x2 games (i.e. two-player games with two strategies available to each player). Note that this is a different idea

3 However, see Bacharach (1997) for a model where, as well as the “I frame” and the “we frame”, there is also an “S (superordinate) frame”, which is active when someone manages to see a problem from both the “I” and the “we” perspectives. Someone in the “S frame” is still compelled only to evaluate the outcomes from either an “I” or a “we” perspective, and the cooperative option is chosen by a player in the “S frame” if it is the best option from the perspective of team reasoning and not worse than any other rational solution in terms of individualistic best-responding.

from Bacharach's strong interdependence and double crossing features (as noted by Bacharach, 2006, p. 83). The measures agree in pure coordination games, where players' interests are perfectly aligned, and in zero-sum games, where players' interests are perfectly opposed. However, in mixed motive games, the two ideas do not always point in the same direction. Bacharach is clear that common interest is strong when the possible gains from coordination are high or the losses from coordination failure are great, which leaves open how consensual players' interests are in general, whereas game harmony is simply a measure of how consensual players' interests are and does not take into account the size of the potential gains from cooperation.

In addition to the structural features of games themselves, priming group or individualistic thinking in decision-makers could be expected to also play an important role in determining which frame of mind the individuals would be in and which mode of reasoning they would use in games. Bacharach (2006) surveys the literature from social psychology on group identity, the effect of social categorization and the minimal group paradigm, and took himself to be contributing to that literature. Group identity may be triggered by players' recognition of belonging to the same social group or a particular category, having common interests, being subject to a common fate or simply having face-to-face contact. For Bacharach, group identity is a "framing phenomenon" (2006, p. 81). To group identify is to conceive of oneself as a group member: to represent oneself as a group member and have group concepts in one's frame. Hence, for him, all these factors that trigger group identity may cause a shift from the "I frame" to a "we frame" (see, in particular, Bacharach, 2006, pp. 76-81).

Sugden (2003, 2011, 2015) takes the second position described above: an individual decision-maker may choose to endorse team reasoning, but there is no basis for rational evaluation of this choice. For Sugden, there may be numerous modes of valid reasoning and an individual decision-maker may choose to endorse any one of them, but none of these modes of reasoning are privileged over others on the basis of instrumental rationality. Instrumental practical reasoning allows an agent to infer the best means to achieve its goals. Therefore instrumental rationality must presume both the unit of decision-making agency as well as its goals and neither of these are amenable to evaluation by the theory of rationality itself. However, Sugden discusses a number of conditions that may need to be satisfied in order for an individual to endorse team reasoning. He sees team reasoning as cooperation for mutual advantage. Hence whether or not a person team reasons will depend on whether it is beneficial for that decision-maker to do so individually (in terms of his or her individual preferences and goals).[4] Further, team play by a particular decision-maker may be conditional on the assurance that other players are reasoning as members of a team as well.

Sugden can still accept a lot of what Bacharach says about the circumstances in which people team reason, as can any theory of team reasoning (Gold, 2012). Sugden's agents still need to conceive of the decision problem as a problem for the team, rather than as a problem for them as individuals, before they can team

4 To understand the idea of mutual benefit it is important to note that the payoff numbers associated with different outcomes in games are meant to represent the interacting players' preferences that, in some sense, mirror their goals and motivations in these games. In this light, higher payoff values represent higher levels of preference satisfaction. This interpretation of payoffs, however, causes a general difficulty in experiments, in which we need to assume that the payoffs presented to participants are correctly aligned with their true motivations and preferences. If games used in experiments are incentivized using monetary payoffs, for example, we need to assume that the interacting participants' true motivations are aligned with the maximization of personal monetary payoffs.

reason (Sugden, 2000, pp. 182-183). The difference is that, in Sugden's theory, people make a choice to team reason and assurance plays a part in this, whereas for Bacharach, team reasoning is the result of a psychological process and may lead team reasoners to be worse off than they would have been if they had reasoned as individuals (for instance they may cooperate in a Prisoner's Dilemma when the other player defects; for more on how this can happen see Gold, 2012).

Bacharach and Sugden agree that all goals are the goals of agents and that it is not possible to evaluate those goals without first specifying the unit of agency. Thus, even though Sugden allows for the unit of agency to be chosen, it is not a matter of instrumentally rational choice. In contrast, Hurley (2005a, 2005b) suggests that there is no need to identify the unit of agency with the source of evaluation of outcomes and that we can identify personal goals prior to identifying the unit of agency. Hurley says that, "As an individual I can recognise that a collective unit of which I am merely a part can bring about outcomes that I prefer to any that I could bring about by acting as an individual unit." (Hurley, 2005a, p. 203). Hence, Hurley suggests that principles of practical rationality can govern the choice of the unit of agency; one should choose the unit of agency that best realizes one's personal goals. If that unit is the team, then one should team reason as a matter of practical rationality.[5]

The problem for theories that allow rational choice of the unit of agency is how to specify the goals that we should be striving for, independent of the unit of agency. Hurley suggests that we should privilege personal goals but, once we recognise that there are other possible units of agency (and evaluation), we might question why it is the case that the personal level takes priority. For a decision-maker in Regan's (1980) theory of cooperative utilitarianism, for example, the goal is always utilitarian and the question is what unit of agency one should be adopting given this goal. However, taking goals as given to us by our theory of value, or moral theory, turns team reasoning from a theory of rational choice into a theory of moral choice, which is not intended by many of its proponents.

The problem is brought out in recent work by Gauthier (2013). Gauthier has long held that it can be instrumentally rational to cooperate in the Prisoner's Dilemma game (Gauthier, 1986). In a recent re-working of his theory, Gauthier (2013) contrasts two opposed conceptions of deliberative rationality: maximization (equivalent to individualistic best-response reasoning) and Pareto-optimization. He suggests that Pareto-optimization is a necessary condition for rationality in multi-player games. A Pareto-optimizing theory "provides only a single set of directives to all the interacting agents, with the directive to each premised on the acceptance by the others of the directives to them" (Gauthier 2013, p. 607). The outcome selected must be both efficient and fair in how it distributes the expected gains of cooperation. Although he does not explicitly use the term "team reasoning", it is clear that Gauthier's theory is similar to ideas of team reasoning for mutual gain. His justification for team reasoning is that it would pass a contractarian test whereby it is "eligible for inclusion in an actual society that constitutes a cooperative venture for mutual fulfilment" (Gauthier, 2013, p. 618).

5 Hurley (2005b) follows Kacelnick (2006) in distinguishing two conceptions of rationality: rationality as consistent patterns of behaviour and rationality as processes of reasoning that underlie that behaviour. Hurley subscribes to the first conception, therefore 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. Hence she investigates local procedures and heuristics from which collective units of agency can emerge. According to her picture, choices can be instrumentally rational even if they result from a crude, low-level heuristic.

As Gauthier (2013, p. 624) puts it, his goal is to show that “social morality is part of rational choice, or at least, integral to rational cooperation”. However, whilst he has sketched out what Pareto-optimization would involve, Gauthier has not provided any argument for its rationality; he concludes that he has not yet been successful in bridging the two and that more needs to be done regarding the connection to rationality (in other words, how instrumental rationality may require us to cooperate in social interactions). But it is hard to see how Gauthier could bring Pareto-optimization within instrumental rationality. If he goes the same route as Hurley and privileges the individual's perspective and goals, then he needs to explain why it is instrumentally rational to cooperate when the individual could do better by deviating in situations that have the double crossing feature. Or, if the idea is that there is some addition to instrumental rationality for choosing the level of agency, then it is hard to see how to characterize such a process. A reasoning process already seems to presume an agent who is doing the reasoning. As Bardsley (2001, p. 185) puts it, “the question `should I ask myself “what am I to do?” or “what are we to do?”?' presupposes a first person singular point of view”.

3. What Do Teams Strive For?

We now turn to reviewing different proposals about a team's goals. The approaches presented differ in whether they require individual decision-makers to sometimes sacrifice their personal interests for the benefit of other members of a team and whether they rely on making interpersonal comparisons of the interacting players' payoffs. Bacharach (2006) mentions Pareto efficiency as a minimal condition, i.e., that if a strategy profile is superior in terms of Pareto efficiency, then it is preferred by the team to the strategy profiles that it is superior to. The exclusion of all Pareto inefficient strategy profiles, however, says nothing about how a team should rank the remaining strategy profiles where there is a conflict of personal interests, such as presented by the pair of outcomes (*C*, *D*) and (*D*, *C*) in the Prisoner's Dilemma game.

In some of the early developments of the theory, e.g., Bacharach (1999, 2006) as well as some of the more recent papers, e.g., Colman et al. (2008, 2014) and Smerilli (2012), the maximization of the average of the interacting players' payoffs is used as an example of a team payoff function. This function—that is, a mathematical representation of a team's goals in an interpersonal interaction—is consistent with the strong interdependence feature and the related Pareto efficiency criterion discussed in the previous section, and it is easy to see that it selects the outcomes (*Hi*, *Hi*) and (*C*, *C*) as uniquely best for a team in the Hi-Lo and in the above Prisoner's Dilemma games respectively. (Specifically, maximising the average payoff will select (*C*, *C*) in any Prisoner's Dilemma game where the average of the payoffs from (*C*, *C*) is higher than the average from any other outcome.) This function, however, sometimes fails with respect to the notion of mutual advantage. Consider a slight variation of the Prisoner's Dilemma game illustrated in Figure 4. Here the maximization of the average of the two players' personal payoffs would prescribe the attainment of the outcome (*D*, *C*). As such, it would advocate a complete sacrifice of the column player's personal interests for the benefit of the row player alone.

	C	*D*
C	2, 2	0, 3
D	5, 0	1, 1

Figure 4: A variation of the Prisoner's Dilemma game

The averaging function also relies on making interpersonal comparisons of the interacting players' payoffs, which suggests, for example, that the row player prefers the outcome (*D*, *C*) to (*C*, *C*) to a greater extent than the column player prefers the outcome (*C*, *D*) to (*D*, *D*) in Figure 4. Strictly speaking, such comparisons go beyond the orthodox assumptions of expected utility theory, which make numerical representations of the interacting players' preferences possible but do not automatically grant their interpersonal comparability. As such, a theory of team reasoning that uses this function as a representation of a team's goals is only applicable in contexts when such interpersonal comparisons of payoffs are possible.[6]

Although not many alternative functional representations of a team's goals have been proposed (perhaps partly because many works on the theory of team reasoning have so far considered examples where team-optimal outcomes seem evident, such as the outcomes (*Hi*, *Hi*) and (*C*, *C*) in the Hi-Lo and the Prisoner's Dilemma games), a number of properties that representations of a team's goals should satisfy have been put forward. One of them is the notion of mutual advantage discussed by Sugden (2011), which suggests that the outcome selected by a team should be mutually beneficial from every team member's perspective. Although he does not present an explicit function of a team's goals, in a recent paper Sugden (2015) proposes to measure mutual advantage relative to a particular threshold. The threshold is each player's personal *maximin* payoff level in a game—the payoff that he or she can guarantee him or herself independently of the other players' chosen strategies. In the Hi-Lo game this is 0 for both players. In the Prisoner's Dilemma game of Figures 2 and 4, this is 1, since it is the lowest possible payoff that either player can attain by playing *D*. A strategy profile is said to be mutually beneficial if (a) it results in each player receiving a payoff that is greater than his or her *maximin* payoff level in a game, and (b) each player's participation in team play is necessary for the attainment of those payoffs.[7]

Karpus and Radzvilas (2016) propose a formal function of a team's goals that is based on the notion of mutual advantage similar to the one above whilst also incorporating the Pareto efficiency criterion (in a weak sense of Pareto efficiency, which means that an outcome of a game is efficient if there is no alternative that is strictly preferred to it by every player in the game). It suggests that an outcome that is optimal for a team is one that is associated with the maximal amount of mutual benefit. The extent of

6 If the numbers in game matrices, for example, represent monetary payoffs and all players value money in the same way (that is, an additional unit of currency is subjectively worth just as much to one player as it is to another), then interpersonal comparisons of payoffs are not problematic. If, however, the payoff numbers in games represent players' personal motivations as von Neumann-Morgenstern utilities, then such comparisons are tricky.

7 Note that according to the above definition, both (*Lo*, *Lo*) and (*Hi*, *Hi*) are mutually beneficial outcomes in the Hi-Lo game, since even (*Lo*, *Lo*) guarantees both players more than their *maximin* payoff. Hence, the definition of mutual advantage does not, by itself, exclude Pareto inefficient outcomes and, for Sugden (2015), which of the mutually beneficial outcomes will be sought by a team depends on which outcome each player in a game will have "reciprocal reason to believe" will be sought by every other player. See also Cubitt and Sugden (2003) for more details on "reason to believe", which is based on a reconstruction of Lewis' (1969) game theory.

mutual benefit is measured by the number of payoff units by which an outcome advances every player's personal interests relative to some threshold points, such as the players' *maximin* payoff levels in games as suggested by Sugden (2015). For example, if both players' *maximin* payoffs (in a two-player game) are 0, an outcome associated with a payoff of 3 to Player 1 and payoff of 2 to Player 2 offers 2 units of mutual benefit (the additional unit of individual benefit to Player 1 is not mutual).[8] As such, the function identifies the outcome (*Hi*, *Hi*) as uniquely optimal for a team in the Hi-Lo game and prescribes the attainment of the outcome (*C*, *C*) in all the versions of the Prisoner's Dilemma game discussed above.[9]

II. Evidence

4. The Difficulties of Empirical Testing

There is a major difficulty that any empirical test of team reasoning will unavoidably face: the fact that a number of separate hypotheses are being tested at once. The main hypothesis to be tested is whether people reason as members of a team in a particular situation. This, however, is intertwined with two additional auxiliary hypotheses. The first is whether the particular situation at hand is one in which people might reason as members of a team in general, and the second is whether the experimenter has correctly specified the goals that the members of the team try to achieve. These may involve assuming particular answers to the "when do people reason as members of a team?" and the "what do people do when they reason as members of a team?" questions that we identified above. Also, if decision-makers do not follow individualistic best-response reasoning in certain situations, we need to be able to distinguish team reasoning from other possible modes of reasoning that they may choose to endorse, e.g., regret minimization or ambiguity aversion, or from factors that influence decisions, like risk aversion.

Despite these difficulties, a number of relatively recent empirical studies have been carried out in an attempt to test the theory of team reasoning. Since the aim is to test the theory of team reasoning tout court, the experiments use situations where it is naturally invoked as an explanation of actual play. They can be broadly divided into two groups: those that focus on team reasoning where it resolves a Nash equilibrium selection problem (coordination problems) and those that focus on team reasoning where it selects outcomes that are not Nash equilibria (as in the Prisoner's Dilemma). We will review both types of studies in turn. The focus is on pitting team reasoning against other explanations of coordination and cooperation in these games, so experimenters hope that the outcome that they identify as the team goal is uncontroversial, although we will see that sometimes there is room for dispute.

5. Tests Based on Nash Equilibrium Selection

8 In Karpus and Radzvilas' function payoffs are first normalized so that, for each player, the least and the most preferred outcomes in a game are associated with payoff values 0 and 100 respectively.

9 There is a connection between the notion of mutual benefit in team play and Gauthier's (2013) idea of rational cooperation discussed earlier. For Gauthier, rational cooperation is attained by maximizing the minimum level of personal gains across players relative to threshold points beyond which individuals would not cooperate. This is similar to the way the maximally mutually beneficial outcomes are identified using the function of team's goals presented by Karpus and Radzvilas (2016). Gauthier, however, does not provide a clear characterization of what the aforementioned threshold points are and his justification for rational cooperation is based on the idea of "social morality" (see earlier discussion in Section 2) rather than the interacting players attempting to resolve games in mutually advantageous ways.

The first category of experiments involves games with multiple Nash equilibria where non-equilibrium outcomes yield no payoffs to the interacting players. As such, they are Nash equilibrium coordination games in which players try to coordinate their actions on one of the available equilibria in order to attain positive payoffs. Team reasoning is said to single out one of the equilibria as uniquely optimal for a team and is tested against other possible modes of reasoning that may be at play. The dominant alternative explanation of behaviour in these experiments (to that of the theory of team reasoning) is assumed to be cognitive hierarchy theory, which posits the existence of individualistic best-response reasoners who differ in their beliefs about what other players are going to do in games. The level-0 decision-makers are said not to reason much at all when playing games and choose any of the available options at random, i.e. they play each available option with equal probability.[10] The level-1 reasoners assume everybody else to be cognitive level-0 and best-respond to the level-0 decision-makers' strategy. The level-2 reasoners assume everybody else to be cognitive level-1 and, similarly, best-respond to the expected strategies of a level-1 player, and so on for higher level cognitive types. Although in principle the cognitive hierarchy theory allows for any number of cognitive types (where each type assumes other players to be of one level lesser type than themselves), in practice it is usually assumed that most decision-makers are level-1 or level-2 reasoners.

	A	*B*	*C*	*D*
A	10, 10	0, 0	0, 0	0, 0
B	0, 0	10, 10	0, 0	0, 0
C	0, 0	0, 0	10, 10	0, 0
D	0, 0	0, 0	0, 0	9, 9

Figure 5: An example of a game from the Amsterdam experiment in the form of a game matrix

Bardsley et al. (2010) conducted a similar experiment at two separate locations—one in Amsterdam and one in Nottingham—using a set of Nash equilibrium coordination games described above. An example is given in Figure 5. In this game, the best response to a player who chooses any of the options with equal probability is to pick one of the options associated with the payoff of 10. This is because somebody who chooses at random is expected to play each of the four available strategies with equal probability of ¼. As such, the expected payoff from choosing *A*, *B* or *C* (when the co-player chooses at random) is 10 x ¼ = 2.5 while the expected payoff from choosing *D* is 9 x ¼ = 2.25. Therefore, a level-1 reasoner would never choose *D*. From this it follows that level-2 reasoners would never choose *D* either, since they are best-responding to the choice of level-1 types, and would, hence, also pick one of the options associated with the payoff of 10.

Bardsley et al. (2010) hypothesized that team reasoners would choose option *D* due to the uniqueness of the outcome (*D*, *D*) and the indistinguishability of the outcomes (*A*, *A*), (*B*, *B*) and (*C*, *C*), which allows players to easily coordinate their actions. In the experiment, games were not presented to participants in the form of a matrix as shown in Figure 5 and there was no way to distinguish between the available strategies and

10 This is assumed in the most frequently occurring version of the cognitive hierarchy theory. For a sightly different version, where level-0 decision-makers randomize between all of the available options, but assign slightly higher probability to the play of the strategy associated with the highest personal payoff or that with the most salient label, see, for example, Crawford et al. (2008).

outcomes other than in terms of payoffs that the players would attain if they managed to successfully coordinate their choices. For example, the outcome (*A*, *A*) could not be identified as being unique due to its top-left position in the matrix or because of being associated with choice options labelled with the first letter of alphabet. Notice that (*D*, *D*) is not Pareto efficient: it is inferior to the outcomes (*A*, *A*), (*B*, *B*) and (*C*, *C*). The reasoning behind the suggestion that team-reasoning decision-makers would opt for the outcome (*D*, *D*) is that, in the case of the three indistinguishable outcomes (*A*, *A*), (*B*, *B*) and (*C*, *C*), a player can only "pick" one of them and hope that the other player would "pick" the same one, whereas in the case of the outcome (*D*, *D*), a player is "choosing" the corresponding strategy *D* because of the uniqueness of that outcome. If both players pick one of the three indistinguishable outcomes, there is a ⅓ chance that they will pick the same one, whereas if they both choose strategy *D*, they can be sure of attaining the outcome *(D, D)*. So the expected payoff from trying to coordinate on one of the outcomes (*A*, *A*), (*B*, *B*) or (*C*, *C*) for a team-reasoning decision-maker is 3⅓ while the certain payoff from coordinating on the outcome (*D*, *D*) is 9. (See Gold and Sugden's introduction to Bacharach (2006) for more on this idea.) To put this differently, it may be said that ex ante, before the uncertainty about the other player's action is resolved and when players take into account the likelihood of coordinating their actions in the computation of their expected payoffs, the optimal outcome in terms of Pareto efficiency is (*D*, *D*). Ex post, once the game has been played, the three outcomes (*A*, *A*), (*B*, *B*) and (*C*, *C*) Pareto dominate (*D*, *D*).[11]

The experimental results, though showing a clear deviation from individualistic best-response reasoning (assuming that it would not discriminate among the available Nash equilibria), are different in the Amsterdam and the Nottingham experiments. The results from Amsterdam seem to suggest the presence of team reasoning rather than cognitive hierarchy reasoning, whereas the results from Nottingham tend to suggest the opposite. In addition to making choices in numerical coordination games, such as the one illustrated above, both experiments asked the participants to complete other non-numerical "text" tasks. These differed between the two experiments and the authors speculate that there may have been spillover effects from the text tasks on the modes of reasoning used in the numerical coordination tasks. In Amsterdam, text tasks involved picking the odd one out, so participants may have tended to pick strategies that were associated with outcomes appearing as odd ones out in the number tasks, while in Nottingham text tasks gave more scope for picking favourites, so participants may have tended to focus on outcomes that were associated with their favourite payoffs.

Another pair of experiments that focus on the Nash equilibrium selection problem was carried out by Faillo et al. (2013, 2016). Both experiments presented the participants with two-player games, in which they had to pick one of three options presented as segments of a pie. See Figure 6 for an example. Upon successfully coordinating on one of the three pie segments participants received positive payoffs, though these were not always the same for the two players. In the game of Figure 6, if we call the top left slice R1, the top right slice R2, and the bottom slice R3, then the outcomes (*R1*, *R1*), (*R2*, *R2*) and (*R3*, *R3*) yielded pairs of payoffs (9, 10), (10, 9) and (9, 9) to the two players respectively. A representation of this game using a game matrix is given in Figure 7.

11 Note that this idea is based on an implicit assumption that decision-makers are not extremely risk-loving. If the interacting decision-makers both preferred the ⅓ chance of receiving a payoff of 10 to a certainty of the payoff of 9 (i.e., if they both were extremely risk-seeking), then the team-optimal choice may be to pick one of *A*, *B*, or *C* in the hope of coordinating on one of the outcomes (*A*, *A*), (*B*, *B*) or (*C*, *C*) respectively.

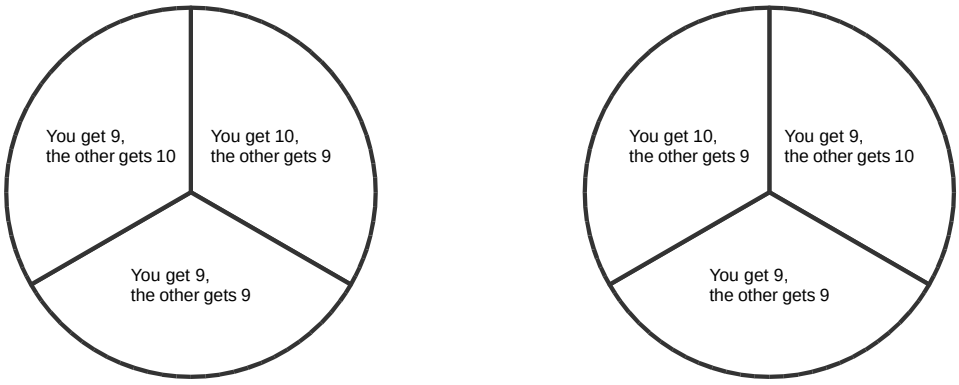

Figure 6: An example of a 3x3 pie game as seen by two interacting players

	R1	*R2*	*R3*
R1	9, 10	0, 0	0, 0
R2	0, 0	10, 9	0, 0
R3	0, 0	0, 0	9, 9

Figure 7: An example of a 3x3 pie game in the form of a game matrix

Like the experiments of Bardsley et al. (2010), these experiments were designed to pit the theory of team reasoning against cognitive hierarchy theory. Faillo et al. (2013, 2016) also followed Bardsley et al. in hypothesizing that team reasoners would take into account the probability of successful coordination when working out the expected payoffs associated with the available options. Pairs of Nash equilibria counted as indistinguishable from the perspective of team reasoning when they were symmetric in terms of payoffs to the two players, such as the pair of outcomes (*R1*, *R1*) and (*R2*, *R2*) in the above example. In fact, the outcomes (*R1*, *R1*) and (*R2*, *R2*) were indistinguishable in all games in the two experiments and the team optimal choice was always associated with the attainment of the outcome (*R3*, *R3*). (The labels *R1*, *R2* and *R3* were hidden from participants and the positions of pie slices were varied across three different treatment groups. The statistical analysis of results showed no significant effects of pie slice positions on the choice of *R3* versus *R1* or *R2*.)

Table 1 summarizes the results of Faillo et al. (2013).[12] Team reasoning is a good predictor in 7 out of the 11 games, where the modal choice was the option *R3*. The observed choices in the remaining 4 games (in addition to 3 of the games in which the theory of team reasoning is a good predictor) can be explained by cognitive hierarchy theory.[13] As such, the results of the experiment are somewhat mixed. Faillo et al. (2013) conclude that team reasoning fails when it predicts the choice of a slice that is ex post Pareto dominated by

12 The type of pie games used and the conclusions drawn in the two experiments are quite similar. We here focus on the results reported in the first study.

13 For example, in the game G3 the cognitive hierarchy theory predicts level-1 reasoners will play the strategy associated with the highest personal payoff. This is the option *R1* for the player who receives the payoff of 10 from the outcome (*R1*, *R1*) and the option *R2* for the player who receives the payoff of 10 from the outcome (*R2*, *R2*). Thus, the level-2 reasoners' best response strategies to the choices of level-1 types will be a mixture of options *R1* and *R2*, depending on which player they are. As a result, the cognitive hierarchy theory predicts a mixture of *R1* and *R2* choices with no play of *R3*.

the other two and this is not compensated by greater equality (games G3, G5, and G7) as well as when the team-optimal outcome yields less equal payoffs than the other options and this is not compensated by Pareto superiority (G10). They suggest that we need a more general theory of team reasoning and offer two ways in which the theory could be amended to explain their results. One is to incorporate the circumstances of group identification (one of the auxiliary hypotheses in any test of team reasoning, as explained above). Ex post Pareto dominance and equality may play an important role in group identification, in which case ex ante Pareto dominance will not be sufficient to trigger team reasoning by itself. The other is to accept that people may not achieve the level of reasoning "sophistication" that would allow them to identify the optimality of the ex ante Pareto efficiency. "Naive" team reasoners may want to pursue the group interest but, because they do not identify the uniqueness of the outcome (*R3*, *R3*), they only use ex post Pareto efficiency and equality of payoffs (when it is not dominated in terms of Pareto efficiency) when determining what the group should do.

Although the aim of this experiment was not to test the claim that game harmony predicts team reasoning, it is clear that the results do not support that idea. Whilst there is a high level of team reasoning in game G6, which has perfect alignment of payoffs, G5 also has perfect alignment of payoffs but relatively little team reasoning. In contrast, G4 and G9 have lower levels of payoff alignment but high levels of team reasoning. So the predictions that payoff alignment leads to team reasoning and payoff conflicts mitigate team reasoning is not supported by this set of games.

Game	**Payoffs**			**Results, %**		
	(*R1*, *R1*)	(*R2*, *R2*)	(*R3*, *R3*)	*R1*	*R2*	*R3*
G1	(9, 10)	(10, 9)	(9, 9)	[CH] 14%	[CH] 11%	74%
G2	(9, 10)	(10, 9)	(11, 11)	0%	1%	[CH] 99%
G3	(9, 10)	(10, 9)	(9, 8)	[CH] 51%	[CH] 45%	4%
G4	(9, 10)	(10, 9)	(11, 10)	16%	4%	[CH] 80%
G5	(10, 10)	(10, 10)	(9, 9)	[CH] 48%	[CH] 34%	18%
G6	(10, 10)	(10, 10)	(11, 11)	1%	3%	[CH] 96%
G7	(10, 10)	(10, 10)	(9, 8)	[CH] 51%	[CH] 31%	18%
G8	(10, 10)	(10, 10)	(11, 10)	26%	22%	[CH] 52%
G9	(9, 12)	(12, 9)	(10, 11)	[CH] 16%	[CH] 11%	73%
G10	(10, 10)	(10, 10)	(11, 9)	[CH] 43%	[CH] 27%	[CH] 30%
G11	(9, 11)	(11, 9)	(10, 10)	[CH] 6%	[CH] 7%	86%

Table 1: Summary of Faillo et al. (2013) results, showing the percentage of subjects making each choice in each game; in all games, team reasoning is assumed to predict the choice of *R3* (highlighted in grey); choices predicted by cognitive hierarchy theory are indicated by [CH]

There is another way to explain the results of Faillo et al. (2013), which challenges their assumption about what the team takes as its goals. Suppose that team-reasoning decision-makers first establish the optimal outcomes from the perspective of the team by identifying those outcomes that maximize the extent of mutual advantage as suggested by Karpus and Radzvilas (2016). These outcomes are always efficient in the weak sense of Pareto efficiency. (Recall that an outcome of a game is said to be Pareto efficient in the weak sense of Pareto efficiency, if there is no alternative that is strictly preferred to it by every player in the

game.) The players then seek ways to coordinate their actions on one of the outcomes in the identified set using unique features of some outcome (if an outcome with unique features exists) as a possible coordinating device. This approach could explain choices observed in games G3, G5 and G7 in addition to the 7 explained originally.[14] For example, in the game G5, the outcomes (*R1*, *R1*) and (*R2*, *R2*) are strictly preferred to the outcome (*R3*, *R3*) by both players and, hence, by Karpus and Radzvilas's approach, they are deemed optimal from the perspective of the team. Since there is no further way to discriminate between the latter two outcomes, team-reasoning decision-makers, according to this interpretation, end up playing a mixture of the two. In the game G1, on the other hand, none of the three equilibria can be excluded from the set of team-optimal outcomes, since they all provide the same extent of mutual benefit to the two players. The outcome (*R3*, *R3*), however, is unique in this set and team-reasoning decision-makers, therefore, opt for this outcome.

6. Tests Involving Non-Nash Equilibrium Play

We now turn to tests of team reasoning where a team selects outcomes that are not Nash equilibria. Although any empirical study of games in which team reasoning prescribes non-equilibrium play can be seen as a test of the theory (e.g., any test involving the Prisoner's Dilemma game) reviewing all historical studies of games of this type is beyond the scope of this chapter. Instead, we will focus on two relatively recent experiments that specifically refer to the theory of team reasoning in their hypotheses.

Colman et al. (2008) conducted an experiment (Experiment 2 in their paper) with five one-shot, 3x3, two-player games with symmetric payoffs (i.e., each game was played once, there were three strategies available to each player and the payoffs to the two players were symmetric). All games had a unique Nash equilibrium and a unique non-equilibrium outcome that was optimal from the perspective of a team. The study assumed team play to be the maximization of the average of players' payoffs. The predicted outcomes, however, would be the same using any of the accounts of a team's goals discussed in Section 3 above. An example of one of their games is given in Figure 8, where the unique Nash equilibrium is the outcome (*E*, *E*) and the optimal outcome for a team is (*C*, *C*).

	C	*D*	*E*
C	8, 8	5, 9	5, 5
D	9, 5	7, 7	5, 9
E	5, 5	9, 5	6, 6

Figure 8: An example of a 3x3 game form
Colman et al. (2008)

The results of the experiment show that in four games (out of five) slightly more than half of participants chose strategies that were associated with the team-optimal outcome and in one of the games this share was higher (86%). An important feature of all games in the experiment was that the team-optimal outcome was superior to the Nash equilibrium in terms of Pareto efficiency (which makes these cases somewhat similar to the Prisoner's Dilemma game). This may suggest that in cases of one-shot interactions with unique Nash equilibria that are not Pareto efficient about half of decision-makers reason as members of a

14 In the game G10 this approach establishes team-optimal outcomes to be (*R1*, *R1*) and (*R2*, *R2*), thus predicting no play of (*R3*, *R3*).

team and play accordingly.

In a different experiment, Colman et al. (2014) used another set of eight one-shot, 3x3 and four 4x4, two-player games where every game (with the exception of one) contained a unique Nash equilibrium and distinct but also unique non-equilibrium predictions based on the theory of team reasoning and cognitive hierarchy theory.[15] Examples of the games are given in Figures 9 and 10.

	A	*B*	*C*
A	3, 3	1, 1	0, 2
B	1, 1	1, 4	3, 0
C	0, 0	2, 1	**2, 5**

Figure 9: An example of a 3x3 game from Colman et al. (2014)

	A	*B*	*C*	*D*
A	**4, 4**	2, 0	3, 2	1, 5
B	2, 2	3, 3	2, 2	2, 0
C	4, 3	2, 4	2, 5	3, 2
D	5, 2	0, 3	0, 0	1, 1

Figure 10: An example of a 4x4 game from Colman et al. (2014)

The study assumed team play to be associated with the maximization of the average of players' payoffs (the corresponding outcomes are indicated in bold in Figures 9 and 10). Sugden's (2015) notion of mutual benefit (see Section 3 above) and the function of team's goals discussed by Karpus and Radzvilas (2016) would yield different predictions in some of these games (for example, in Figure 9 the optimal outcome for the team based on the notion of maximal mutual advantage would be the outcome (*A*, *A*)). The results of the experiment are mixed, with at least two out of three or three out of four available strategies played quite frequently. This, combined with uncertainty about which outcome is the team reasoning solution, makes it difficult to identify which mode of reasoning predominates.

Furthermore, many of these results could be explained by a combination of level-0 and level-1 reasoning, which simply corresponds to random picking and best-responding to a random choice of the other player (also see Sugden, 2008, who suggests that these results would be obtained with a population consisting of 50% team-reasoners, 40% level-1 and 10% level-0 types). There is some evidence that increasing the difficulty of a task increases the amount of randomizing (Bardsley and Ule, 2014).[16] Since the games that Colman et al. (2014) used had numerous strategies and non-symmetric variable payoffs, and appear to be quite complex and cognitively demanding in the identification of rational outcomes, random picking and the principle of insufficient reason (which means best-responding to a random choice) may provide a good explanation of the actual choices.

Conclusion and Further Directions

In this chapter we reviewed some of the recent developments of the theory of team reasoning in games. Since its early developments, which were triggered by orthodox game theory's inability to definitively resolve certain types of games with multiple Nash equilibria (such as the Hi-Lo game) and explain out-of-

15 The study also refers to a mode of reasoning called the strong Stackelberg reasoning, but, since the latter always predicts the play of a Nash equilibrium, in all (but one) of the studied cases it is indistinguishable from individualistic best-responding.

16 Bardsley and Ule (2014) test for team reasoning vs. cognitive hierarchy and the principle of insufficient reason in a 'risky' coordination game, where players may experience losses as well as gains. Their results favour team reasoning. (We learned of this paper too late to review it in detail in this chapter.)

equilibrium play in others (such as the Prisoner's Dilemma game), the theory has advanced in a number of different directions. From the theoretical point of view, different answers were proposed to the two fundamental questions that the theory of team reasoning needs to address: “when do people reason as members of a team?” and “what is it that they do when they reason in this way?”. In response to the first question, it has been suggested that the mode of reasoning that an individual decision-maker adopts may depend on that decision-maker's psychological make-up, it may be endorsed by the decision-maker depending on a number of conditions that need to be satisfied, such as the assurance of others' participation in team play and the notion of mutual benefit, or it may be a result of rational deliberation about which mode of reasoning is instrumentally most useful in any given situation. In response to the second question, one aspect that differentiates the suggested answers is whether they allow team play to advocate a potential sacrifice of some members of a team for the benefit of others.

The results of the nascent developments in empirical testing of the theory, a number of which we reviewed in the second part of this chapter, are, at best, mixed and further research in this field is needed. The studies start from the assumption that the games they use are situations where people could be expected to team reason. Nevertheless, some of them can be seen as providing indicative answers to the first question, “when do people reason as members of a team?”, because they arguably identify circumstances in which people are likely to team reason. One interpretation of Faillo et al. (2013) is that ex post Pareto dominance and equality play an important role in group identification. One interpretation of Colman et al. (2014) is that the team reasoning outcome needs to be simple and clear, as complex or cognitively demanding games lead people to randomise. However, these are speculative hypotheses which were developed post hoc to explain the experimental results and they still need to be put to the test. None of the experiments aim to test the mechanism by which people adopt team reasoning: whether it is caused by a psychological process or a decision, and the role of assurance and players' beliefs about what others will do.

With regards to the second question, “what is it that team reasoning decision-makers strive for?”, in some of the games that have been studied, the predictions of the various functional representations of team interests coincide. This is often so in Nash equilibrium coordination games. But even then some differences are possible (recall, for example, the interpretation of results discussed by Faillo et al. (2013) based on ex ante vs. ex post optimality of the considered outcomes and the idea of coordination among outcomes that are maximally mutually beneficial). In more complex scenarios studied by Colman et al. (2008, 2014), the differences between various predictions of team play loom larger, which may therefore offer a better ground to test the competing assumptions about team reasoning decision-makers' goals, keeping in mind that if games get too complex that may mitigate against team reasoning.

Any experimental test of the theory of team reasoning is complicated by the multiplicity of hypotheses that are to be tested simultaneously in connection with the above questions. It may thus be necessary to apply methods that go beyond mere observation of decision-makers' choices in games, e.g., asking the participants to explain the reasons behind their choices, or encouraging the adoption of one or another mode of reasoning through the use of additional pre-play tasks. One possibility for further experimental work is to study how priming group or individualistic thinking affects people's choices in simple Nash equilibrium coordination games where the team-optimal outcome seems to be obvious. Such a test would accord with a number of versions of the theory with respect to what is assumed to trigger the shift in individuals' adopted mode of reasoning as well as a number of suggested functional representations of a

team's goals.

Acknowledgements

We are extremely grateful to Nicholas Bardsley, Julian Kiverstein, Guglielmo Feis and Mantas Radzvilas for their invaluable suggestions which we used to improve earlier versions of this work. We are also grateful to James Thom for a large number of insightful discussions on the topic during the course of preparing this chapter. Our 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

Bacharach, M. (1997): ''We' Equilibria: A Variable Frame Theory of Cooperation', Oxford: Institute of Economics and Statistics, University of Oxford, 30.

Bacharach, M. (1999): 'Interactive Team Reasoning: A Contribution to the Theory of Co-operation', *Research in Economics*, 53, pp. 117-147.

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

Bardsley, N. (2000): *Theoretical and Empirical Investigation of Nonselfish Behaviour: The Case of Contributions to Public Goods'*, University of East Anglia.

Bardsley, N. (2001): 'Collective Reasoning: A Critique of Martin Hollis's Position', *Critical Review of International Social and Political Philosophy*, 4, pp. 171-192.

Bardsley, N., Mehta, J., Starmer, C. and Sugden, R. (2010): 'Explaining Focal Points: Cognitive Hierarchy Theory versus Team Reasoning', *The Economic Journal*, 120, pp. 40-79.

Bardsley, N. and Ule, Aljaz (2014): 'Focal Points Revisited: Team Reasoning, the Principle of Insufficient Reason and Cognitive Hierarchy', Munich Personal RePEc Archive (MPRA) Working Paper No. 58256.

Colman, A. M., Pulford, B. D. and Rose, J. (2008): 'Collective Rationality in Interactive Decisions: Evidence for Team Reasoning', *Acta Psychologica*, 128, pp. 387-397.

Colman, A. M., Pulford, B. D. and Lawrence, C. L. (2014): 'Explaining Strategic Cooperation: Cognitive Hierarchy Theory, Strong Stackelberg Reasoning, and Team Reasoning', *Decision*, 1, pp. 35-58.

Crawford, V. P., Gneezy, U. and Rottenstreich, Y. (2008): 'The Power of Focal Points is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures', *The American Economic Review*, 98, pp. 1443-1458.

Cubitt, R. P. and Sugden, R. (2003): 'Common Knowledge, Salience and Convention: A Reconstruction of David Lewis' Game Theory', *Economics and Philosophy*, 19, pp. 175-210.

Faillo, M., Smerilli, A. and Sugden, R. (2013): 'The Roles of Level-k and Team Reasoning in Solving Coordination Games', Cognitive and Experimental Economics Laboratory Working Paper (No. 6-13), Department of Economics, University of Trento, Italy.

Faillo, M., Smerilli, A. and Sugden, R. (2016): 'Can a Single Theory Explain Coordination? An Experiment on Alternative Modes of Reasoning and the Conditions Under Which They Are Used', CBESS [Centre for Behavioural and Experimental Social Science] Working Paper 16-01, University of East Anglia.

Gauthier, D. (1986): *Morals by Agreement*, Oxford: Oxford University Press.

Gauthier, D. (2013): 'Twenty-Five On', *Ethics*, 123, pp. 601-624.

Gold, N. (2012): 'Team Reasoning, Framing and Cooperation', in S. Okasha and K. Binmore (eds), 'Evolution and Rationality: Decisions, Co-operation and Strategic Behaviour', Cambridge: Cambridge University Press, ch. 9, pp. 185-212.

Gold, N. (in press): 'Team Reasoning: Controversies and Open Research Questions', in K. Ludwig and M. Jankovic (eds), 'Routledge Handbook of Collective Intentionality', Routledge.

Gold, N. and Sugden, R. (2007a): 'Collective Intentions and Team Agency', *Journal of Philosophy*, 104, pp. 109-137.

Gold, N. and Sugden, R. (2007b): 'Theories of Team Agency', in F. Peter and H. B. Schmid (eds), 'Rationality and Commitment', Oxford: Oxford University Press, pp. 280-312.

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

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

Kacelnik, A. (2006): 'Meanings of Rationality', in S. Hurley and M. Nudds (eds), 'Rational Animals?', Oxford: Oxford University Press, pp. 87-106.

Karpus, J. and Radzvilas, M. (2016): 'Team Reasoning and a Rank-Based Function of Team Interests', manuscript under review.

Lewis, D. (1969): *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.

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

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

Smerilli, A. (2012): 'We-Thinking and Vacillation Between Frames: Filling a Gap in Bacharach's Theory', *Theory and Decision*, 73, pp. 539-560.

Sugden, R. (1993): 'Thinking as a Team: Towards an Explanation of Nonselfish Behavior', *Social Philosophy and Policy*, 10, pp. 69-89.

Sugden, R. (2000): 'Team Preferences', *Economics and Philosophy*, 16, pp. 175-204.

Sugden, R. (2003): 'The Logic of Team Reasoning', *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, 6, pp. 165-181.

Sugden, R. (2008): 'Nash Equilibrium, Team Reasoning and Cognitive Hierarchy Theory', *Acta Psychologica*, 128, pp. 402-404.

Sugden, R. (2011): 'Mutual Advantage, Conventions and Team Reasoning', *International Review of Economics*, 58, pp. 9-20.

Sugden, R. (2015): 'Team Reasoning and Intentional Cooperation for Mutual Benefit', *Journal of Social Ontology*, 1, pp. 143-166.

Tversky, A. and Kahneman, D. (1981): 'The Framing of Decisions and the Psychology of Choice', *Science*, 211, pp. 453-458.

Zizzo, D. J. and Tan, J. H. (2007): 'Perceived Harmony, Similarity and Cooperation in 2× 2 Games: An Experimental Study', *Journal of Economic Psychology*, 28, pp. 365-386.