Team Reasoning, Framing and Cooperation

Cooperation is puzzling for orthodox game theory because, although cooperation is arguably rational and a substantial number of people cooperate in real life, game theory cannot explain why cooperation is rational nor predict that rational players will cooperate in one-shot games. Solving these puzzles of cooperation is one of the motivations behind the development of a body of decision-theoretic literature on *team agency.*[1] This seeks to extend standard game theory, where each individual asks separately 'What should I do?', to allow teams of individuals to count as agents and for players to ask the question 'What should we do?'. This leads to *team reasoning*, a distinctive mode of reasoning that is used by members of teams, and which may result in cooperative actions. The basic idea is that, when an individual reasons as a member of a team, she considers which *combination* of actions by members of the team would best promote the team's objective, and then performs her part of that combination. The rationality of each individual's action derives from the rationality of the joint action of the team.

One can distinguish between the generic idea of team reasoning and specific versions proposed by different people. The most fully developed theories of team reasoning are those of Michael Bacharach (1999; 2006) and Robert Sugden (1993; 2003). They differ in important ways regarding: what happens when there is not common knowledge of group membership, what the group agent should take as its goals and how group agency comes about. In this chapter, I will explore some of the ramifications of each of these differences.

After showing how a basic version of team reasoning with common knowledge solves problems of cooperation (sections 1 and 2), by comparing what Bacharach and Sugden advocate that players should do in the prisoner's dilemma when there is not common knowledge of group identification, I will clarify the concept of a group goal in each of their theories (section 3). By comparing how group agency comes about, and the role of framing in each of Bacharach and Sugden's theories, I will offer some insights into framing in decision-making in general (section 4). Finally, I compare team reasoning with payoff-transformation theories of cooperation (section 5).

One can draw an analogy between rational choice theory and natural selection. Whilst the rational agent of decision theory maximise utility, evolutionary processes maximise fitness. There is also an analogy between team reasoning and group selection: in team reasoning, the team (or the team utility function) is an additional primitive to those found in standard decision theory, whilst theories of group selection add groups to the ontology of evolutionary biology. At points throughout the paper I will draw the analogy. However, note that it is only a structural analogy, suggesting a common mathematical framework, and, as I show in section 5, the truth or falsity of group selection has no bearing on the truth or falsity of team reasoning and vice versa.

---

[1] See Hodgson (1967), Regan (1980) Gilbert (1987), Hurley (1989), Sugden (1993; 2003), Hollis (1998), Bacharach (1999; 2006), and Anderson (2001).

1. The Puzzle of Cooperation: Stag Hunts and Prisoner's Dilemmas

Common models of cooperation are examples of what Michael Bacharach (2006) called *games with scope for common gain*, or *games with scope* for short. In games with scope, there exists a Nash equilibrium whose outcome could be improved upon for both players if at least one of them plays a different strategy. (Kollock (1998a) gives this as the definition of a *social dilemma*.) Games with scope offer the possibility of mutual benefit and, hence, we may think of them as offering the possibility of cooperation. Two games that are closely identified with the problem of cooperation are the prisoner's dilemma and the stag hunt.[2]

The prisoner's dilemma, whose generic form is illustrated in figure 1, has been the *locus classicus* in the study of cooperation (e.g. Taylor (1987)). In specifying the payoffs of this game, we require only that they are symmetrical between the players and that they satisfy two inequalities. The inequality $a > b > c > d$ encapsulates the central features of the Prisoner's Dilemma: for each player, the best outcome is that in which he chooses *defect* and his opponent chooses *cooperate*; the outcome in which both choose *cooperate* is ranked second; the outcome in which both choose *defect* is ranked third; and the outcome in which he chooses *cooperate* and his opponent chooses *defect* is the worst of all. The inequality $b > (a + d)/2$ stipulates that each player prefers a situation in which both players choose *cooperate* to one in which one player chooses *cooperate* and the other chooses *defect*, each player being equally likely to be the free-rider. This condition is usually treated as a defining feature of the Prisoner's Dilemma.

		Player 2	
		cooperate	*defect*
Player 1	*cooperate*	*b*, *b*	*d*, *a*
	defect	*a*, *d*	*c*, *c*

$a > b > c > d$

$b > (a + d)/2$

Figure 1: The Prisoner's Dilemma

For each player, *defect* strictly dominates *cooperate*. Each individual player can reason to the conclusion, 'The action that gives the best result *for me* is *defect*'. Thus, in its explanatory form, conventional game theory predicts that both players will choose *defect*. In its normative form, it recommends *defect* to both players. Yet both would be better off if each chose *cooperate* instead

[2] Games with scope also include co-ordination games where the equilibria have different payoffs. One example is the game of Hi-Lo, which has also attracted attention in the team reasoning literature, e.g. Sugden (1993), Bacharach (2006), Gold and Sugden (2007), Bardsley (2007), Gold and Sugden (2008).

of *defect*. Thus, each player can also reason to the conclusion: 'The pair of actions that gives the best result *for us* is not (*defect*, *defect*)'.[3]

The game theoretic puzzle of cooperation is isomorphic to the biological problem of its evolution, often referred to by biologists as the problem of the evolution of altruism, with the payoffs now being units of fitness (often taken to be number of offspring). Natural selection favours individuals with a relatively high fitness compared to the rest of the population. Biological altruism, by definition, involves advantaging others at a cost to self. So the altruist is at an evolutionary disadvantage compared to non-altruists and altruism should never evolve.

Arguably, many situations which have been analysed as prisoner's dilemmas can be better thought of as stag hunts.[4] Brian Skyrms (2004) considers the stag hunt, rather than the prisoner's dilemma, to be the paradigm problem of cooperation. The *stag hunt* is is named for Rousseau's story about the beginnings of human society and the dilemma faced by hunters in the state of nature, as they begin to learn the benefits of cooperation. The generic payoff matrix for a stag hunt is given in figure 2. For the story, consider two hunters who can each chase either stag or rabbit. Between the two of them, they could catch a stag which provides a large amount of meat, but if one of them chases stag alone then he will fail and go hungry. Rabbits can be caught by one person but provide less than half the meat of a stag. What each hunter cares about is how much meat he gets, so each prefers the outcome where both hunt stag to hunting rabbit (together or alone). In turn, hunting rabbit is preferred to hunting stag alone.

The game has two Nash equilibria, the outcome where both hunt stag and the outcome where both hunt rabbit. In other words, if player 2 hunts stag then player 1 does best if she also hunts stag and if player 2 hunts rabbit then player 1 does best if she hunts rabbit, and vice versa.

		P2	
		stag	*rabbit*
P1	*stag*	*s, s*	*q, r*
	rabbit	*r, q*	*r, r*

s > r > q

Figure 2: A stag hunt

We might think of hunting stag as cooperating. Unlike the prisoner's dilemma, if one player cooperates then the other player is best off cooperating too. Both cooperating is an equilibrium.

---

[3] In order to conclude that (*cooperate*, *cooperate*) is the best pair of strategies for them, the players have to judge the payoff combinations (a, d) and (d, a) to be worse 'for them' than (b, b).

[4] See Joshi *et al* (2005) for an illustration from everyday life, Kollock (1998b) for an example from the lab, and Kollock (1998a) and Skyrms (2004) for theoretical arguments.

Further, hunting stag is better for both players than hunting rabbit so it may seem obvious that the players should coordinate on the equilibrium that they prefer.

However from the assumptions that the players are perfectly rational (in the normal decision theoretic sense of maximising expected payoff) and that they have common knowledge of their rationality, we cannot deduce that each will choose *stag.* Or, expressing the same idea in normative terms, there is no sequence of steps of valid reasoning by which perfectly rational players can arrive at the conclusion that they ought to choose *stag.* This is because, from the assumption of rationality, all we can infer is that each player chooses the strategy that maximises her expected payoff, given her beliefs about what the other player will do. All we can say in favour of *stag* is that, if either player expects the other to choose *stag*, then it is rational for the first player to choose *stag* too; thus, a shared expectation of *stag*-choosing is self-fulfilling among rational players. But exactly the same can be said about *rabbit*. And, in a one-shot game with common knowledge of rationality, there is no prior reason to expect the play of either *stag* or *rabbit*.[5]

This is an example of the problem of *equilibrium selection.* Classical game theory cannot make a unique prediction of play unless it is supplemented with new criteria to select between the equilibria. So even if the outcome where both play the cooperative strategy is a Nash equilibrium, orthodox game theory cannot predict its play - and sometimes it is not an equilibrium, as in the prisoner's dilemma. Analogously, although normative game theory assumes that if there is a rational way to play the game, then it will be a Nash equilibrium (Nash, 1953), where there is more than one Nash equilibrium standard game theory has no way to advocate one over the other without introducing new assumptions - and, by this criteria, playing a strategy that is not a part of a Nash equilibrium is never rational.

2. Team Reasoning and Cooperation

The prisoner's dilemma and the stag hunt are both games with scope. In the Prisoner's Dilemma, the Nash equilibrium is (*defect, defect*) but each player does better if the outcome is (*cooperate, cooperate*). In the Stag Hunt, (*rabbit*, *rabbit*) is a Nash equilibrium but the players do better under (*stag*, *stag*). The Prisoner's Dilemma seems more intractable because (*cooperate, cooperate*) is not a Nash equilibrium, but standard game theory has nothing to say about when or why hunters chase stag instead of rabbit without adding further equilibrium selection criteria. The equilibrium selection problem is particularly perplexing in the stag hunt. Intuitively, it may seem

[5] The stag hunt is sometimes called an 'assurance' game, as each player would play *stag* if assured that the other player were going to play stag too. It is rational for a player to play *stag* if she believes that the probability that the other player will play stag is greater than (r - q) / (s - q), but the rational players of standard game theory have no reason to believe that.

obvious that each player should choose *stag* because both prefer the outcome of (*stag*, *stag*) to that of (*rabbit*, *rabbit*); but that 'because' has no standing in the formal theory.[6]

The source of both puzzles seems to be located in the mode of reasoning by which, in the standard theory, individuals move from preferences to decisions. In the syntax of game theory, each individual must ask separately 'What should *I* do?' In stag hunt, the game-theoretic answer to this question is indeterminate. In the prisoner's dilemma, the answer is that *defect* should be chosen. Intuitively, however, it seems possible for the players to ask a different question: 'What should *we* do?' In stag hunt, the answer to *this* question is surely: 'Choose (*stag*, *stag*)'. In the Prisoner's Dilemma, 'Choose (*cooperate*, *cooperate*)' seems to be at least credible as an answer.

The reasoning that occurs in game theory is *instrumental practical reasoning*, where conclusions about what an agent ought to do are inferred from premises that include propositions about what the agent is seeking to achieve. Such reasoning is *instrumental* in that it takes the standard of success as given; its conclusions are propositions about what the agent should do in order to be as successful as possible according to that standard.[7] So instrumental practical reasoning presupposes a unit of agency that pursues its own objectives. Standard game theory presumes that the unit of agency is the individual. Theories of team agency generalize game theory to allow that teams can be agents.[8]

The basic idea is that, when an individual reasons as a member of a team, she considers which *combination* of actions by members of the team would best promote the team's objective, and then performs her part of that combination. The rationality of each individual's action derives from the rationality of the joint action of the team. Gold and Sugden show that, in a group of agents, if there is common knowledge that each member group identifies, common knowledge that each member aims to maximise the team payoff function and a unique profile of actions that does this, then each individual can reach the conclusion that she should choose her component of that profile (Gold and Sugden 2007; 2008).

This can lead to cooperation in the prisoner's dilemma and the stag hunt. First consider the prisoner's dilemma. We need to define a payoff function for the group consisting of Player 1 and Player 2. We shall assume that, when a player identifies with a group, she wants to promote the combined interests of its two members, at least in so far as those interests are affected by the

---

[6] However there is a counter-argument which suggests that rational hunters chase rabbit, formalized in Harsanyi and Selten's criteria of risk dominance (Harsanyi & Selten, 1988). Intuitively, hunting rabbit means a player will never get the big prize but it also ensures that she will never go hungry.

[7] Bacharach, in his unpublished *Scientific Synopsis* describing his initial plans for *Beyond Individual Choice*, defines a mode of reasoning as valid in games if it is *success-promoting*: given any game of some very broad class, it yields only choices which tend to produce success, as measured by game payoffs.

[8] Individual reasoning is a special case of team reasoning, where the team has only one member. See the analysis in Gold and Sugden (2007; 2008).

game that is being played. If we assume that the payoff function treats the players symmetrically, we need to specify only three values of this function: the payoff when both players choose *cooperate*, which we denote u_C, the payoff when both choose *defect*, which we denote u_D, and the payoff when one chooses *cooperate* and one chooses *defect*, which we denote u_F (for 'free riding'). It seems unexceptionable to assume that the team payoff function is increasing in individual payoffs, which implies $u_C > u_D$. Given the condition $b > (a + d)/2$, it is natural also to assume $u_C > u_F$. Then the profile of actions by Player 1 and Player 2 that uniquely maximises the team payoff is (*cooperate*, *cooperate*). If there is common knowledge of the rules of the game, each player can use team reasoning to reach the conclusion that she should choose *cooperate* (Gold and Sugden, 2008).

Now consider the stag hunt. Again we need to specify only three values of this function: the payoff when both players choose *stag*, which I denote u_S; the payoff when both choose *rabbit*, which I denote u_R; and the payoff when one chooses *stag* and the other chooses *rabbit,* which I denote u_{SR}. It seems unexceptionable to assume that *U* is increasing in individual payoffs, which implies $u_S > u_R > u_{SR}$. Then (*stag*, *stag*) is the profile that uniquely maximises the payoff function, and so (provided there is common knowledge of the rules of the game), each player can use team reasoning to reach the conclusion that she should choose *stag*.

Thus team reasoning can predict the play of the the payoff dominant equilibrium. Payoff dominance is often used as a criterion for equilibrium selection (e.g. Harsanyi & Selten, 1988). It is intuitively compelling but, previously, it has not been justified by standard game theoretic rationality assumptions. Rather it is an additional supposition, included purely for the purposes of equilibrium selection. In contrast, team reasoning can explain why rational agents play the strategies that lead to the payoff dominant equilibrium (given a reasonable assumption about the team payoff function, discussed further below).

The idea that cooperating is better for the group also plays a fundamental role in the answer provided by theories of 'group' or multi-level selection to the evolutionary puzzle of altruism. These theories allow that natural selection can act on groups as well as individuals. Although altruism puts the individual altruist at an evolutionary disadvantage, it can be advantageous at the level of the group, putting the group at an advantage compared to groups composed of non-altruists. Depending on the relative strength of inter-individual and inter-group evolutionary pressures, altruism may evolve because it is good for the group *(Sober & Wilson, 1998)*.

3. **Team Agency, Group Goals and Expected Utility Theory**

A minimal constraint on being an agent, for the purposes of instrumental practical reasoning, is that the (group) agent has an objective, which can be the basis of instrumental reasoning for its

members. So theories of team reasoning need to assume that there is a group objective.[9] The basic idea of team reasoning places no constraints on this objective but, in order to operationalize the theory, it is necessary to make some assumptions about the group's goals. For example, in the discussion of cooperation above, I assumed that the collective utility function was increasing in individual payoffs, that it was symmetrical and, in the case of the prisoner's dilemma, made an assumption about how it would rank outcomes with different distributions of individual utility.

A key issue is the relation between the individual agents' utility and the group's utility. Theories of team agency differ in the constraints they impose on this relationship. In this section, I compare what Bacharach and Sugden's theories advocate that players should do in the prisoner's dilemma when there is not common knowledge of group identification. I then use the comparison to correct two misconceptions about what the theory of team reasoning is committed to, as regards expected utility maximization and the group goal.

Team reasoning without common knowledge of group membership

Bacharach's theory of 'circumspect team reasoning' was formulated to explain cooperation in situations where there is not common knowledge of group identification. For Bacharach, whether a particular player identifies with a particular group is a matter of 'framing'. A *frame* is the set of concepts a player uses when thinking about her situation. In order to team reason, a player must have the concept 'we' in her frame. Bacharach proposes that the 'we' frame is normally induced or *primed* by games that have the property of *strong interdependence*. Roughly, a game has this property if it has a Nash equilibrium which is Pareto-dominated by the outcome of some feasible strategy profile. So both the prisoner's dilemma and the stag hunt have the property of strong independence. (In the stag hunt, (*rabbit*, *rabbit)* is dominated by (*stag, stag*) - itself a Nash Equilibrium. In the prisoner's dilemma, *(defect, defect)* is Pareto-dominated by *(cooperate, cooperate).*

Although Bacharach proposes that the perception of strong independence increases the probability of group identification, he does not claim that games with this property *invariably* prime the 'we' frame. More specifically:

> **'**In a Prisoner's Dilemma, players might see only, or most powerfully, the feature of common interest and reciprocal dependence which lie in the payoffs on the main diagonal. But they might see the problem in other ways. For example, someone might be struck by the thought that her coplayer is in a position to double-cross her

[9] Extant theories also assume that the group's objective is common knowledge amongst team reasoners, but that assumption could be relaxed.

by playing [*defect*] in the expectation that she will play [*cooperate*]. This perceived feature might inhibit group identification.' (Bacharach, 2006, p.86)[10]

The implication is that the 'we' frame *might* be primed but, alternatively, a player may see the game as one to be played by two separate individual agents.

In Bacharach's theoretical framework, this dualism is best represented in terms of *circumspect team reasoning*. Suppose there is a random process which, independently for each member of the group, determines whether or not that individual identifies with the group. Let ω (where $0 < \omega \leq 1$) be the probability that, for any individual player, the 'we' frame comes to mind; if it does, the player identifies with the group. Then an individual who group-identifies will maximize the *expected value* of the group payoff function given the probabilities that other group members fail to identify. (Any individual who identifies is assumed to also know the value of ω. The idea is that, in coming to frame the situation as a problem 'for us', an individual also gains some sense of how likely it is that another individual would frame it in the same way; in this way, the value of ω becomes common knowledge among those who use this frame.)

We can apply this to the prisoner's dilemma played by the group of Player 1 and Player 2. Define the group payoff function as before. If the 'we' frame comes to mind, with probability ω, the player identifies with the group. Assume that, if this frame does *not* come to mind, the player conceives of herself as a unit of agency and thus, using best-reply reasoning, chooses the dominant strategy *defect*. We can now ask which protocol maximises the group payoff function, given the value of ω. Viewed from within the 'we' frame, the protocol (*defect*, *defect*) gives a payoff of u_D with certainty. Each of the protocols (*cooperate*, *defect*) and (*defect*, *cooperate*) gives an expected payoff of $\omega u_F + (1 - \omega)u_D$. The protocol (*cooperate*, *cooperate*) gives an expected payoff of $\omega^2 u_C + 2\omega(1 - \omega)u_F + (1 - \omega)^2 u_D$. There are two possible cases to consider. If $u_F \geq u_D$, then (*cooperate*, *cooperate*) is the team utility-maximising protocol for all possible values of ω. Alternatively, if $u_D > u_F$, which protocol maximises the team utility function depends on the value of ω. At high values of ω, (*cooperate*, *cooperate*) is uniquely optimal; at low values, the uniquely optimal protocol is (*defect*, *defect*).[11]

If we assume *either* that $u_F \geq u_D$ *or* that the value of ω is high enough to make (*cooperate*, *cooperate*) the uniquely optimal protocol, we have a model in which players of the Prisoner's

---

[10] For Bacharach these features compete to be noticed and it is their relative salience that determines whether or not an agent will group identify. Another natural reading has the agent deliberating about whether or not to use the we-frame. As explained below, reasoning about frames has no place in Bacharach's framework. But Smerilli (2008) takes this 'double-crossing intuition' and proposes an extension to the theory, where agents adjudicate between the outcomes of individual and team reasoning, based on reasoning about deviation from equilibrium.

[11] We can normalise the payoff function by setting $u_C = 1$ and $u_D = 0$. Then, given that $u_F < 0$, the critical value of ω is $\omega^* = 2u_F/(2u_F - 1)$. The protocol (*cooperate*, *cooperate*) is optimal if and only if $\omega \geq \omega^*$, (*defect*, *defect*) is optimal if and only if $\omega \leq \omega^*$. There is no non-zero value of ω at which (*cooperate*, *defect*) or (*defect*, *cooperate*) is optimal.

Dilemma choose *cooperate* if the 'we' frame comes to mind, and *defect* otherwise. Bacharach offers this result as an explanation of the observation that, in one-shot Prisoner's Dilemmas played under experimental conditions, each of *cooperate* and *defect* is usually chosen by a substantial proportion of players. He also sees it as consistent with the fact that there are many people who think it completely obvious that *cooperate* is the only rational choice, while there are also many who feel the same about *defect*.

If $\omega = 1$ then there is common knowledge of group identification amongst group members, if $\omega < 1$ then there is not. Since Bacharach's team reasoners act based on their ex ante probability that other members will group identify, if ω is less than 1 but high enough to make (*cooperate*, *cooperate*) the uniquely optimal protocol, a team reasoner may find ex poste that she has cooperated when the other player has defected. In other words, she may find that she has been 'suckered' - as sometimes happens in the experiments whose results Bacharach sought to explain. This can happen because Bacharach's team reasoners cooperate without assurance that other group members are also team reasoning.[12]

Further, if $u_F \geq u_D$, then the team reasoner will cooperate regardless of whether she expects the other player to group identify, regarding it as better to be suckered by the other player than to defect. For Bacharach, the way the group utility function ranks the off-diagonal payoffs is an open question. In a very brief discussion of the group objective, Bacharach claims that it is likely to be Paretian and to embody principles of fairness (Bacharach, 2006, p.88). However these are not specified as conceptual constraints, but as testable hypotheses. Bacharach thought that group identification could explain why members of nationalistic movements are ready to sacrifice their lives for the cause (Bacharach 2006, footnotes, p.91), so he allowed in principle that the group objective might be welfare-decreasing for some members.

Sugden disagrees with Bacharach on both these points. For Sugden, the purpose of the theory of team reasoning is to explain how people cooperate for mutual advantage, so he takes exception to the idea that the team utility function might make some individuals worse off by their individual lights (for example, ranking u_F above u_D), and to the possibility that a team reasoner might cooperate without assurance that other group members will act likewise. For Sugden, a person should not be made worse off by team reasoning. Hence he would place more constraints on the group objective than Bacharach. In Sugden's theory the team should pursue outcomes that are advantageous to all its members.

---

[12] Bacharach also countenances a version of team reasoning that he calls *restricted team reasoning*, where the team reasoners know in advance that some of the team members will not team reason, and optimise the team utility function as best they can given that some team members will not function. But he prefers circumspect team reasoning because it is more general; we often don't know for certain whether or not other team members will group identify, or what those who do not group identify will do.

It also follows that Sugden's team reasoners will not risk being suckered. On Sugden's account of *mutually assured team reasoning,* a person will not commit herself to team reasoning unless she has assurance that other team members will also act on team reasoning. Sugden uses a theoretical framework in which the central concept is *reason to believe*. To say that a person has reason to believe a proposition *p* is to say that *p* can be inferred from propositions that she accepts as true, using rules of inference that she accepts as valid. In mutually assured team reasoning, team members will not act on the results of team reasoning unless each has reason to believe of all the others that (1) they identify with the group and acknowledge the group payoff function as the objective of the group, and (2) they endorse and act on mutually assured team reasoning. So if Sugden's group members are not sure that they will all cooperate to achieve what they all take to be best for the group, then they will not team reason.

Team reasoning and expected utility theory

Some philosophers are uneasy about the association between team reasoning and expected utility theory. For instance, Raimo Tuomela (2009, p. 298), has complained of Gold and Sugden that 'the only collective goal that they consider and seem to take to be possible in their account is maximization of collective utility'. Hence Tuomela claims that team reasoning is not applicable to all cooperative contexts because its conception of a goal is not applicable. It is quite difficult to work out what Tuomela's complaint is, since there is a trivial sense in which virtually any reasoning can be represented as the maximization of a function so, when it occurs, team reasoning must involve maximization of a team objective function.[13] But his remarks reflect a tendency of some philosophers to wonder whether expected utility theory is the right framework to apply in all occurrences of team agency.

In response to that worry I note that, although Bacharach's theory of team reasoning explicitly incorporates expected utility theory, team reasoning per se is not committed to the idea of expected utility maximization. Whilst Bacharach's team reasoners maximise a group objective given that some members do not group identify, Sugden's cooperate in a mutually advantageous enterprise. Hence Bacharach's theory implies that group members act to get the best outcome for the group even when other members fall short, whereas in Sugden's they only team reason when they are sure that their cooperative actions will be reciprocated. So, in the presence of uncertainty, Sugden's version of team reasoning does not involve maximization of expected utility.

Team reasoning and averaging

---

[13] In making his inference Tuomela also implies that the conscious goal of the agents is to 'maximize utility'. But that would be a mistake because modeling a situation using game theory does not imply that agents have the conscious goal of maximizing utility. I address this mistake elsewhere (Natalie Gold, unpublished manuscript).

There has been some presumption that team utility is the sum or average of the group utilities, the 'utilitarian' payoff function. Bacharach (1999), in an illustrative example of his theory, took the group payoff to be the mean of the individual payoffs, as does Alessandra Smerilli (2008). Andrew Colman, Briony Pulford and Jo Rose (2007), in their test of team reasoning, followed Bacharach's later suggestion that team reasoners will aim for outcomes that Pareto dominate Nash equilibria, even when those outcomes are not equilibria themselves. Colman *et al* point out that, as consequence of this, the team reasoning outcomes in their experiment maximize the sum of the payoffs of the players.

However, it should be obvious from the discussion above that team reasoning does not require that the team utility is the sum or average of the utility of the individual members and that some specific versions may reject this constraint.

As we saw, Bacharach hypothesized that the team utility function would be Paretian, i.e. if every individual agent gets at least as much utility in outcome *x* than outcome *y*, and at least one agent does strictly better, then the group function will rank outcome *x* above outcome *y*. The utilitarian payoff function is Paretian but so are many others. The Pareto criterion alone cannot resolve the question of how the team function ranks outcomes in situations of partial, but not complete, harmony of interest - for instance it does not give any guidance about how to rank the off-diagonal payoffs in the prisoner's dilemma. In contrast the utilitarian function can provide a ranking. However, the utilitarian function is more informationally demanding than the Pareto criterion. Applying the Pareto criterion only requires that utility is ordinally measurable and does not require interpersonal comparability, whereas the utilitarian function requires inter-personal comparability of utility.

Although Bacharach posited the Pareto criterion as an empirical hypothesis, he did not specify any conceptual constraints on the group goal. In contrast, Sugden would impose conceptual constraints on the team utility function which lead to the rejection of the idea that it is the average of the utility of the team members. Sugden assumes that, when a player identifies with a group, she wants to promote the combined interests of its two members, at least in so far as those interests are affected by the game that is being played. Since his theory is concerned with cooperation for mutual advantage, team members will not take on a team utility function that would make them worse off than they would be as individual agent (*cf.* the discussion in section 2 of the team payoff function in the prisoner's dilemma). Since a team utility function that maximizes the sum of the individual payoffs takes no account of the distribution of payoffs between players, or of whether individual players improve their lot by team reasoning, the the team utility function in Sugden's theory cannot simply be the sum of individual payoffs.

There is an obvious analogy between the notion of a team payoff in the theory of team reasoning and the idea of group fitness in evolutionary theories of group selection.[14] The issue of the relation between individual and team payoff also has an interesting analogy with evolutionary biology. In evolutionary theory, most formal models define group fitness as total or average individual fitness. However, Rick Michod (2005) has argued that the fitness of the group cannot be equated with the total fitness of its parts. Michod allows that one situation can have more group fitness than another, even when the fitness of the individual units is exactly the same in each. Hence his model violates the Pareto criterion of social choice theory (Okasha, 2009) - the one criterion that it seems advocates of team reasoning agree on!

4. Framing and Team Reasoning

Bacharach and Sugden's theories also differ in the way that group agency comes about. Gold and Sugden (2007; 2008) labelled Bacharach's theory 'team agency as the result of framing', in contrast to Sugden's emphasis on assurance. But that should not be taken to imply that framing has no role to play in Sugden's theory. By comparing the how agents come to team reason in Bacharach and Sugden's theories, I will unpick the various processes involved in framing in decision-making in general.

Although Bacharach's theory is most strongly associated with framing, there is a framing step in Sugden's theory as well. Sugden (2000) makes an analogy between the way that, in the theory of team reasoning, people have different preferences from the perspectives of different units of agency and the way that, in standard decision theory, preferences are relative to particular conceptions of, or framings of, decision problems. He also explains group identification in terms of framing:

> 'The idea is that, in relation to a specific decision problem, an individual may conceive of herself as a member of a group or team, and conceive of the decision problem, not as a problem for her but as a problem for the team. In other words, the individual frames the problem, not as `What should I do?', but as `What should we do?'' (2000, p. 182-3)

Sugden describes the framing of a decision-problem as reflecting the agent's subjective perceptions - what she takes her decision-problem to be - and her preferences as her all-things-considered choice-relevant reasons. He says that her preferences are defined relative to her framing of the problem. So when Sugden says that an agent frames the problem as a problem for the team, that implies that the agent sees team considerations and achieving the group goal as

---

[14] This was suggested to me by Samir Okasha, who himself draws an analogy between individual and group fitness, and individual and social preference orderings as investigated in social choice theory (Okasha, 2009).

choice-relevant. Hence Sugden's account of team reasoning also involves framing, in the sense that it presupposes that the potential team reasoner conceptualizes the situation as a problem for the team, i.e. sees the team utility function as representing choice-relevant reasons. However, this is not enough to ensure that the agent will team reason. The agent also has to decide that she endorses team reasoning.

Sugden likens endorsing mutually assured team reasoning to making a unilateral commitment to a certain form of practical reasoning, where this reasoning does not generate any implications for action unless one has assurance that others have made the same commitment. Such assurance could be created by public acts of commitment or induced by repeated experience of regularities of behaviour in a population. But the questions of assurance and endorsement are separate: even if each individual were assured that others would choose their components of the group payoff-maximizing profile, each would still have to decide whether team reasoning was a mode of reasoning that she wanted to endorse. It is possible for a person either to have assurance but not endorse team reasoning, or to endorse team reasoning but not to have assurance.

Assurance seems to be bound up with group identifying: 'to construe oneself as a member of a team, one must have some confidence that the other members of that team construe themselves as members too' (Sugden 2000, p.194). This suggests that an agent could see the relevance of team-directed reasons but, because she does not have assurance that other agents will team reason, she does not group identify, in the sense of conceiving the group as a unit of agency, acting as a single entity in pursuit of some single objective (Gold and Sugden 2007, 2008).

This contrasts with Bacharach's approach in *Beyond Individual Choice*, where he is interested in 'the role of spontaneous group identification in decision making' (Bacharach 2006, p. 81), and there is no gap between framing the decision-problem as a problem for 'us' and group identifying, or between group identifying and team reasoning, in which the agent can choose whether to group identify or to team reason.[15] Bacharach does recognise that these are simplifications. When mooting the interpretation of ω as the probability that someone group identifies he footnotes an unexplored subtlety, that 'This assumes that group identification, if it happens, primes [team reasoning] with probability 1. More generally, if this probability is $p < 1$, and that of group identification is v; then $\omega = pv$.' (2006, footnotes p.152). Never-the-less, reasoning about whether or not to team reason does not enter the picture.

Nor does an agent who notices we-reasons get to decide whether or not she identifies with the group. When talking about group identification, Bacharach makes a distinction between the 'salience' of features that tend to promote group identity and their 'effectiveness', i.e. their tendencies, if and when perceived, to stimulate or inhibit group identity. However Bacharach speculates that there may be 'a positive relationship between salience and effectiveness'; that if

[15] Bacharach set out an earlier version of the theory in a 1997 working paper. Discussion of differences between the two will be confined to the footnotes.

features tending to promote group-identity are highly salient then, when noticed, they are also highly effective (Bacharach, 2006 p.87). So there is conceptual space in Bacharach's theory for a gap between noticing the group and group identifying. Elsewhere Bacharach says that this gap is filled by *affiliation*, 'a psychological process in which a person who does think about a certain group, defined by some shared property, comes to think about it as 'us'' (Bacharach, 1997, p. 2). Thus the gap is filled by a psychological process, not a choice.

For both Bacharach and Sugden there are potential gaps between noticing the group and group identifying, and between group identifying and team reasoning. For Sugden these gaps are both bridged by decisions, for Bacharach they are filled by psychological processes. As explained below, for neither of them are these points subject to standards of practical, instrumental rationality.

Now we are in a position to delineate the various steps involved in framing and decision-making. In the context of team reasoning: first the agent must see the possibility of cooperation and notice the potential for team reasoning, then she must group identify and see the group goal as providing a choice-relevant reason; third she must decide to act on team reasoning. The same steps occur in framing in general. We do not notice all the ways that we, counter-factually, could distinguish between objects or between actions. For a distinction to be the basis of action, an agent must first notice it. The move from 'noticing' *simpliciter* to 'noticing as choice-relevant' corresponds to Bacharach's 'effectiveness'. It also takes us to Sugden's start-point, where an agent conceptualizes her decision-problem using considerations that she takes to be choice-relevant. The move from 'noticing as choice-relevant' to deciding to act on team reasons corresponds to Sugden's 'endorsement'. We might call the two steps together, that take the agent from noticing a feature/ concept cluster/ family to acting on the reasons that it provides, 'motivational grip'. Then the two steps provide two reasons why a concept cluster, even when noticed, would not have motivational grip and hence would not be the basis for action: because it is not perceived as choice-relevant or because it is seen as choice-relevant but, nonetheless, does not form the basis of the agent's action. One reason an agent might not act on a feature that she acknowledges is choice-relevant is that she decides that it is not a reason that she wants to endorse. Other reasons why she might fail to move from seeing as choice-relevant to acting include making mistakes and akrasia (weakness of will).

These steps - noticing, noticing as choice relevant, and acting on - are involved in framing more generally, not just in the framing that leads to team reasoning.[16] To illustrate this with an example, inspired by Sugden's discussion of how economists model demand for consumption goods (Robert Sugden, 2000), consider someone in a supermarket deciding whether to buy an apple or an orange. It is natural to think of her as conceptualizing her decision-problem as

[16] I rely here on an intuitive grasp of the notion of 'noticing' but I note that it is complicated and there are conceptual issues regarding noticing that I do not have space to go into here (e.g. we tend to notice things that are choice-relevant).

choosing whether to ‘buy an apple’ or ‘buy an orange’. In fact, she is probably confronted by a crate of apples and a crate of oranges, each containing numerous pieces of fruit. But, although we can distinguish many pieces of fruit according to their position, a normal person wouldn’t see her choice as being between ‘buy the apple on the top left’, ‘buy the apple, top left but one’, ... ‘buy the orange on the bottom right but one’, ‘buy the orange on the bottom right’. The position of the pieces of fruit is not relevant to the decision problem and is not a part of its conceptualization. But position is a feature that someone might notice and discard as irrelevant to the choice.

As Sugden points out, questions about how to frame decision-problems do not always have easy answers. Another way that we can distinguish fruit in supermarkets is by the identity of its supplier. Like position, the supplier does not usually enter models of purchasing decisions. However, in the 1980s whether the orange came from South Africa was, for many people, choice-relevant information. Now-a-days some people choose, e.g., not to buy Israeli products, to buy products with fewer food miles, or to buy products that are fair trade. But other consumers are not bothered by the place of origin of their food. Supermarkets often label where food is from but many people do not assimilate this information, they do not ‘notice’ it. In order for place of origin to influence purchasing decision, a shopper must first notice the origin, then take place of origin to be a choice-relevant feature and, finally, act on that feature. Further, people who are already committed to not buying from certain suppliers are more likely to notice and even seek out information about the place of origin of the goods in their shopping basket.

Should consumers discriminate oranges according to their suppliers? For Sugden (2000, p. 200), ‘there is no objective answer to this question, independent of consumers' subjective perceptions: what matters is whether consumers take them to be the same.’ This reflects Sugden’s general approach, which does not acknowledge agent-neutral concepts of ‘validity’ and ‘rationality’. I won’t comment upon his approach here. I will merely note that, in saying that there are no objective answers to questions about framing, Sugden places himself at odds with most of the literature on framing in individual decision-making which, at least implicitly - and sometimes explicitly - assumes that there is a correct way to frame a decision-problem. I also note that, despite many researchers’ implicit commitments, there is at present no theory of rational framing that tells us whether a particular framing of a decision-problem is ‘correct’ or not.[17]

In *Beyond Individual Choice* Bacharach does not admit questions about the rationality of frames either. He is investigating valid instrumental practical reasoning. The agent can only reason from premises about the world that are accessible to her, so frames set the parameters of reasoning. Frames are supplied by involuntary psychological processes, to which the concept of

---

[17] James Joyce (1999) makes some intriguing comments about how such a theory might proceed. John Broome (1991) has argued that we should individuate outcomes by justifiers, where a *justifier* is a difference between two putative outcomes that makes it rational to have a preference between them. This implies that framing is a amenable to rational assessment, but Broome does not have any further discussion of what differences it is rational to have a preference between.

practical rationality does not apply. In other places, Bacharach suggested that a complete theory of rational choice would address questions about whether particular framings are rational or not (Bacharach, 2000) and discussed various problem cases (Bacharach, 1998). But any principles of rationality that are concerned with how agents should frame the world will not be principles of instrumental rationality. John Broome's rational requirements of indifference (Broome, 1991), Susan Hurley's agent neutral goals (Hurley, 1989) and Elizabeth Anderson's principles of rational self-identification (Anderson, 2001) all rely on thicker notions of rationality than means-end reasoning.

5. **Team Reasoning vs Other Theories of Cooperation**

Team reasoning has been advanced as a theory that can explain why cooperation is rational, as a theory that can predict cooperation, and as a psychological theory about how people make decisions. In this section I will compare team reasoning with other theories of cooperation on these dimensions. Again, the starting point for this is a comparison of the theories' analyses of cooperation in the prisoner's dilemma.

The prediction of rational cooperation

In the analysis above, a rational player of the one-shot Prisoner's Dilemma can choose *cooperate*. For many game theorists, this conclusion is close to heresy. For example, Ken Binmore (1994, pp. 102-117, quotation from p. 114) argues that it can be reached only by 'a wrong analysis of the wrong game': if two players truly face the game shown in Figure 1, then it follows from the meaning of 'payoff' and from an unexceptionable concept of rationality that a rational player must choose *defect*. Binmore recognises that rational individuals may sometimes choose *cooperate* in games in which *material payoffs* – that is, outcomes described in terms of units of commodities which people normally prefer to have more of rather than less, such as money, or years of not being in prison – are as in Figure 1. But that just shows that the payoffs that are relevant for game theory – the payoffs that govern behaviour – differ from the material ones. The first stage in a game-theoretic analysis of a real-life situation should be to find a formal game that correctly represents that situation.

Thus in response to the problem of explaining why *cooperate* is sometimes chosen in games whose material payoffs have the Prisoner's Dilemma structure, the methodological strategy advocated by Binmore is that of *payoff transformation*: we should look for some way of transforming material payoffs into game-theoretic ones that makes observed behaviour consistent with conventional game-theoretic analysis. This strategy has been followed by various theorists who have proposed transformations of material payoffs to take account of psychological or moral motivations that go beyond simple self-interest. For example, Ernst Fehr and Klaus Schmidt (1999) and Gary Bolton and Axel Ockenfels (2000) have proposed that, for any given level of material

payoff for any individual, that individual dislikes being either better off or worse off than other people. Matthew Rabin (1993) proposes that each individual likes to benefit people who act with the intention of benefiting him, and likes to harm people who act with the intention of harming him. Cristina Bicchieri (2006) proposes a theory of norms such that, when an individual recognises that she is in a situation where a norm exists, she prefers to conform to the norm if she believes that a sufficiently large number of other people will also conform and that a sufficiently large number expect her to conform.

The theory of team reasoning can accept Binmore's instrumental conception of rationality, but rejects his implicit assumption that agency is necessarily vested in individuals. We can interpret the payoffs of a game as showing what each player wants to achieve *if she takes herself to be an individual agent*. In this sense, the interpretation of the payoffs is similar to that used by Binmore: payoffs are defined, not in material terms, but in terms of what individuals are seeking to achieve. The theory of team reasoning can replicate Binmore's analysis when it is applied to players who take themselves to be individual agents: if Player 1 frames the game as a problem 'for me', the only rational choice is *defect*. However, the theory also allows the possibility that Player 1 frames the game as a problem 'for us'. In this case, the payoffs that are relevant in determining what it is rational for Player 1 to do are measures of what she wants to achieve *as a member of the group {Player 1, Player 2}*; and these need not be the same as the payoffs in the standard description of the game.

Thus, there is a sense in which team reasoning as an explanation of the choice of *cooperate* in the Prisoner's Dilemma depends on a transformation of payoffs from those shown in Figure 1. However, the kind of transformation used by theories of team reasoning is quite different from that used by theorists such as Fehr and Schmidt. In team reasoning, the transformation is not from material payoffs to choice-governing payoffs; it is from payoffs which govern choices for one unit of agency to payoffs which govern choices for another. Thus, payoff transformation takes place as part of a more fundamental *agency transformation*. Once we accept the possibility of agency transformation, in a situation where there is common knowledge of group identification all that is needed to make a unique prediction of rational cooperation is the assumption that (*cooperate*, *cooperate*) is the best profile of actions for the two players together. That assumption is hardly controversial in the games under consideration and in many everyday situations.[18]

---

[18] There is also a sense in which team reasoning is more efficient than payoff transformation: it is more efficient at maximizing the team payoff function than a game between two *benefactors*, (individuals who have each taken on the team payoff function but still use individual reasoning. Bacharach (1999) proves that optimal team decision rules give Nash equilibra for benefactors but not all Nash equilibria for benefactors define optimal team decision rules. Intuitively, team reasoners are concerned with co-ordination, and co-ordination is important for efficiency - there is usually a lesser total payoff when agents fail to co-ordinate in games with scope. The team reasoning equilibrium singles out the best of the possible ways of co-ordinating, whereas benefaction generally does not lead to a unique equilibrium, leaving the possibility of mis-co-ordination.

Then, the idea is that the rationality of each individual's action derives from the rationality of the joint action of the team. Bacharach says,

> 'The remainder of the team reasoning procedure is then inevitable. Once I have computed the best team profile and identified my component in it, team reasoning prescribes that I should choose to perform this component... I am rationally obliged to follow the remainder of the procedure. If I believe that *we* should do a certain combination of actions, it is logically required that I also believe that I should do the bit that falls to me. If I am convinced that we should pass each other on the left, I must also think that I should pass you on the left (and that you should do likewise). The underlying general principle is that I cannot coherently will something without willing what I know to be logically entailed by it. This is a standard inference rule of deontic logic, the logic of what ought to be.' (2006, chapter 4, section 4)[19]

Not everyone agrees with Bacharach about the rules of deontic logic. But even those who disagree that it follows from the rationality of a profile *for the group* that it is rational for the *individual members* to play their part, must acknowledge that team reasoning is no worse off than standard decision theory in this respect. Actions are taken by temporal parts of people and, as is notorious in the theory of dynamic choice, the incentives of a temporal part may differ from the interests of her later selves or from the interests of the agent conceived as a person over time. Any difficulty that the theory of team reasoning has when explaining why it is rational for individuals to play their part in the team plan is analogous to the difficulty that standard decision theory has to overcome, if it is to explain why it is rational for temporal parts to act in the interests of the person over time.[20]

In contrast, even if payoffs have been transformed to go beyond self interest, conventional game theory still does not necessarily provide an explanation of why each individual chooses *cooperate* in a prisoner's dilemma. The payoff transformation theories listed above all tend to function by decreasing the relative appeal of the 'off-diagonal outcomes' where one player cooperates and the other doesn't. Depending on the theory, the individual does not want: to get a different payoff from the other player; to cooperate and hence benefit another player who is defecting and hence harming himself, and vice versa; or to follow a norm when no-one else does. This can make *(cooperate, cooperate)* into a Nash Equilibrium. But it does not tend to alter the

---

[19] Anderson also argues that, when a group of agents see their actions as jointly advancing a common goal, then it is rational for an agent 'to do my part in what we are willing together' (Anderson 2000, pp. 28–30).

[20] For more on the analogy between individuals and teams, and time slices and persons-over-time, see Gold (forthcoming).

equilibrium status of *(defect, defect)*.[21] There are now two Nash Equilibria, *(cooperate, cooperate)* and *(defect, defect).* So payoff transformation theories face the problem of equilibrium selection, explained above. They cannot show that that rational players of this game choose *cooperate* or make a unique prediction of cooperation. The situation of payoff transformation theories is similar regarding cooperation in the stag hunt, except that *(stag, stag)* was already an equilibrium anyway before any payoff transformation occurred.[22]

The problem faced by theories of individual reasoning in explaining cooperation is that recommendations for action are conditional on the actor's beliefs about what the other individuals will do. Within classical theory, there are no grounds for assigning such beliefs. However payoff transformation theories can also predict cooperation *if* the prior probability that each player assigns to the other cooperating is suitably high or, for Bicchieri, if the probability that the other player is sensitive to the norm is suitably high. So, in order to make the prediction, they must be supplemented with an account of the probabilities that the agent assigns. The agents of classical game theory have no grounds for assigning probabilities one way or the other so, if the proponents of payoff transformation theories wish to claim that they predict rational cooperation, then they also owe us an account of how rational agents acquire their probability estimates. It is usually claimed that rational probabilities are formed by Bayesian updating, and Bicchieri explicitly appeals to the idea that the probability that the other player will be norm-sensitive is high in a Bayesian game.[23] In contrast, the theory of team reasoning generates recommendations for action that are not conditional on the actor's beliefs about what the other individuals will do. Where there is common knowledge of group identification, the theory of team reasoning can predict rational cooperation. Recommendations are conditional on beliefs about group identification, but these are outside the purview of instrumental rationality. Whilst payoff transformation theories of cooperation can predict cooperation given suitable beliefs, there is a sense on which only team reasoning predicts rational cooperation, within the restrictions of classical game theory.

The explanation of cooperation

---

[21] For a more detailed exposition of team reasoning versus payoff transformation theories of cooperation in the prisoner's dilemma, including a numerical example, see section 3 of Gold and Sugden (2008).

[22] It is possible that, in the stag hunt, a player who was suckered when playing stag would not consider that the rabbit player *intended* to harm her. However,even if Rabin's theory does not downgrade the off-diagonal outcomes in this case, it is not going to show that *(stag, stag)* is uniquely rational because the theory does not change the equilibrium status of (*rabbit, rabbit).*

[23] Whilst considering Bayes and rationality, we might also note that framing is problematic for Bayesianism. For the operation of Bayes rule, we must assume that the agent assigns a strictly positive probability to each alternative. There is no space for a agent to be simply unaware of an alternative.

Team reasoning is proposed by both Bacharach and Sugden as a mode of reasoning that people actually use. Advocates of payoff transformation theories also claim that their favoured payoff transformation describes what people do. All these theories can, in principle, explain cooperation. But which provides the best explanation of what people do? One way to adjudicate the question is to compare the theories with respect to evidence about their auxiliary hypotheses.[24]

One type of auxiliary hypothesis relates to evolution. Any proposed proximate mechanism for cooperation must be a possible result of evolution, so having a plausible evolutionary story speaks in favour of any particular proximate mechanism. (Conversely, if we have independent reasons for favouring a particular proximate mechanism and if that mechanism has implications for the correct ultimate mechanism, then we might be able to use what we know about the proximate mechanism to adjudicate between ultimate, evolutionary explanations.) There is not space here to go into detail about the relation between proximate and ultimate explanations, but I will compare Bacharach and Sugden's views about the evolution of the capacity for team reasoning.

Bacharach (2006, ch.3) argued that group selection can explain cooperative behaviour and that group identification is a key proximate mechanism in producing well functioning groups. Sugden, on the other hand, thinks that group selection is unlikely to be part of any ultimate explanation because the conditions that are required for group selection to occur were never fulfilled (personal communication).

Sugden (2002; 2005) has suggested that team reasoning is related to generalised reciprocity, which is at least partly sustained by agents taking pleasure in the correspondence of their sentiments with the sentiments of others in the group, or *fellow-feeling*. A natural corollary would be to hypothesise that reciprocity is the appropriate ultimate explanation. Alternatively, Sugden's account of fellow-feeling might be compatible with kin selection. We might have evolved fellow-feeling because it was helpful for increasing the number of descendants left by our kin - who in Pleistocene times would also have been our fellow group (or *deme*) members. But, in modern times, the trait 'mis-fires' and we apply it to group members who are not related to us.

Alternatively, given the right environmental conditions, team reasoning might have evolved by individual selection. Although he mainly focussed on group selection, Bacharach himself raises this intriguing possibility. He claims that our ancestors faced *ludic diversity*, or an environment in which they had to play many types of games. Team reasoning gets good results in a variety of games and, since we have limited cognitive resources, team reasoners might have had an evolutionary advantage because team reasoning is a parsimonious way of achieving good

---

[24] Another way is to run experiments to test the theories. However, since all the theories claim to explain cooperation, when trying to discriminate between them, looking at behaviour in standard games is not sufficient. Some effort has been put into identifying which reasoning or payoff transformations people actually use and some ingenious experiments have been devised. Colman *et al* (2007) test team reasoning by constructing games in which there are unique Nash Equilibria that would not maximize a team payoff function. Guala *et al* (2009) compare team reasoning with other theories of cooperation by considering the role of expectations in various theories and manipulating subjects' beliefs.

outcomes across a diverse range of situations. In this story, team reasoning emerges by individual selection.

Hence it should be obvious that it is neither the case that group selection being wrong would be problematic for Bacharach's theory, nor that group selection being right would be problematic for Sugden. When Bacharach proposed that team reasoning evolved by group selection, he argued that a 'how possible' evolutionary explanation provided evidential support for the existence of team reasoning. The fact that there are multiple 'how possible' explanations available that support the evolution of team reasoning is not damaging to the theory. Although Bacharach and Sugden may prefer different 'how possible' explanations, it is not clear that either of their theories requires their preferred evolutionary explanation. In particular, whether or not team reasoning is a mechanism that people use is independent of whether or not there was group selection.

A second type of auxiliary hypothesis is about the psychological capacities that each theory requires. For example, sympathy is required for altruism, intention detecting for Rabin's kindness, a sense of obligation or desire for conformity for norms, group identification or fellow feeling for team reasoning. Evidence for all of these capacities can be found. This supports the intuitively appealing view that humans are creatures with heterogenous motivations, so in practice a catholic approach is to be favoured. Gold and Sugden (2007) explicitly say that a given pattern of behaviour could be generated either collectively, using team reasoning, or individually, using individual reasoning supplemented with appropriate beliefs.[25]

However, if we want to encourage cooperation then we need more precise answers, as different theories indicate different strategies for increasing cooperation: should we institute policies that affect expectations, increase group identification, or enable transparent perception of people's intentions? Even accepting that all the above motivations are used in some circumstances, a complete theory would specify which motivation is used in which circumstances.

Another type of auxiliary hypothesis that any of these theories could be supplemented with is a hypothesis about framing. In any particular situation that offers the potential for cooperation, it might be possible to construe the situation in several ways, so that more than one of altruism, inequality aversion, kindness, norms, team reasoning, etc. could be seen as choice relevant. As discussed above, a pre-requisite for modeling a situation is deciding which considerations are choice-relevant for the agent. Those theories that do not explicitly include framing assume it implicitly. If we accept that humans have heterogeneous motivations, then we need to know under what circumstances people frame the situation so that each potential motivation is seen as be

---

[25] Another type of auxiliary hypothesis involves support for the cognitive primitives implied by a theory. With respect to team reasoning, Gold and Harbour (forthcoming) argue that the primitives needed for team reasoning enjoy direct and diverse support from linguistic theory. They use this approach to discriminate between theories of collective intentions but, at present, this approach has not been used to discriminate between theories of cooperation.

choice-relevant. A general theory of framing might give us a theoretical grip on when each theory can predict choice, and provide testable hypotheses.

6. Conclusion

I have explained why games with scope pose both a normative and an explanatory problem for classical game theory, and I have shown how team reasoning can explain cooperation in prisoner's dilemmas and stag hunts. By comparing Bacharach and Sugden's theories of team reasoning, I have shown some of the ways in which theories of team reasoning can differ.

I clarified that neither Bacharach nor Sugden is committed to the idea that the team payoff function is the average of the individual utilities and that Sugden is not committed to expected utility maximization. In doing so, I drew a structural analogy between models of team reasoning and models of group selection, although I demonstrated that the truth of the theory of team reasoning does not rely on the truth of the theory of group selection.

I also showed how framing has a role to play in both Bacharach and Sugden's theories and, indeed, in all theories of cooperative motivations. A complete theory of cooperative behaviour would require an account of when people 'notice' particular features of situations, when people take features as 'choice relevant' and when noticing a feature as choice relevant leads to choosing in accordance with it.

Acknowledgments: Thanks to Samir Okasha for comments and discussion, and to audiences at the University of British Columbia and the CUNY Seminar in Logic and Games, where preliminary versions of some of this material were presented. This work was supported by a British Academy Small Research Grant, which I gratefully acknowledge.

Bibliography

Anderson, E. (2001). Unstrapping the Straitjacket of "Preference": A Comment on Amartya Sen's Contributions to Philosophy and Economics *Economics and Philosophy*(17), 21-38.

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

Bacharach, M. (1998). Preferenze razionali e descrizioni In M. Galarotti & G. Gambetta (Eds.), *Epistemologia et Economia*. Bologna: CLUEB.

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

Bacharach, M. (2000). Scientific Synopsis. Unpublished Unpublished manuscript (describing initial plans for Beyond Individual Choice).

Bacharach, M. (2006). *Beyond Individual Choice: Teams and frames in game theory*. Princeton: Princeton University Press.

Bardsley, N. (2007). On Collective Intentions: Collective Action in Economics and Philosophy. *Synthese, 157*(2), 18.

Bicchieri, C. (2006). *The Grammar of Society: the Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Bolton, G., & Ockenfels, A. (2000). ERC – a theory of equity, reciprocity and competition. *American Economic Review, 90*, 166-193.

Broome, J. (1991). *Weighing Goods*. Oxford: Basil Blackwell.

Colman, A., Pulford, B., & Rose, J. (2007). Collective rationality in interactive decisions" Evidence for team reasoning. *Acta Psychologica 128*, 387-397.

Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics 114*, 817-868.

Gilbert, M., P. (1987). *On Social Facts*. New York: Routledge.

Gold, N. (forthcoming). Framing and Self-Control. In N. Levy (Ed.), *Self-Control and Addiction: Lessons from Philosophy and Cognitive Science* Oxford: Oxford University Press.

Gold, N. (unpublished manuscript). Group Goals, Game Theoretic Reasoning and Spontaneous Collective Intentions.

Gold, N., & Harbour, D. (forthcoming). Cognitive Primitives of Collective Intentions: Linguistic Evidence of our Mental Ontology. *Mind and Language*.

Gold, N., & Sugden, R. (2007). Collective Intentions and Team Agency. *Journal of Philosophy 104* (3), 109-137.

Gold, N., & Sugden, R. (2008). Theories of Team Agency. In F. Peter & S. Schmidt (Eds.), *Rationality and Commitment* Oxford Oxford University Press.

Guala, F., Mittone, L., & Ploner, M. (2009). Group membership, team preferences, and expectations.

Harsanyi, J., & Selten, R. (1988). *A General Theory of Equilibrium Selection in Games* Cambridge, Ma.: MIT Press.

Hodgson, D. (1967). *Consequences of Utilitarianism* New York: Oxford University Press.

Hollis, M. (1998). *Trust Within Reason*. Cambridge: Cambridge University Press.

Hurley, S. (1989). *Natural Reasons*. New York: Oxford University Press.

Joshi, M. S., Joshi, V., & Lamb, R. (2005). The Prisoners' Dilemma and city-centre traffic. *Oxford Economic Papers, 57*(1).

Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

Kollock, P. (1998a). Social dilemmas: The anatomy of cooperation. *Annual review of sociology, 24*, 183-214.

Kollock, P. (1998b). Transforming Social Dilemmas: Group Identity and Cooperation In P. A. Danielson (Ed.), *Modeling rationality, morality and evolution* (pp. 186-210). Oxford: Oxford University Press.

Michod, R. (2005). The group covariance effect and fitness trade-offs during evolutionary transitions in individuality. *Proceedings of the National Academy of Science 103*(24), 9113–9117.

Nash, J. (1953). Two-Person Cooperative Games. *Econometrica, 21*, 128-140.

Okasha, S. (2009). Individuals, groups, fitness and utility: multi-level selection meets social choice theory. *Biology and Philosophy, 24*, 561-584.

Rabin, M. (1993). Incorporating Fairness Into Game Theory and Economics. *The American Economic Review, 83*, 1281-1302.

Regan, D. (1980). *Utilitarianism and Cooperation* New York Oxford University Press.

Skyrms, B. (2004). *The Stag Hunt and Evolution of Social Structure*. Cambridge: Cambridge University Press.

Smerilli, A. (2008). We-thinking and 'double-crossing': frames, reasoning and equilibria. Munich Personal RePEc Archive Paper 11545. University Library of Munich, Germany.

Sober, E., & Wilson, D. S. (1998). *Unto Others: The evolution and psychology of unselfish behavior*. Cambridge, Ma: Harvard University Press.

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

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

Sugden, R. (2002). Beyond Sympathy and Empathy: Adam Smith's Concept of Fellow-Feeling. *Economics and Philosophy, 18*, 63-87.

Sugden, R. (2003). The Logic of Team Reasoning. *Philosophical Explorations, 6*, 165-181.

Sugden, R. (2005). Fellow-feeling. In B. Gui & R. Sugden (Eds.), *Economics and Social Interaction* (pp. 52-75). Cambridge Cambridge University Press.

Taylor , M. (1987). *The Possibility of Cooperation*. Cambridge Cambridge University Press.

Tuomela, R. (2009). Collective intentions and Game Theory. *Journal of Philosophy 106*, 292–300.