University of Groningen

Team reasoning and group identification

Hindriks, Frank

Published in:
Rationality and Society

DOI:
[10.1177/1043463111429274](https://doi.org/10.1177/1043463111429274)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2012

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Hindriks, F. (2012). Team reasoning and group identification. *Rationality and Society*, 24(2), 198-220.
<https://doi.org/10.1177/1043463111429274>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Team reasoning and group identification

Rationality and Society

24(2) 198–220

© The Author(s) 2012

Reprints and permission:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1043463111429274

rss.sagepub.com **Frank Hindriks**

University of Groningen, Groningen, The Netherlands

Abstract

The team reasoning approach explains cooperation in terms of group identification, which in turn is explicated in terms of agency transformation and payoff transformation. Empirical research in social psychology is consistent with the significance of agency and payoff transformation. However, it also reveals that group identification depends on social categorization processes to a greater extent than is currently acknowledged within the team reasoning approach. In light of this, Bacharach's claim that group identification is prompted by a perceived conflict between individual and collective interests has to be rejected. Instead, it is triggered by the salience of a social category. Sugden's account of the role of trust in team reasoning needs to be modified: rather than by evidence of behavior, it is induced by common knowledge of shared membership of a particular group. The upshot is that the empirical adequacy of the team reasoning approach can be substantially enhanced by incorporating the notion of category salience as a key explanatory variable.

Keywords

category salience, cooperation, group identification, social categorization, social identity, team reasoning

Many situations involve a tension between individual and collective interests. In such situations people can try to take advantage of those with whom they interact. However, in practice people often cooperate, putting their own interests at risk. In light of this, it is unsurprising that many social scientists have conceptualized cooperation as a non-rational process. In particular,

Corresponding author:

Frank Hindriks, Faculty of Philosophy, University of Groningen, Oude Boteringestraat 52, 9712GL Groningen, the Netherlands.

Email: f.a.hindriks@rug.nl

social psychologists have proposed explanations that focus on group identification as the key to cooperation, and rationality plays no role in these explanations. Recently, Michael Bacharach (1999, 2006) and Robert Sugden (2000, 2003) have also proposed explanations of cooperation that rely on group identification. At the same time, however, they maintain that people act rationally when they cooperate due to group identification. They propose that people who identify with a certain group tend to engage in a mode of reasoning called ‘team reasoning’, meaning that the agents conceive of themselves as a team and pursue the collective interests of the group rather than their own individual interests.

The team reasoning approach carries with it the promise of combining the theoretical framework of rational choice theory with empirical findings from social psychology. I argue, however, that the proposals that Bacharach and Sugden have put forward are unsatisfactory as they stand: they do not do sufficient justice to the power of social categorization, which is a non-rational process. Both versions of the approach can be improved in this respect. Bacharach claims that a perceived tension between individual and collective interests triggers team reasoning and thereby generates cooperation. Empirical research in social psychology suggests instead that cooperation is often relatively arbitrary and is prompted by the salience of the fact that the individuals belong to the same social category. The impact that the salience of social categories has on cooperation is also relevant to Sugden’s account of team reasoning. Sugden claims that team reasoning requires observation-based assurance. Findings concerning depersonalized trust, however, suggest that people often rely on others even if they have not observed any relevant behavior. Again, basically all they need to know is that others belong to the same social category. The upshot is that the empirical adequacy of the team reasoning approach can be substantially improved by incorporating social categorization as a key explanatory factor.

Team reasoning

The team reasoning approach

Bacharach and Sugden present the team reasoning approach as an attempt to solve some problems with which conventional non-cooperative game theory is confronted. Consider the Prisoner’s Dilemma (Figure 1). The rational choice for an individual in a Prisoner’s Dilemma is to defect and not to cooperate. This holds in spite of the fact that both players would be better off if both of them were to cooperate. Given this scope for common gain, it appears that it is in some sense rational for them to cooperate. Moreover,

	Player 2	
Player 1	Cooperate	Defect
Cooperate	3, 3	1, 4
Defect	4, 1	2, 2

Figure 1. The Prisoner's Dilemma.

	Player 2	
Player 1	Hi	Lo
Hi	2, 2	0, 0
Lo	0, 0	1, 1

Figure 2. Hi-Lo.

experiments reveal that a substantial number of people do in fact cooperate in Prisoner's Dilemma situations. Bacharach and Sugden argue that conventional game theory cannot account for this.¹ A similar puzzle arises in relation to the game of Hi-Lo (Figure 2). This is a common-interest game in which the players gain by converging on the same strategy, regardless of the strategy they select. However, one strategy is superior to the other in that it provides both players with a higher payoff. Again there is scope for common gain. Even though it appears to be uniquely rational to select the strategy that delivers the higher payoff, conventional game theory does not accommodate this intuition and does not exclude the other strategy.

Bacharach and Sugden suggest that the source of these problems is that conventional game theory lacks the notion of doing something together, or that of joint action. They present the team reasoning approach as a generalization of game theory that is able to accommodate this notion. It permits one to ask the question 'What should *we* do?' in addition to 'What should I do?'.² Answering the former question requires a mode of reasoning that differs from the individual mode of reasoning employed to answer the latter. When they adopt this non-individual mode of reasoning, which Bacharach and Sugden call 'team reasoning', the individuals conceive of themselves as a group and think in terms of 'we' rather than 'I'. In other words, they transform the unit of agency from independent individuals to that of an interdependent group or team of individuals of which each of them makes up only a part. Furthermore, they have to conceptualize the payoffs in terms of the

interest of the team as a whole rather than in terms of the interests of independent individuals. This comes down to adopting a common or shared utility function that is based on collective rather than individual interests. The first step is one of *agency transformation*, the second is one of *payoff transformation*. Once these steps have been taken it turns out to be rational to cooperate, which means that the scope for common gain can be exploited.³

The way in which agency and payoff transformation enable cooperation can be elucidated in terms of team reasoning schemas. Team reasoning from a group viewpoint is explicated as follows (with S for a set of individuals, A for any profile of actions, and U for some payoff function):⁴

- (1) We are the members of S .
- (2) Each of us identifies with S .
- (3) Each of us wants the value of U to be maximized.
- (4) A uniquely maximizes U .

Each of us should choose her component of A (see Bacharach, 2006: 158; Gold and Sugden, 2007: 125).

Identifying with a set of individuals involves agency transformation, so premise 2 implies that the members of the set of individuals conceive of themselves as a group or team. In order to make sense of this schema, we can imagine the members of S to be in an open meeting at which the relevant premises are announced and subsequently acknowledged as true by each person; a similar procedure is applied to the inference, which is acknowledged as valid by all (Bacharach, 2006: 158; Gold and Sugden, 2007: 125). This reveals that team reasoning requires common knowledge of identification and utility maximization, as is made explicit in the schema for team reasoning from an individual viewpoint:

- (1) I am a member of S .
- (2) It is common knowledge in S that each member of S identifies with S .
- (3) It is common knowledge in S that each member of S wants the value of U to be maximized.
- (4) It is common knowledge in S that A uniquely maximizes U .

I should choose my component of A (see Bacharach, 2006: 159; Gold and Sugden, 2007: 126).

The second premise of these schemas indicates that agency transformation has occurred. The third premise implies that payoff transformation has taken place. This transforms the decision problem from one of selecting the

best reply to the actions of other team members to a maximization problem. In other words, the choice situation ceases to be one of strategic interaction and becomes one in which several individual agents maximize the same payoff function and act as one.

Let me illustrate team reasoning using the Footballer's Problem, which is a version of the Hi-Lo game. Player *A* has to pass the ball to player *B* because an opposing player is converging on him. Player *A* faces a choice between moving the ball left or right, and player *B* has to run left or right to intercept the pass. The chance of scoring a goal is slightly higher when both choose right as compared to left. The players must choose simultaneously and are not in a position to communicate. It is easy to see that *A* should choose right if *B* does so, and that *A* should choose left if *B* does so (similarly for player *B*). At the same time, it is not a trivial matter to go beyond this and arrive at an unconditional recommendation (other than randomization). Intuitively it is obvious that the players should choose right. It is less obvious, however, how this response can be rationalized in a convincing manner.

Team game theory, Bacharach and Sugden argue, delivers the recommendation to choose right in a way that fits our intuitions. If the players engage in team reasoning, they acknowledge that they are members of the same team, and that each of them identifies with that team, which has scoring goals (or maximizing the chance that a goal will be scored) as its objective. The profile of moving the ball right and running to the right uniquely maximizes this chance. So each player should choose his component of this profile.

The Footballer's Problem reveals that payoff transformation by itself does not always lead to cooperation. Suppose the payoffs of those involved in a Hi-Lo game are added to one another (cf. note 3). This does not change the structure of the game, so an agent who relies on best-response reasoning cannot settle on right as the uniquely rational choice. An agent who also engages in agency transformation, however, will assume that other players maximize the team utility function just as she does. This enables her to abandon best-response reasoning herself, and regard the situation as a problem of maximizing team utility. As a consequence, a categorical or unconditional recommendation follows (Bacharach, 2006: 173).

Interdependence and assurance

According to the team reasoning approach, group identification involves both agency transformation and payoff transformation. Gold and Sugden in fact identify group identification with these two transformations, as is suggested by the following passage: 'We will say that *i* identifies with *G* if *i*

conceives of G as a unit of agency, acting as a single entity in pursuit of some single objective' (Gold and Sugden, 2007: 125).⁵ In this section I discuss what Bacharach and Sugden add to this core idea in their respective conceptions of how group identification leads to cooperation. This prepares the way for evaluating their proposals in the next two sections.

Bacharach proposes that group identification is a mechanism that is triggered by the presence of common interests or scope for common gain. He develops this idea in terms of a feature he calls 'strong interdependence'. Individuals are interdependent in this strong sense just when 'they will do well only if the other does something that does not seem to be assured' (Bacharach, 2006: 84). More precisely, a set of individuals is *strongly interdependent* exactly when a strategy-profile exists that serves their individual interests better than the outcome that results from individualistic best-response reasoning. This is the case in Prisoner's Dilemma's and Hi-Lo games (Bacharach, 2006: 84 and 143 respectively).

Bacharach maintains that the perception of strong interdependence prompts group identification, or makes its occurrence probable. He refers to this as 'the *interdependence hypothesis*'. Group identification produces what he calls the 'reasoning effect', which consists of the relevant individuals engaging in team reasoning. As explained above, team reasoning involves agency transformation and payoff transformation. Bacharach regards both of these as matters of framing, that is, as non-rational processes: under certain conditions people just happen to switch from an 'I-frame' to a 'we-frame', and from 'I-reasoning' to 'we-reasoning' (Bacharach, 2006: 90 and 135–137). The upshot, then, is that the perception of strong interdependence triggers agents to re-conceptualize their situation in terms of membership of a team and to adopt the objective of the team as their own.

In contrast to Bacharach, Sugden has no ambition to provide an explanation of team reasoning or to offer an account of the conditions under which it takes place. He claims that, as a consequence, his theory has no empirical content (Sugden, 2000: 203). This appears to be something of an exaggeration. Even though he provides no sufficient conditions for group identification, Sugden does present assurance as a necessary condition. In his view an individual will engage in team reasoning only if she is confident or has assurance that others will do so too. Surely this is an empirical claim. I shall refer to it as 'the *assurance hypothesis*'.

Sugden explicates assurance in terms of having a reciprocal reason to believe that every member of the team endorses and acts on mutually assured team reasoning (2003). Sugden suggests that having such a reason requires observation or experience: 'Mutual confidence in the use of

a particular mode of reasoning can be brought about *merely* by mutual observation of behaviour, combined with inferences about how other people reason' (Sugden, 2000: 196, emphasis added; see also 2003: 178). Assurance need not be based on observation of the behavior of all the individuals with whom one interacts, but can instead be based on experience of regularities of behavior in a population: 'assurance might be induced by the players' common experience that, in some analogous class of recurrent interactions, people like them do in fact engage in team reasoning' (Sugden, 2003: 180).⁶ As an example, Sugden mentions his experience in British traffic, which he uses to predict how drivers he has never met will behave (2000: 196). In such cases the public nature of the practice provides the source of assurance (Sugden, 2003: 179). The key thing to appreciate is that the kind of trust involved in assurance is observation- or evidence-based.⁷

Bacharach and Sugden have very different ideas about the rationality of team reasoning and group identification.⁸ Bacharach takes the question of whether people engage in team reasoning to be entirely a matter of how they happen to frame the situation, which is a non-rational issue. For Sugden the question of whether people actually engage in team reasoning depends not only on framing, but also on assurance. As he conceives of team reasoning in terms of reciprocity, assurance functions as a rational precondition for team reasoning on Sugden's account. Considerations of rationality do not enter into the way Bacharach conceives of the process that triggers people to engage in team reasoning.⁹

This explication of what these proponents of the team reasoning approach mean by group identification facilitates the evaluation of their proposals. I shall ask to what extent it makes sense to conceptualize group identification in terms of agency and payoff transformation, and I shall explore whether what Bacharach and Sugden say about the conditions under which people engage in team reasoning is plausible. As Bacharach (2006) and Sugden (2000) acknowledge, the best place to look for answers to these questions is social psychology. The notion of group identification plays an important role in that discipline, and numerous empirical studies have been performed to investigate the processes involved and the conditions that trigger those processes. For this reason I shall review the major developments in social psychological research about group identification. I first introduce the social identity approach and the notion of social categorization. Furthermore, I describe how the salience of a social category has surfaced in social psychology as a key determinant of cooperation. In the following section I criticize Bacharach's interdependence hypothesis and argue that the team reasoning approach should incorporate the notion of category salience. Once I have introduced the notion of depersonalized trust, I then turn to

Sugden's assurance hypothesis. The empirical evidence, I argue there, favors the notion of depersonalized trust over Sugden's conception of evidence-based trust. The overarching conclusion that emerges from these discussions is that the empirical adequacy of the team reasoning approach can be enhanced by incorporating the salience of social categories as a key explanatory factor.

The interdependence hypothesis

Group identification and category salience

One of the most prominent and well-developed theoretical frameworks in social psychology is the social identity approach, a core tenet of which is the idea that social identity and group identification can explain cooperation. As I discuss below, this idea extends to situations in which collective and individual interests are in tension with one another, which psychologists commonly call 'social dilemmas' (economists tend to refer to them as 'collective action problems'). The notion of social categorization plays a central role in the social identity approach. The evidence reveals that social categorization is conducive to cooperation. And I shall argue below that neither Bacharach nor Sugden does sufficient justice to this finding. In this section I discuss how and under what conditions social categorization leads to cooperation, and more specifically what role what I shall refer to as 'category salience' plays in this process.

The social identity approach was initiated by Henry Tajfel (1970) who proposed the minimal group paradigm. The experiments that Tajfel and his colleagues conducted suggest that the introduction of arbitrary and virtually meaningless categorizations in a situation results in people favoring others who belong to the same category and increasing the extent of cooperation insofar as other members of the category are concerned. This phenomenon is known as 'in-group favoritism'. Tajfel's research was particularly surprising in that it revealed that interaction between group members or with out-group members is not required for in-group favoritism. Cognitive categorization processes suffice to generate this effect. Apparently people rely on categorizations to simplify, structure, and give meaning to their social environment, and this in turn affects their motivation and behavior. Minimal categorizations that have been found to lead to in-group favoritism include preferring Klee to Kandinsky and being labeled as someone who systematically overestimates the number of dots on slides consisting of large clusters of dots.

In order to explain how introducing arbitrary categorizations leads to in-group favoritism, Tajfel and John Turner proposed the social identity theory, a more recent component of the social identity approach (see Tajfel, 1978, 1981; Tajfel and Turner, 1979; Turner, 1975). The point of departure for social identity theory is the idea that there is a universal need for people to maintain a positive self-identity. A further assumption made is that people derive their self-worth in part from the social categories with which they are associated. They are taken to be motivated to establish a positive evaluation for those categories as compared to rival categories. In-group bias provides the means for making those categories positively distinct, which enhances the sense of self-worth that members have. In addition to positive relative evaluation, in-group bias can manifest itself in terms of differential behavior or in-group favoritism. According to social identity theory, then, in-group favoritism has a cognitive and an affective or motivational component. People perceive social categorizations and tend to minimize perceived differences within categories and accentuate differences between them. Furthermore, they assign affective significance to features of the in-group. At the heart of social identity theory lies the idea that self-enhancement motivates people to regard other group members favorably and treat them accordingly.

Marilynn Brewer (1991) proposes an alternative explanation based on the idea that group identification is caused by security motives, in particular the need to belong to a group and the need to distinguish oneself from others. The former motivates immersion in a social group, while the latter discourages it: 'social identity derives from a fundamental tension between human needs for validation and similarity to others (on the one hand) and a counter-vailing need for uniqueness and individuation (on the other)' (Brewer, 1991: 477). The evidence concerning the minimal group paradigm that she had surveyed earlier suggests that in-group favoritism is triggered in particular by perceived intra-group similarity and that perceived dissimilarity to the out-group makes little or no difference (Brewer, 1979). In contrast to what proponents of the self-enhancement explanation maintain, in-group favoritism does not require out-group derogation. What is more, in-group favoritism does not even require a positive view of the group to which one belongs.

Brewer uses the notions of personal identity and social identity that were proposed within the context of social categorization theory (Turner, 1985) to develop her alternative. Someone's personal identity consists of the features that uniquely identify her, while her social identity or collective self concerns the features she has in common with others. Brewer maintains that which of the two security motives prevails depends in part on whether personal identity is salient or whether the collective self has been activated.

She transforms the idea that members come to view their social category as positively distinct using the notion of optimal distinctiveness, and maintains that the extent of identification is determined by a number of conflicting needs.

Brewer (1979) critically examines research within the minimal group paradigm, and scrutinizes in particular the claim that any arbitrary categorization leads to in-group favoritism. She argues that categorization has a significant effect only if membership of social categories is sufficiently salient. Such salience, in turn, is influenced by a number of factors, a significant one being that of having a common fate. When a selection of individuals within a larger group wins a prize in a lottery, for instance, this leads them to evaluate the traits of fellow prizewinners more favorably in comparison to individuals who did not win a prize. What will prove to be important below is that, as the lottery example reveals, a situation in which people have a common fate need not be a social dilemma. Instead, it can be anything that gives people the idea that they are 'in this together', and it is not at all obvious that having opposing individual interests has this effect. A more recent discovery concerning the role that the salience of a social category plays is that in-group favoritism does not occur if group membership is not common knowledge (Karp et al., 1993; Yamagishi and Kiyonari, 2000). Furthermore, it is not even expected to occur when people know that the people with whom they interact do not know which social categories they belong to (Foddy et al., 2009). These findings suggest an important qualification of the central claim of the minimal group paradigm. I shall use the term 'category salience' to formulate the qualification, and use it to refer to the salience of a social category that is common knowledge among the individuals concerned. The qualification is that, rather than mere categorization, it is category salience that leads to in-group favoritism.

The notion of a collective self can be used to shed light on the process by which category salience induces in-group favoritism. Brewer and Gardner (1996: 87) maintain that the collective self that is involved in group identification 'is symbolically represented by the shift from *I* to *we* as a term of self-reference' (see also Schroeder et al., 1995 on 'the "We-ness" factor'). In a similar vein, Brewer claims the following: 'When a distinctive social identity is activated, the collective self dominates the individuated self' (1991: 479). The way in which interests are perceived also plays an important role in the process. Brewer (1979) had already suggested that a reduction in perceived distance on some trait leads to a reduction in perceived distance between individual and collective outcomes. The idea is that people increase the weight they attach to collective outcomes

when some category to which they belong is made salient: outcomes for the group as a whole come to be perceived as one's own. In other words, category salience induces people to focus on collective rather than individual interests, or to perceive 'outcomes for other group members, or for the group as a whole... as one's own' (Kramer and Brewer, 1984: 1045). In a similar vein, Turner et al. (1987: 65) argue that cooperation follows from 'the shared and mutual perception by ingroup members of their interests as interchangeable'.

Brewer's view of the role that social identity plays with respect to ingroup favoritism is hospitable to the idea of agency transformation as it is conceived of in the literature on team reasoning. Moreover, what she says about a switch of focus from personal to collective interests is consistent with the notion of payoff transformation. This means that the key notions of the team reasoning approach survive unscathed from a critical confrontation with relevant findings from social psychology. The question to which I turn now is whether Bacharach's interdependence hypothesis is consistent with the evidence discussed here.

Category salience and the interdependence hypothesis

Bacharach's hypothesis states that strong interdependence stimulates group identification and hence cooperation. Roughly, the claim is that the perception of scope for common gain prompts team reasoning and thereby facilitates cooperation. The empirical findings reviewed in above do not support this hypothesis. The initial claim of the social identity approach was that any arbitrary categorization could induce cooperation. This claim was too strong. The more nuanced claim that has survived empirical scrutiny is that a category needs to be sufficiently salient to induce cooperation. Thus the evidence conflicts with Bacharach's hypothesis. It is not the case that team reasoning is triggered by some characteristic of the interests at stake in a strategic interaction situation (scope for common gain). Instead, it is prompted by a property of the agents present in that situation: a shared social category (that is common knowledge). The Footballer's Problem discussed previously can be used to illustrate this. The very fact that the players of a particular football team wear the same shirt can already trigger group identification. The players need not realize that they have something to gain from working together in order to identify with their team and cooperate.

In an attempt to save the interdependence hypothesis, one might want to point out that cooperation does take place in social dilemma situations in the absence of category salience. This observation, however, cannot be

used to bolster Bacharach's hypothesis. That hypothesis concerns group identification, and there is no reason to believe that the cooperation at issue involves agency transformation or payoff transformation. A wide variety of factors other than group identification can lead to cooperation. Think, for instance, of direct reciprocity. Cooperation can also be explained in terms of people's sense of efficacy, or in terms of framing a situation as involving a gain rather than a loss (see Kerr and Park, 2001 for discussion). Another factor that has recently gained prominence is social value orientation. While 'proselfs' focus on their own gain, possibly in comparison to that of others, 'prosocials' are disposed to cooperate.¹⁰ None of these factors, however, provides a reason for conceptualizing the process that generates cooperation in terms of agency and payoff transformation. Research on category salience does.¹¹

Bacharach (2006: 73–81) discusses the literature on group identification, and suggests that it supports his theory. At the same time, however, he sets aside the empirical findings that social psychologists have produced concerning social categorization. Bacharach discusses a number of factors that have played a prominent role in the social identity approach, including being a member of a social category, having face-to-face contact, and the presence of another group. However, when he gets to the point of deciding what to take away from his review of the literature, he ignores all of the cognitive factors that feature prominently in the social identity approach and zooms in on factors that concern interests, in particular interdependence and common interests (Bacharach, 2006: 82). His argument is that these two features 'connect group identity to characteristics of decision problems' (Bacharach, 2006). This argument, however, is question-begging. One of the questions at issue is exactly whether group identification is triggered by characteristics of decision problems or by the salience of some social category. Against the background of this question, the only conclusion that can be drawn is that Bacharach arbitrarily leaves out features he acknowledges to be important.

Bacharach (2006: 75 and 83–84) also maintains that empirical findings concerning the role of interests in group identification support his interdependence hypothesis. The relevant literature concerns the notion of common fate. Bacharach interprets common fate in terms of collective or common interests. Such an interpretation is apt given the meaning that was attached to the term when it was first introduced. The idea was that having common interests induces a sense of having a fate in common (Rabbie and Horwitz, 1988). More recently, however, common fate has come to be conceived of differently: when two people have a fate in common, their attention has somehow been drawn to the fact that they are in a situation together

(Brewer, 2003). Often this has nothing to do with having common interests. So, due to the fact that the term 'common fate' has changed in meaning, the evidence concerning common fate does not in fact support the interdependence hypothesis.

Recall the earlier example of having won a lottery as a group. The fact that people have this in common says nothing about the payoff structure of the situation they subsequently face. It can, however, readily create a sense of collectivity. The evidence suggests that having a common fate induces a perception of common interests. In this vein Brewer writes: 'Perceived interdependence may be a product of common fate, rather than its source' (2003: 37). This implies that common fate is not a characteristic of the interests at stake in a situation of strategic interaction, but a property of the agents present in such a situation.

Bacharach (2006: 75) supports his interpretation by reference to a passage from a paper by Linnda Caporael (1995) that concerns people stuck in an elevator between floors. Caporael's point, however, is that the very fact that a number of otherwise unrelated people are stuck in an elevator creates a sense of togetherness. This is a matter of common fate and is independent of the interests that the individuals concerned might have in common (for instance, in being rescued). In fact, Caporael contrasts common fate to common interest.¹² The notion of common fate is backward-looking, whereas that of common interest is forward-looking.¹³ Given this more up-to-date interpretation of the term 'common fate', the evidence presented in the literature at issue is in fact in conflict with Bacharach's interdependence hypothesis.

The first conclusion that I draw, then, is that Bacharach's interdependence hypothesis should be rejected. The preceding discussion also allows me to draw the constructive conclusion that the empirical adequacy of the team reasoning approach can be improved significantly by incorporating some notion of category salience. Rather than strong interdependence, the claim should be that the salience and common knowledge of belonging to one and the same social category stimulate group identification. Such category salience triggers the thought 'we are in this together', which commonly results in cooperative behavior.

The assurance hypothesis

Group identification and depersonalized trust

An evaluation of Sugden's assurance hypothesis requires a review of the literature that concerns the role of assurance or trust in group identification

and cooperation. Normative factors have played a role in the explanation of in-group favoritism from the beginning. As early as 1971, Tajfel suggested that people might act on a norm dictating that they should cooperate with fellow group members (Tajfel et al., 1971).¹⁴ Brewer (1979, 2007) emphasizes the importance of normative expectations about in-group members for in-group favoritism, arguing that category salience leads people to trust other members and develop feelings of loyalty. In this context people trust other group members as such, regardless of whether they have interacted with them or whether they are even acquainted with them. Brewer (1981) has coined the term 'depersonalized trust' for such group-based trust that is induced by membership of a social category. It plays an important role in group identification.

Kramer and Brewer (1984) report that when the boundaries of social categories are manipulated and changes are made to the collection of individuals who belong to a particular social category, the trustworthiness ratings of these individuals is also affected (but not other perceptions and evaluations). They also note that individuals 'seem to have been consciously aware of their decision to exercise personal restraint and to feel that this was the right thing to do even though they were also aware that the other group members were not necessarily doing the same' (Kramer and Brewer, 1984: 1056). This suggests that moral convictions play a role in this context. In a similar vein, Kramer and Goldman observe that 'the greater cooperativeness of high trusters was not actually contingent upon reciprocity' (1995: 63). In order to explain this they invoke the notion of group-based or depersonalized trust.

Karp et al. (1993) and Yamagishi and Kiyonari (2000) contrast an expectation of in-group reciprocity with an expectation of direct reciprocity to make a similar point. Direct reciprocity occurs when someone cooperates in response to someone else cooperating. In-group reciprocity is conditional on belonging to the same social category. The psychologists mentioned argue that an expectation of in-group reciprocity, which is a matter of depersonalized trust, is necessary for categorization to lead to in-group favoritism. In a rather ingenious way they separate the effect of an expectation of direct reciprocity from that of an expectation of in-group reciprocity. They let people play simultaneous as well as sequential Prisoner's Dilemmas and compared the amount of cooperation when people played with in-group members to that when people played with out-group members. It turns out that people cooperate more with in-group members as compared to out-group members only in simultaneous Prisoner's Dilemma games. Interestingly, cooperation in the sequential games was in fact greater than in the simultaneous games, which suggests that expectations of direct

	Player 2	
Player 1	Cooperate	Defect
Cooperate	4, 4	1, 3
Defect	3, 1	2, 2

Figure 3. The Assurance Game.

reciprocity had a greater effect than expectations of in-group reciprocity. Yamagishi and Kiyonari conclude: 'In the sequential game, this expectation of in-group reciprocity would be overridden by the much stronger expectation based on direct reciprocity' (2000: 122).

These findings provide a basis for a narrower circumscription of the role of group identification regarding in-group favoritism. Category manipulations had an effect on group identification in the sense that ratings concerning belongingness, commonality, closeness, and liking were all significantly different between in-group and out-group members (except for liking in the sequential game). However, given that in-group favoritism only occurred in simultaneous games, high ratings on these features do not as such explain cooperation. The alternative conclusion that Yamagishi and Kiyonari draw is that 'the social category is the container of the expectations of in-group reciprocity' (Yamagishi and Kiyonari, 2000: 127). This means that category salience leads to in-group favoritism because it generates expectations of in-group reciprocity or depersonalized trust.

The upshot is that category salience affects cooperation in two ways. As discussed in earlier, category salience tends to bring about a switch of focus from individual to collective interests. Secondly, category salience induces depersonalized trust, which means that people cooperate without relying on expectations of direct reciprocity. Both of these effects have been incorporated in a new framework for conceptualizing cooperation in Prisoner's Dilemma situations (Kollock, 1998; Simpson, 2006; Yamagishi and Sato, 1986). The first effect has been developed into the claim that category salience reduces greed and freeriding, leaving fear of being a 'sucker' intact. More technically, the idea is that people are inclined to transform Prisoner's Dilemmas into Assurance Games (Figure 3). The reduction in greed implies that defection as a response to cooperation ceases to be tempting (cf. the literature on goal transformation: De Cremer and Van Vugt, 1999; De Cremer and Van Dijk, 2002; De Cremer et al., 2008; and Kramer and Goldman, 1995). As fear remains in place, people are still motivated to

defect in response to defection. The second effect, that of inducing depersonalized trust, entails that the Pareto-optimal outcome in an Assurance Game is generated. In effect, then, the idea is that some form of payoff transformation explains mutual cooperation when combined with depersonalized trust.¹⁵

The picture that emerges from social psychology, then, is that group identification is a process or mechanism that has category salience as its primary input and cooperation as its output. Category salience leads to an increase in perceived similarity and to conceiving of others and the self as members of a group or a 'we'. Under certain conditions, including common knowledge of membership, this leads to depersonalized trust and a transformation of perceived interests in which collective interests take priority. Again, these ideas are hospitable to the idea that agency transformation and payoff transformation are inherent to group identification. What remains to be seen is whether Sugden's assurance hypothesis is consistent with the evidence.

Depersonalized trust and the assurance hypothesis

According to Sugden's assurance hypothesis, an individual will engage in team reasoning only if she is confident or has assurance that others will do so too. As discussed above, Sugden explicates assurance in terms of having a reciprocal reason to believe that every member of the team endorses and acts on mutually assured team reasoning, and having such a reason requires evidence. The discussion of depersonalized trust above suggests that evidence is not required for the kind of trust that facilitates cooperation. There we saw that category salience suffices for this. Behavioral evidence is not needed for people to rely on others. Consider the Footballer's Problem once again. The very fact that members wear the same shirt can induce group identification and thereby cooperation. Observation of how other players act is not needed for a player to choose right for intercepting a pass. This suggests that the assurance hypothesis should be rejected.

Some of the things Sugden writes, however, reveal that his evidence requirement is not particularly strong. Experience of behavioral regularities in a population suffices as a basis for assurance. The thing to note, though, is that even such a weak requirement is too strong. When a certain category is salient in a particular context, people tend to expect cooperation and assume that other members treat them favorably. The formation of the expectation is a cognitive effect that is not based on observation. Perhaps people are at some level aware of the fact that things work in this way. This

much is suggested by the finding that people prefer to rely on in-group membership as a basis for trust rather than on stereotypes about groups in forming their expectations. This holds even when the in-group stereotype is more negative than the image people have of the out-group (Foddy et al., 2009). Note, however, that even though this line of reasoning supports the idea that empirical evidence is relevant to the role that category salience plays in generating cooperation, the evidence does not concern the behavior of members of a particular population. Instead, it pertains to shared membership of a social category. So it falls short of supporting Sugden's claim that the kind of cooperation he is concerned with requires assurance that is grounded in behavioral evidence.

Another respect in which the social psychological findings differ from Sugden's proposal concerns the content of the expectations. The expectations that people form need not be expectations of actual cooperation; they can also be normative expectations to the effect that members of the social category at issue should cooperate. As discussed earlier, this normative element provides an explanation of the fact that people sometimes cooperate even when they realize the other might not. Because the relevant expectations and behaviors are bound by the social categories that are salient in the context, this line of reasoning also supports the importance of category salience for cooperation.

Sugden has responded to my first point of critique by suggesting that depersonalized trust induced by category salience is a (subjective) form of assurance. If this suggestion is accepted, and I see little reason not to do so, my proposal amounts to a rival characterization of assurance. The evidence suggests that group identification depends on framing rather than on rational and empirical notions such as reason to believe, observation, and experience. And framing is a non-rational process. Note, however, that Sugden's explication of the induction of assurance as a rational process is what makes his proposal particularly attractive to rational choice theorists. The evidence is inconsistent with this explication.¹⁶

The upshot is – and this is the second conclusion of this paper – that the assurance hypothesis should be rejected, or at least qualified. Furthermore, the empirical adequacy of the team reasoning approach can be improved by incorporating depersonalized trust as a central facilitating condition of cooperation. Rather than by evidence of behavior, trust or assurance is induced by common knowledge of shared membership of a particular social category. Because of the link between depersonalized trust and category salience, this supports the earlier conclusion that existing contributions to the team reasoning approach do not yet do sufficient justice to the importance of social categorization processes in cooperation.¹⁷

Conclusion

The team reasoning approach explains cooperation in terms of group identification and takes group identification to be a matter of agency transformation and payoff transformation. Someone who identifies with a particular group switches from thinking in terms of 'I' to thinking in terms of 'we' and comes to prioritize collective interests to individual interests. The empirical literature on group identification within the social identity approach is consistent with this way of explicating group identification. In other words, the evidence is hospitable to the idea that agency transformation and payoff transformation are core ingredients of a group identification mechanism that explains cooperation.

The empirical evidence, however, does not support the claims that have been made within the team reasoning approach about the conditions under which team reasoning takes place. In particular, Bacharach's interdependence hypothesis has to be rejected, and Sugden's assurance hypothesis has to be modified. The source of the problem is the same in both cases: neither Bacharach nor Sugden does sufficient justice to the significance that social categorization plays in cooperation. *Pace* Bacharach, it is not a conflict between individual and collective interests as such that triggers group identification. Instead, the salience of a social category is the key explanatory factor in this process. Category salience also leads to in-group reciprocity or depersonalized trust. Empirical findings suggest that insisting on assurance or observation-based trust, as Sugden does, is too demanding. The upshot is that the empirical adequacy of the team reasoning approach can be substantially enhanced by incorporating the notion of category salience as a key explanatory variable.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgements

I gratefully acknowledge helpful comments from Marilyn Brewer, Laetitia Mulder, Robert Sugden, an anonymous referee, and the participants of the first European Network of Social Ontology meeting in Konstanz in October 2009.

Notes

1. Some game theorists will in fact say that it is a matter of definition that people do not cooperate in situations that are adequately modeled as Prisoners' Dilemmas (Binmore, 1994). From this perspective it makes no sense to try and

account for cooperation in such situations. For such game theorists the puzzle to be explained should be reformulated as follows: people sometimes cooperate in situations in which it appears that it would be individually rational to defect.

Proponents of the team reasoning approach argue that situations that initially have the features of a Prisoner's Dilemma can be transformed into situations in which cooperation is rational. This can be squared with the idea that cooperation is never rational in Prisoners' Dilemmas by maintaining that, even though strictly speaking the situations are not Prisoners' Dilemmas, they are like Prisoners' Dilemmas in that they are situations in which there is scope for common gain.

2. Arguably, joint actions are based on joint intentions. Bacharach (2006) and Gold and Sugden (2007) argue that the team reasoning approach explains how joint intentions are formed (cf. Bardsley, 2007). See Tuomela (2009a, 2009b) for critical discussions of this claim.
3. The terms 'agency transformation' and 'payoff transformation' are due to Bacharach (2006; as discussed above, similar ideas have been proposed earlier in social psychology). Note that agency transformation does not carry any ontological commitments to supra-individual entities. Sugden (2000) emphasizes this by using the term 'team-directed reasoning' rather than the more established term 'team reasoning'. Team-directed reasoning is carried out 'by individuals who take themselves to be members of teams' (Sugden, 2000: 184). Bacharach (2006: 59) proposes that payoff transformation is a matter of ranking act-profiles according to 'a Paretian criterion' that affects the payoff structure of games such as the Prisoner's Dilemma in such a way that mutual cooperation is ranked highest. Sugden (2003: 168; 2008: 403) mentions a more specific proposal according to which payoff transformation is a matter of summing up the payoffs of the team members.
4. The schemas presented are restricted to utility functions that have a unique maximum.
5. Given this conception of group identification it would perhaps be more natural if the team reasoning schemas mentioned agency transformation rather than group identification in their second premises. The second and third premises together would then entail that the relevant agents identify with the group. Note that Sugden (1993, 2000, 2003) does not use the term 'group identification'. Gold and Sugden (2007), however, use it in their presentation of his position.
6. So Sugden does not insist on common knowledge (Gold and Sugden's reconstruction of the team reasoning approach discussed earlier suggests otherwise).
7. As an alternative means for generating assurance, Sugden mentions agreement and the public acts of joint commitment that figure in Gilbert's (1989) work. He also mentions work by D.W. Hodgson, Donald Regan, and Susan Hurley, who rely on moral principles as a basis for team reasoning. I follow Bacharach, Gold, and Sugden in focusing on group identification.
8. Another respect in which their views differ concerns the status of team reasoning itself. Bacharach regards team reasoning as a valid form of reasoning (he takes a schema of practical reasoning to be valid exactly if it is success-promoting,

by which he means that it is instrumental for maximizing utility (2006: 7–10 and 163–164). Sugden does not regard it as either valid or invalid. Instead, he holds that it is a mode of reasoning that one can, but need not, endorse.

9. In fact Bacharach's position on the role of assurance is not fully clear. This may be due to the fact that he passed away before he had finished his book. Gold and Sugden have edited the book and comment on aspects that remain somewhat vague. Commenting on assurance, they suggest that it does play some role in Bacharach's theory given that common knowledge of team reasoning plays a role in the team reasoning schema from the individual point of view (see the conclusion they wrote to Bacharach, 2006: 162). However, they also make the following observation: 'Sugden presents assurance as a fundamental component of team reasoning. ... If team reasoning is to be interpreted in terms of reciprocity, assurance seems to be essential. In contrast, reciprocity does not seem to be central to Bacharach's understanding of team reasoning' (Bacharach, 2006: 165).
10. This coarse-grained distinction was preceded by a fine-grained distinction that included pure individualists and individuals with a competitive orientation on the one hand, and altruists and people with a cooperative orientation on the other. A fifth category that has been distinguished consists of people who are concerned with equality (see Loewenstein et al., 1989; Messick and Sentis, 1985; Van Lange, 1999).
11. In their contribution to the literature on goal transformation, De Cremer et al. (2008) provide empirical support for the claim that group identification triggers payoff transformation in particular in proselves. The notion of goal transformation is rather similar to that of payoff transformation and can be traced back to Kelley and Thibaut (1978).
12. Personal communication with Linnda Caporael confirms this reading of her paper.
13. I thank Robert Sugden for suggesting this way of drawing the distinction to me.
14. Explanations of cooperation in terms of group-specific norms are sometimes opposed to those that invoke group identification. However, it can also be argued that group-specific norms play a role in the process of group identification. This is the view I pursue in this paper. See Bicchieri (2002) for a useful discussion of the relation between group identification and group-specific norms.
15. This proposal concerning payoff transformation differs from those advocated within the team reasoning approach in that it does not require people to adopt identical utility functions. Instead, it involves an asymmetry concerning the ranking of unilateral defection. What the proposals have in common is that collective interests are prioritized over individual interests in that all agents rank mutual cooperation highest.
16. A more radical proposal would be to say that the belief or the presumption that members from the same social category are inclined to cooperate is rational because we know that most people have this inclination as a matter of course. Assurance could then be rational without being based on observation.

17. The team reasoning approach has recently been tested directly. Colman et al. (2008a) present evidence that they take to support the approach. As discussed in Krueger (2008), Sugden (2008), and Van Lange (2008), the evidence is inconclusive, in particular because the findings are amenable to other interpretations (see Colman et al., 2008b for a response).

References

- Bacharach M (1999) Interactive team reasoning: a contribution to the theory of cooperation. *Research in Economics* 53: 117–147.
- Bacharach M (2006) *Beyond Individual Choice: Teams and Frames in Game Theory* (edited by N Gold and R Sugden). Princeton, NJ: Princeton University Press.
- Bardsley N (2007) On collective intentions: collective action in economics and philosophy. *Synthese* 157: 141–159.
- Bicchieri C (2002) Covenants without swords: group identity, norms, and communication in social dilemmas. *Rationality and Society* 14: 192–228.
- Binmore K (1994) *Playing Fair: Game Theory and the Social Contract* (Volume 1). Cambridge, MA: MIT Press.
- Brewer MB (1979) Ingroup bias in the minimal intergroup situation: a cognitive–motivational analysis. *Psychological Bulletin* 86: 307–324.
- Brewer MB (1981) Ethnocentrism and its role in interpersonal trust. In: Brewer MB and Collins BE (eds) *Scientific Inquiry and the Social Sciences*. New York: Jossey-Bass, 345–359.
- Brewer MB (1991) The social self: on being the same and different at the same time. *Personality and Social Psychology Bulletin* 17: 475–482.
- Brewer MB (2003) *Intergroup Relations*. 2nd ed. Berkshire: Open University Press.
- Brewer MB (2007) The importance of being we: human nature and intergroup relations. *American Psychologist* 62: 728–738.
- Brewer MB and Gardner W (1996) Who is this ‘we’? Levels of collective identity and self representations. *Journal of Personality and Social Psychology* 71: 83–93.
- Caporael L (1995) Sociality: coordinating bodies, minds, and groups. *Psychology* 6(1). Available at: <http://www.cogsci.ecs.soton.ac.uk/cgi/psyc/newpsy>
- Colman AM, Pulford BD and Rose J (2008a) Collective rationality in interactive decisions: evidence for team reasoning. *Acta Psychologica* 128: 387–397.
- Colman AM, Pulford BD and Rose J (2008b) Team reasoning and collective rationality: piercing the veil of obviousness. *Acta Psychologica* 128: 409–412.
- De Cremer D and Van Dijk E (2002) Reactions to group success and failure as a function of group identification: a test of the goal-transformation hypothesis in social dilemmas. *Journal of Experimental Social Psychology* 38: 435–442.
- De Cremer D and Van Vugt M (1999) Social identification effects in social dilemmas: a transformation of motives. *European Journal of Social Psychology* 29: 871–893.
- De Cremer D, Van Dijk E, Van Knippenberg D, et al. (2008) Cooperating if one’s goals are collective-based: social identification effects in social dilemmas as

- a function of goal transformation. *Journal of Applied Social Psychology* 36: 1562–1579.
- Foddy M, Platow MJ and Yamagishi T (2009) Group-based trust in strangers: the role of stereotypes and expectations. *Psychological Science* 20: 419–422.
- Gold N and Sugden R (2007) Collective intentions and team agency. *Journal of Philosophy* 104: 109–137.
- Gilbert M (1989) *On Social Facts*. Princeton: Princeton University Press.
- Karp D, Jin N, Yamagishi T, et al. (1993) Raising the minimum in the minimal group paradigm. *Japanese Journal of Experimental Social Psychology* 32: 231–240.
- Kelley HH and Thibaut JW (1978) *Interpersonal Relations: A Theory of Interdependence*. New York: Wiley.
- Kerr NL and Park ES (2001) Group performance in collaborative and social dilemma tasks: progress and prospects. In: Hogg MA and Tindale RS (eds) *Blackwell Handbook of Social Psychology: Group Processes*. Oxford: Blackwell, 107–138.
- Kollock P (1998) Social dilemmas: the anatomy of cooperation. *Annual Review of Sociology* 24: 183–214.
- Kramer RM and Brewer MB (1984) Effects of group identity on resource utilization in a simulated commons dilemma. *Journal of Personality and Social Psychology* 46: 1044–1057.
- Kramer RM and Goldman L (1995) Helping the group or helping yourself? Social motives and group identity in resource dilemmas. In: Schroeder DA (ed.) *Social Dilemmas: Perspectives on Individuals and Groups*. Westport, CT: Praeger, 49–67.
- Krueger JI (2008) Methodological individualism in experimental games: not so easily dismissed. *Acta Psychologica* 128: 398–401.
- Loewenstein GF, Thompson L and Bazerman MH (1989) Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology* 57: 426–441.
- Messick DM and Sentis KP (1985) Estimating social and non-social utility functions from ordinal data. *European Journal of Social Psychology* 15: 389–399.
- Rabbie JM and Horwitz M (1988) Categories versus groups as explanatory concepts in intergroup relations. *European Journal of Social Psychology* 18: 117–123.
- Schroeder DA, Sibicky ME and Irwin ME (1995) A framework for understanding decisions in social dilemmas. In: Schroeder DA (ed.) *Social Dilemmas: Perspectives on Individuals and Groups*. Westport, CT: Praeger, 183–199.
- Simpson B (2006) Social identity and cooperation in social dilemmas. *Rationality and Society* 18: 443–470.
- Sugden R (1993) Thinking as a team: towards 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 (2003) The logic of team reasoning. *Philosophical Explorations* 6: 165–181.
- Sugden R (2008) Nash equilibrium, team reasoning and cognitive hierarchy theory. *Acta Psychologica* 128: 402–404.

- Tajfel H (1970) Experiments in intergroup discrimination. *Scientific American* 223: 96–102.
- Tajfel H (ed.) (1978) *Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations*. London: Academic Press.
- Tajfel H (1981) Social stereotypes and social groups. In: Turner J and Giles H (eds) *Intergroup Behaviour*. Oxford: Blackwell, 144–167.
- Tajfel H and Turner JC (1979) An integrative theory of intergroup conflict. In: Austin W and Worchel S (eds) *Social Psychology of Intergroup Relations*. Chicago, IL: Nelson, 33–47.
- Tajfel H, Billig M, Bundy R, et al. (1971) Social categorization and intergroup behaviour. *European Journal of Social Psychology* 1: 149–178.
- Tuomela R (2009a) Collective intentions and game theory. *Journal of Philosophy* 106: 292–300.
- Tuomela R (2009b) Beyond individual choice, Michael Bacharach. *Economics and Philosophy* 25: 125–133.
- Turner JC (1975) Social comparison and social identity: some prospects for intergroup behaviour. *European Journal of Social Psychology* 5: 5–34.
- Turner JC (1985) Social categorization and the self-concept: a social cognitive theory of group behaviour. In: Lawler EJ (ed.) *Advances in Group Processes* (Volume 2). Greenwich, CT: JAI Press, 77–122.
- Turner JC, Hogg M, Oakes P, et al. (1987) *Rediscovering the Social Group: A Self-Categorization Theory*. Oxford: Basil Blackwell.
- Van Lange PAM (1999) The pursuit of joint outcomes and equality in outcomes: an integrative model of social value orientation. *Journal of Personality and Social Psychology* 77: 337–349.
- Van Lange PAM (2008) Collective rationality: the integrative model explains it (as) well. *Acta Psychologica* 128: 405–408.
- Yamagishi T and Kiyonari T (2000) The group as the container of generalized reciprocity. *Social Psychology Quarterly* 63: 116–132.
- Yamagishi T and Sato K (1986) Motivational bases of the public goods problem. *Journal of Personality and Social Psychology* 50: 67–73.