Theories of Team Reasoning

Alessandra Smerilli
Università Pontificia Salesiana
(Italia)

1. Introduction

The idea of we-thinking, or we-reasoning, is increasingly drawing the attention of more and more economists. In its general formulation, it has been proposed by David Hodgson (1967), Donald Regan (1980), Margaret Gilbert (1989), Susan Hurley (1989), Raimo Tuomela (1995), and Martin Hollis (1998). Within this body of literature, Robert Sugden (1993, 2000, 2003) and Michael Bacharach (1995, 1997, 1999, 2006) have developed analytical frameworks from an economic point of view. The main claim of scholars that analyze we-thinking is that it can be endorsed by people when they face a decision problem. In fact, experimental evidence shows that, especially in some kind of games, such as coordination games, people do endorse we-thinking[1]. However, the way in which we-thinking arises and how it brings people to behave in a particular way in games is a matter that requires further investigation.

[1] See Tan and Zizzo (2008) for a review of experiments.

The two main contributors are Bacharach and Sugden, and they approach the topic in two different ways. Sugden's aim is to show that we-reasoning is a consistent and logical way of thinking, but he does not face the problem of how we-reasoning can arise. He gives some intuitions only about a psychological background based on Smith's analysis of correspondence of sentiments.

Bacharach's theory is based on frames and his never reached aim (because of his death) was to explain we-thinking in terms of Variable Frame Theory (Bacharach 1993). But, as we shall see later, some of his intuitions conflict with the logical analysis he proposes. His logical analysis was developed by Zizzo and Tan (2003, 2008): They introduce a 'game harmony' measure, as a proxy of the extent of cooperation or conflict in games. Game harmony could represent a step forward towards the endogenization of the probability that we-thinking could arise in a decision problem.

In the present paper I offer a review of the most known theories of we-thinking and I focus in particular in Bacharach's theory.

2. Theories of Team Reasoning

All the existing theories of team reasoning share the idea that people can group-identify in front of a decision problem and that they try to reach the best outcome for the whole group, doing their part in the best combination of actions.

This kind of reasoning is called team or we-reasoning and it is the effect of group identification.

Theories of we-reasoning are basically divided into two kinds of approaches: some scholars consider group formation as rational, others do not.

For Bacharach, for example, team-reasoning is a result of a psychological mechanism – the group identification; for Gilbert (1989) and Tuomela (1995) group formation is a result of a mutual commitment; Regan (1980), in his "Utilitarianism and co-operation" proposes a normative theory for moral and rational agents in which the rule to follow is: "What each agent ought to do is to cooperate, with whoever else is cooperating, in the production of the best consequences possible given the behaviour of non cooperators" (p.124).

Susan Hurley offers a theory of 'rationality' of we-thinking. Defining the unit of agency as "the unit the causal consequences of the activity of which are in question" (1898, p.140), Hurley identifies as units the subsystem ('each') or the system ('we') and claims that those units have not to be taken as fixed.

In fact, in front of a decision problem an agent firstly must ask herself: which is the objective in this situation? Subsequently She can choose the unit of agency, which is the most appropriate for the objective: "An adequate theory should help us to understand what the appropriate unit of agency is in various circumstances" (p.146). The consequence of this reasoning is that it is rational to allow different units of agency, hence we-thinking is also rational. A question, however, remains open: "If units of agency are not exogenously fixed, how are units formed and selected? Is centralized information or control required, or can units emerge as needed from local interactions? At what points are unit formation and selection rationally assessable?" (2003, p.165). The last question is very important, because the fact that unit formation is rational has still to be demonstrated. In fact, as Hollis and Sugden point out, in answering the question whether the formation of a unit is a requirement of rationality: the question whether the formation of a unit is a requirement of rationality: "If

we are to stay at all close to account of rationality that derives form Hobbes, Hume, Bentham, Pareto and Savage, we must answer 'No' (Hollis and Sugden 1993, p. 13)". In fact, they explain that according to their account of rational choice theory, a choice is rational in relation to the desires or preferences of the agent who is making the choice: "a choice can be rational only for a particular agent" (ib.). It follows that a theory of rationality cannot give an account of the formation of the unit of agency. Differently from Hurley, who claims that there must be agent neutral goals to be pursued, Elisabeth Anderson (2001) states that the determination of personal identity, which can be plural or individualistic, precedes the choice of the kind of reasoning to be adopted. She makes use of team reasoning in order to give an account of the 'rationality of committed action': "regarding themselves as members of a single collective agency, the parties are

committed to acting only on reasons that are universalizable to their membership"(p.29). She then states the 'Priority of Identity Rational Principle': "what principle of choice it is rational to act on depends on a prior determination of personal identity, of who one is" (p.30). Following the previous principle, Anderson shows that either acting on maximization of expected utility or on team reasoning is a rational act, depending on regarding oneself as an isolated individual or a member of a team. In Anderson's account, then, the determination of personal identity comes before the decision of what principle of choice is in play.

In the following section Bacharach's theory will be analysed.

3. Bacharach's theory of we-thinking

*T*he answers to fundamental questions about coordination and cooperation ... lie in the agent's conception not of the objects of choice, nor of the consequences, but of herself and of the agents with whom she is interacting (Bacharach, M., N. Gold and R. Sugden 2006, p. 70).

This sentence is the starting point of Bacharach's analysis of we-reasoning[2].

---

[2] It is interesting to see how and when Bacharach developed the idea of we-thinking. He started by building the Variable Frame theory (Bacharach 1993), in parallel he was developing a theory of cooperation. In 1995 he introduced the category of 'fellow member reasoner': "Someone who is a member of a natural type T and chooses a certain strategy

if she is sufficiently sure that her interactants are also member of T" (1995, p.1). In this context he tries to link T-membership to variable frame theory and, at for the first time he introduces the 'we' category: "The present paper has made type T membership an issue which type T members think about, and nuanced their capacity to recognize it. An alternative development would make T membership a variable element in players' frames in the sense of variable frame theory: that is, a player might or might not think about the game in terms of whether she and her coplayers belong to T. In the case in which T is the player set, we may put this by saying that a player may or may not think in 'we' terms about how to play the game. The more inclined a player is to 'we' thinking, and the more inclined she takes coplayers to be, the more will fellow-member reasoning be favoured" (p.17). In 1997 Bacharach formally introduces we-thinking, but, at the same time, he claims that we-reasoning is different from team-thinking: "the essence of team thinking is that each player participates in the group-best profile in conditions of common knowledge that they form a 'team' - that is, that they all act in this way" (p.13). The theory of 'team-reasoning' he has in mind was proposed by Robert Sugden in 1993. This was a seminal paper in Sugden's research project on we-thinking. In fact later versions of his theory (see Sugden 2000, 2003) do not assume the participation of all subject to participate and the common knowledge hypothesis. The first published paper in which Bacharach formalizes his theory is the 1999's article about 'interactive team reasoning'. In it Bacharach introduces some elements that we can find in the book, such as, group identification, team reasoning as the effect of group identification, unreliable team interaction, that in the book becomes circumspect team reasoning, etc. Between the '99 article and the book we may find some lecture notes, in which the concepts of agency and 'superagency' begin to ap-

We-reasoning is seen as a powerful 'mechanism' (in Bacharach's words) for solving puzzles about cooperation and co-ordination in game theory (i.e. games like Hi-Lo and Prisoner Dilemma). More in general, in his work Bacharach tries to demonstrate, by showing some evidence[3], that we-reasoning is a valid mode of reasoning and people do endorse it.

His theory is based on frames: if we-frame comes to mind, the subject will group-identify and then she will start to we-reason. A frame can be defined as a set of concepts that an agent uses when she is thinking about a decision problem. It cannot be chosen, and how it comes to mind is a psychological process: "Her frame stands to her thoughts as a set of axes does to a graph; it circumscribes the thoughts that are logically possible for her (not ever but at the time). In a decision problem, everything is up for framing. . . also up for framing are her co-players, and herself". (ib. p. 69).

In Bacharach's framework, then, a person may start to we-reason only if she has 'we' concepts in her frame: in other words, a person firstly recognizes the we-perspective, and then endorses it.

The 'framing' perspective is not the only way in which the theory of we-thinking has been proposed, however. Robert Sugden, for example, has developed a different framework for looking at the problem. In his framework the central concept is the 'common reason to believe'[4]: people who group identify are not committed to reason as a team unless there is a common reason to believe that other agents are doing the same. The psychological side of group identifca-tion in Sugden's theory might be found in his analysis of Smith's 'correspondence of sentiments' (Sugden, R. 2005): 'fellow-feeling' could be seen as the source of group identification. In sum, in Bacharach's framework if people group identify they automatically start to reason like a team, whereas in Sugden's theory people may group-identify, but team reasoning does not follow automatically.

Bacharach's aim is to explain situations in which some people may 'we'- reason and some other may not. In order to model the previous intuition, he assumes that a frame 'we' can come to mind with a probability

---

pear. The book represents an (incomplete, because of his death) attempt to build a complete theory of we-thinking.

[3] Bacharach claims that there are five kinds of evidence: logical, introspective, evolutionary, transcendental and experimental (see Bacharach 2006, pp.145-146).

[4] Common reason to believe is defined as follows: "there is common reason to believe

to a preposition p in a set of individual T if: (i) for all individuals i in T, i has reason to

believe p (i.e. p can be inferred from propositions that she accepts as true n.d.a.); (ii) for

all individuals I and j in T, I has reason to believe that j has reason o believe p; (iii) for

all individuals I, j, and k in T, I has reason to believe, that j has reason to believe that k

has reason to believe p; and so on" (Gold and Sugden 2008, p. 302).

'ω', which represents the probability for a subject to group-identify. The probability ω is common knowledge[5], in fact: "in coming to frame a 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"(Bacharach, Gold and Sugden 2006 p. 163). A context in which some people may group-identify and some may not is seen by Bacharach as an unreliable coordination context, and team reasoning in this context is called *circumspect team reasoning*.

In Bacharach's theory group identification is a framing phenomenon that determines choices by "changing the logic by which people reason about what to do"(ib). If, by reasoning in the individual standard mode (I-reasoning), an agent looks at a decision problem by thinking what it would be the best for her to do, when there is group identification, the agent will think: "What would the best be for us to do?". Basically then, "Somebody 'team reasons' if she works out the best feasible combination of actions for all the members of her team, then does her part in it"(Bacharach 2006 p.121).

Sugden describes team reasoning in a similar way: "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, pp. 182-183).

Trying to go deeper into the matter, Bacharach aims to explain how a frame, in particular a we-frame, may come to mind. In his earlier works (Bacharach, M. 1997, 1999) he proposes that the possibility of team reasoning is related to having 'scope for cooperation' and to the 'harmony of interests'.

In his last book he suggests the (strong) Interdependence Hypothesis, that roughly states: "perceived interdependence prompts group identification".

The perception of interdependence between two agents in a game is given by three factors:

✓ common interest (the agents have common interest in s* over s, if both prefer s* to s, where s*, s are possible state of affairs, or, in a game, possible outcomes)

✓ co-power (nobody can reach s* alone, but both can together)

✓ standard solution (basically Nash equilibrium that may realise s).

Bacharach says that if the two agents have common interest in s* over s and co-power over it, and the standard solution of the game contains outcomes in s (this is the case of PD), then people are stimulated to group-identify. In other words, if s is a possible outcome of the individual rationality, and s* is strictly Pareto-preferred by both the agents, and they have common power to

---

[5] In a previous work (1999), Bacharach has developed a more formalized model, in which each agent can participate or lapse in a team and everyone, before choosing, receive a signal knowing the joint probability distribution of this signal and agent's state (i.e. an agent's signal includes her participation state).

bring about s*, then group identification is stimulated. Or, if the outcome

that can be reached by an individual way of reasoning is Pareto-dominated by another outcome achievable only by thinking as a group, there is space for group identification.

The main Bacharach's purpose is to explain cooperation, seen as a successful group activity (see 2006 p. 69), and the core mechanism for doing that comprehends 'framing', 'common purpose', and `cooperation': "(i) we frame ourselves as members of groups; (ii) . . . perceived agreement of individual goals among a set of individuals favours framing as members of a group with this common goal; (iii) the group framing tends to issue in efficient cooperation for the group goal"(p.90). People then cooperate because they group identify.

4. We-thinking and Variable Frame Theory

Bacharach's (never reached) aim was to explain we-reasoning in terms of Variable Frame Theory (VFT), which he had developed b [6]. Concisely, in VFT a player can intentionally choose an object, or an action, if she has a way of thinking about that object or that action, i.e. he has a frame. Frames can be more or less salient or available, depending on a probability measure on them. A decision

---

[6] See Bacharach 1993, 2001.

rule in VFT is "a mapping from frames to options induced by those frames"(Bacharach, M. 2001a), and an equilibrium for symmetrical games is defined as follows: "the pair (δ,δ) is a variable frame equilibrium if, for each Frame F, the option $\delta(F)$ is subjectively best from the perspective of F against δ as perceived in F" (ib.). In other words the decision rule $\delta(F)$ has to be the best reply against δ.

The intersection between VFT and we-thinking would have been called by Bacharach 'Variable agency theory' (Bacharach 2006, p.59). However, he could not complete the description of 'we'-reasoning in terms of VFT. In fact, there are at least two problems to solve, in order to complete Bacharach's theory: one is related to the way he conceives the 'we' frame, the other is the endogenization of ω.

In Bacharach's circumspect team reasoning, as I have said before, if people group-identify, then the we-frame comes to their mind and they start to we-reason. It seems as though in Bacharach's framing theory there are two aspects that are deeply linked: in framing a situation, the first step is to recognize a frame, that is coming to see it; the second step is endorsing that frame, i.e. reasoning as the frame allows you to do. In Bacharach's theory group identification means not only endorsing a particular way of reasoning, but also coming to see it. The 'compression' between the two aspects of framing is due to the VFT. In it in fact, changing frame does not mean to change the way of reasoning and the decision problem

for a subject is fully determined by the inter-play of his frame and the objective world.

Because of this 'compression', Bacharach in his theory of we-thinking cannot allow people to use more than one frame at a time. In a certain sense, as it has been noticed by Gold and Sugden (Bacharach, Gold and Sugden 2006), in we-frame people become committed to we-reason: "In the theory of team reasoning, an individual who reasons in the 'we' frame is aware of the 'I' frame too (as one of that other players might use) but acknowledge only 'we' reasons. It seems that group identification involves something more that framing in the sense of variable frame theory: the group-identifier does not merely become aware of group concepts, she also becomes committed to the priority of group concepts over individual ones" (p.199).

The fact that people cannot use more that one frame at time and they cannot reason about frames whilst thinking of which frame to endorse, prevents Bacharach to develop analytically one of his intuitions. In fact, taking the most famous game in terms of co-operation, the PD game, as an example, Bacharach says: "In a Prisoner's Dilemma, players might see only, or most powerfully, the feature of common interest and recipro-cal dependence which lie in the payoffs on the main diagonal" (p.86). If this happens, players do cooperate. But, it might be the case that "they might see the problem in other ways. For example, someone might be struck by the thought that her co-player is in a position to double-cross her by playing D in the expectation that she will play C. This

perceived feature might inhibit group identi-fication" (ib).

Here Bacharach seems to have in mind some psychological process which inhibits group identity which is not quite represented by his own concept of interdependence - the idea of 'double-crossing'. The reason this idea does not fit his framework is that double-crossing is the incentive to act on in-dividual reasoning when one believe the other is acting on team reasoning. This re-quires that the player uses both frames at the same time, while thinking about which one to use. Or, in other words, a player, in order to recognize the 'double-cross' threat, should be allowed to imagine himself in a we-frame, and then deliberating to cooper-ate, but at the same time he should use the I-frame by thinking that the other player would take advantage of her. In the first player's conjecture, the other player too should use the we-frame in order to think that the first player could choose to cooper-ate, and, at the same time, she should use I-frame in order to think how 'double cross' the first player.

In the theory of we-thinking the way in which a person reasons (I-mode or we-mode) is a consequence of the perceived frame. So, if a person is in we-frame she cannot reason in the standard theoretical mode, and then she cannot 'see' the double-crossing threat. She may switch from I-mode of reasoning to we-reasoning (if we-frame comes to mind), or not. Bacharach, then, does not seem to take into account the possibility that once we are in the we-frame, we may switch to I-mode of reasoning, or

better, he allows the possibility of switching frame, but does not allow a person to be able to visualize switching frames. And this is why he cannot represent his 'double- crossing' intuition. It seems that when the "we" frame is perceived, it is also perceived as the correct frame or dominant frame, so that once a person sees the world this way she cannot visualize going back to seeing it the other (compare illusions, myths, lies - 'the scales fell from my eyes').

In effect, in one of his unpublished papers (Bacharach, M. 1997), Bacharach allowed for the possibility of the existence of three frames: the I frame, the We frame and the 'S' (superordinate) frame. We and I are called simple frames: "players in them begin their reasoning with the two basic conceptualization of the situation, as 'what shall we do?' problem and 'what shall I do?' problem respectively"(p. 5). A S frame is active when someone manages "during deliberation to see the problem from both the we and the I/she perspectives" (p.14). Although Bacharach allows for the existence of S, based on psychological attainments, he states that we and I perspectives cannot be held simultaneously: "Although we can switch self-identities rather easily, we appear to be unable to inhabit more than one at a time"(p.15). This seems to be a contradiction, and in fact this assumption leads Bacharach to introduce a solution concept that imposes, in his words "a stringent requirement"(p.21). He assumes that I thoughts in S frame generate a personal evaluation, whereas we thoughts generate a group evaluation. The solution concept in the model roughly states that the cooperative option is chosen by a player in S if it is the best (against the mix of the co-player's generated by his varying frame - p.21 -) in group evaluation and not worse than the other option in personal evaluation.

The S-frame intuition of 1997 unpublished paper, however, disappeared in subsequent pieces of work, but also the hypothesis that agent can 'vacillate' between the two frames does not appear in the most recent Bacharach's effort: the book.

Later on, in developing the VFT Bacharach faces the issue of integrability of frames. He says that normally frames are integrable: "It is easy to integrate frames which consist of classifiers such as shape, colour and position: we can easily see a mark as a triangle, as a blue triangle, as a blue triangle on the left,. . . on the other hand. . . a person can see the marks as letters and as geometric shapes, but not at the same time - you can't integrate these two perceptions" (2001, p.6). There exist frames, then, that are non-integrable. 'I' and 'we' frames appear non integrable in Bacharach's words, and when this happens, "the agent may find herself vacillating between the judgments that she should do"(ib.). In spite of these attempts, then, 'double-cross' threat has not been enclosed in Bacharach's analytical framework.

The second unsolved problem is the endogenization of ω. Bacharach tries to endogenize ω, because he sees that the fact that ω is exogenous represents a lacuna in his theory. We may see a clear evolution in his thoughts along the years: in 1997 and 1999

papers he sees **ω** as a function of the gain from cooperation and the harmony of interests: "To endogenize **ω**, and other feature of **ω**, one must show that the payoffs and other constitutive features of the basic game make collective identity salient or otherwise tend to induce team-thinking. The laboratory evidence is promising, as it suggests that group identification may be induced by the 'common problem' mechanism. In addition, it is plausible that **ω** may be an increasing function of certain quantitative features of the payoff structure, such as 'scope for cooperation' and 'harmony of interest'" (1999, p.144). In 2001 he turns to other two features: "We need a link from the game parameters to the onset of group identification. Two of the classic favouring conditions are commonness of interests and commonness of predicament" (2001 b, p.8). In his book he goes deeper into the matter, as we have seen, by stating the interdependence hypothesis. But he does not complete the work: "It may also be that there is a positive relationship between salience and effectiveness: when a feature tending to promote self-identity is highly salient, then if and when it is noticed it is also highly effective. These are empirical speculations; their investigation will be an important part of the future development of the theory of group action" (2006, p. 87).

A step forward on this topic has been made by Tan and Zizzo (Tan, J. and D. Zizzo 2008): in their paper there is an attempt to investigate the relationship between harmony of interests ('game harmony' for them), group identification and cooperation. They claim that game harmony is a good measure of the extent of cooperation or conflict in games. However, in experiments and in real life, we often observe what they call excess conflict or cooperation relative to theoretical predictions. In their framework, it is group identification that can explain excess in-group cooperation and excess out-group conflict. They also believe that the so called 'perceived harmony' can capture the effect of team reasoning on cooperation. This is a theoretical construct and it has not been tested yet. At the same time 'perceived harmony' cannot be inferred by the payoffs of the game, because it is a sort of 'experimental' measure (similarity index in experiments - how subjects think a game is similar to a pure coordination game or to a zero-sum game - is a proxy for perceived harmony). Anyway, the simple game harmony (not the perceived one) is the best existent proxy for what Bacharach has called 'the harmony of interest', and it is entirely derived from the payoffs of the game. Game harmony, defined as "a generic property describing how harmonious or disharmonious the interests of players are, as embodied in the payoffs" (Tan and Zizzo 2008, p. 3), is based on the correlation coefficient between payoff pairs. This measure can be a solution of Bacharach's problem of endogenization of **ω**.

5. Conclusion

In this article I offered a review of theories of team reasoning. In particular I have analysed Bacharach's theory of we-thinking. I focused on his attempt to for-

malize we-thinking in terms of Variable Frame theory. I found that two main problems arise trying to do it. One is linked to the endogenization of ω, the probability that a person may group-identify in a determinate situation. The other one is the impossibility, in Bacharach's framework, to use more than one frame at the same time. This problem prevents Bacharach from representing his 'double-crossing' intuition in the Prisoner Dilemma game, because it requires an agent to have I-frame
in mind when he is we-thinking.

References

BACHARACH, M. (1993): "Variable Universe Games," in Frontiers of Game Theory, ed. by K. Binmore, A. Kirman, and P. Tani. Massachusetts: MIT Press.

BACHARACH, M. (1995): "Co-Operating without Communicating," London.

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

BACHARACH, M. (1999): "Interactive Team Reasoning: A Contribution to the Theory of Cooperation," Research in Economics, 53, 30.

BACHARACH, M. (2001): "Framing and Cognition in Economics: The Bad News and the Good," ISER Workshop, Cognitive Processes in Economics.

BACHARACH, M., N. GOLD, and R. SUGDEN (2006): Beyond Individual Choice. Princeton University Press.

GILBERT, M. (1989): On Social Facts. Routledge.

GOLD, N., and R. SUGDEN (2008): "Theories of Team Agency," in Rationality and Commitment, ed. by P. Di Fabienne, and S. H.: Oxford University Press.

HODGSON, D. H. (1967): Consequences of Utilitarianism. Oxford: Clarendon Press.

HOLLIS, M. (1998): Trust within Reason. Cambridge: Cambridge University Press.

HUME, D. ([1739] 1978): A Treatise of Human Nature. Oxford: Oxford University Press.

HURLEY, S. (1989): Natural Reasons. Oxford: Oxford University Press.

REGAN, D. (1980): Utilitarianism and Co-operation. Oxford: Clarendon 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. (2003): "The Logic of Team Reasoning," Philosophical explorations, 16, 165-181.

SUGDEN, R.(2005): "Fellow-Feeling," in Economics and Social Interactions, ed. by B. Gui, and R. Sugden: Cambridge University Press.

TAN, J., and D. ZIZZO (2008): "Groups, Cooperation and Conflict in Games," The Journal of Socio-Economics, 37, 1-17.

TUOMELA, R. (1995): The Importance of Us: A Philosophical Study of Basic Social Notions. Stanford University Press.

ZIZZO, D., and J. TAN (2003): "Game Harmony as a Predictor of Cooperation in 2 X 2 Games: An Experimental Study," Oxford: Department of Economics, University of Oxford.

ZIZZO, D. (2004): "Positive Harmony Transformations and Equilibirum Selection in Two-Player Games," Oxford: Department of Economics, University of Oxford.