What Does 'We' Want? Team Reasoning, Game Theory, and Unselfish Behaviours

Guilhem Lecouteux

WHAT DOES 'WE' WANT? TEAM REASONING, GAME THEORY, AND UNSELFISH BEHAVIOURS

Documents de travail GREDEG
GREDEG Working Papers Series

GUILHEM LECOUTEUX

GREDEG WP No. 2018-17
https://ideas.repec.org/s/gre/wpaper.html

What Does 'We' Want? Team Reasoning, Game Theory, and Unselfish Behaviours*

Guilhem Lecouteux

GREDEG Working Paper No. 2018-17

Abstract: This editorial presents the main contributions of the theory of team reasoning in game theory, and the issues that remain to be solved before this theory could become a credible alternative to 'orthodox' game theory. I argue in particular that an approach based on collective agency rather than rational choice theory and social preferences offer a scientifically preferable theory of unselfish behaviours, both in terms of parsimony and empirical validation. I review the economic literature on team reasoning, and highlight the contributions of the papers of the present volume to tackle the open issues of the theory of team reasoning.

Keywords: team reasoning, preferences, rationality, cooperation, coordination.

JEL Codes: B41, C72, D70.

* Introduction to a special issue of the *Revue d'économie Politique* on Team Reasoning (REP 128(3), May-June 2018. Affiliation: Université Côte d'Azur, CNRS, GREDEG, France. Postal address: CNRS – GREDEG, Campus Azur; 250 rue Albert Einstein, CS 10269, 06905 Sophia Antipolis Cedex, France. Telephone number: +33 (0)4 93 95 43 74. Email: guilhem.lecouteux@unice.fr ORCID: 0000-0001-6602-7247

'Game theorists of the strict school believe that their prescriptions for rational play in games can be deduced, in principle, from one-person rationality considerations without the need to invent collective rationality criteria, provided that sufficient information is assumed to be common knowledge' (Binmore 1994, 142)

1. Introduction

Classical game theory has trouble explaining how individually rational agents could participate in collectively rational practices, such as contributing to a public good or coordinating on a focal point. The main approaches suggested so far to account for the empirical evidence that people do cooperate in social dilemmas and manage to coordinate in matching games, consist in assuming that players' utility functions are not reducible to their material payoff, but could also depend on others' payoffs[1] or beliefs.[2] In parallel to those approaches – which keep the core assumption that players should ultimately maximise their own utility, whatever defines it – Sugden (1993, 2003) and Bacharach (1999, 2006) suggested as an alternative the theory of *team reasoning* (henceforth 'TR'). A distinctive feature of the theory of TR is that it explicitly allows the possibility of *collective agency*, i.e. that a group of players can be an agent on its own, in addition to the fact that the players constituting the group are themselves agents (Stapleton and Froese 2015). In this regard, the theory of TR constitutes a serious challenge to methodological individualism, and may offer a relatively natural explanation of collective action phenomena.

This editorial will present the main contributions of the theory of TR, and the issues that remain to be solved before TR could become a credible alternative to 'orthodox' game theory. I will start by emphasising the core differences between TR and

[1] Edgeworth already discussed in the 19th century the possibility of 'mixed forms of utilitarianism'. In such cases, the utility function of an individual could be deduced from the material payoffs of all the individuals, 'by multiplying each pleasure, except the pleasures of the agent himself, by a fraction – a factor doubtless diminishing with what may be called the social distance between the individual agent and those of whose pleasures he takes account' (Edgeworth, 1881, p.103). See Sobel (2005) for a review on interdependent preferences.

[2] See Attanasi and Nagel (2008) for a review on 'psychological' games.

the approaches mentioned above, and will argue that a theory of collective agency offers a more *parsimonious* solution to collective action puzzles. The fundamental point is that game theorists should distinguish between the 'theory of preferences' of the players (how they evaluate the outcomes) and their 'theory of choice' (how they choose, once the evaluation of the outcomes is given). This distinction is crucial for the development of an empirically – and scientifically – robust theory of games (section 2). I will then briefly review the economic literature on TR, and highlight the contributions of the papers of the present volume to tackle the open issues of the theory (section 3). Section 4 concludes.

2. Collective rationality or social preferences?

Much of economic analysis is traditionally based on two fundamental assumptions, rationality and self-interest: individuals are assumed to act rationally in pursuit of their objectives, and these objectives are assumed to be defined in terms of their own well-being. Those two assumptions describe (i) how the individual ranks the different options in a situation of choice, and (ii) how the individual chooses, given her ranking of the available options. I will refer to those two points as (i) a theory of *preferences*, and (ii) a theory of *choice*. According to the 'standard' model, preferences are selfish, and individual choice is the result of acting in an instrumentally rational way (the individual chooses her preferred outcome) – I will designate the theory of choice of the standard model as the *rational choice theory* (henceforth 'RCT'). The standard model however leads to puzzling predictions in two simple games, the prisoner's dilemma (PD) and the Hi-Lo game (Gold and Sugden 2007b, pp. 281-285). Consider firstly the PD:

PD	C	D
C	(3;3)	(0;4)
D	(4;0)	(2;2)

It is well known that D is a strictly dominant strategy, and therefore that both players are expected to play D. The puzzle is that experimental evidence highlight that real individuals actually cooperate in unrepeated prisoner's dilemma almost half of the time

(e.g. Sally 1995). The common explanation of such unselfish behaviours is that players are individually rational (the theory of choice is therefore the right one), but that we must extend the definition of preferences to include unselfish motives (see e.g. Fehr and Schmidt (1999) and Charness and Rabin (2002) on inequity aversion). If the standard model fails to predict the right outcome, it is only because the utility functions (the theory of preferences) were ill-defined: the observed choice of the players therefore *reveal* unselfish preferences. I will argue in this section that altering the utility function to include such motives is unsatisfactory for two complementary reasons:

(i) if individual preferences are defined to match the observed behaviours, then the model merely *describes* observed behaviours, and cannot *explain* them;

(ii) if individual preferences are defined as genuine psychological traits (which could therefore explain individual behaviours), then the model will lead to puzzling predictions in apparently trivial games.

2.1 The interpretation of preferences in game theory

If we define payoffs in game theory as revealed preferences, then explaining cooperation in a PD is a nonsense (see e.g. Binmore (2007, pp. 13-15; 2009, pp. 26-29)): cooperating indeed means that the preferences revealed by the players through their choice are not compatible with the payoff structure of a PD! This interpretation of payoffs requires that the payoff structure of a game is a mere *representation* of their choices. The primitive of the analysis is therefore not the preferences of the players, but their choices: the preferences are defined *ex post*, such that the observed choice is *compatible* with RCT. The PD seems puzzling only because we try to deduce the choice of the players from the payoffs (in which case the theory of choice according to which players are individually rational given their selfish preferences is indeed invalidated), while there is no genuine *explanation* of individual behaviour with the revealed-preference interpretation of payoffs. A PD is a mere tautology in the sense that payoffs and choices are conceptually inextricable (Lehtinen 2011, Heidl 2016).

Consider now the Hi-Lo game:

Hi-Lo	H	L
H	(2;2)	(0;0)
L	(0;0)	(1;1)

There are two pure-strategy Nash equilibria: HH is payoff-dominant, while none of the equilibria is risk-dominant. The apparent puzzle of the Hi-Lo game is that, although almost everyone plays H in this game (Bacharach 2006, p.43; Lahno and Lahno, this issue), LL is also a Nash equilibrium. Again, the puzzling aspect of the Hi-Lo game is that the theory according to which players are individually rational given the strategy of the other seems invalidated. The main issue is however not about the theory of preferences (altruistic players would also fail to recognise HH as the only 'correct' outcome), but about the theory of choice. It could indeed be perfectly rational for player 1 to choose L if she believes that player 2 chooses L (and it would be rational for player 2 to choose L if she believes that player 1 believes it, and so on). But if we interpret those payoffs as revealed preferences, then there is no puzzle at all: a game in which all the players systematically choose H is simply a game in which H is always the only payoff maximising strategy (it is therefore not a Hi-Lo game!).

Before exposing the theory of TR, it is therefore essential to clarify the notion of 'preferences' (so as to avoid any confusion, I will also refrain from using the term 'utility'). Heidl (2016) suggests that preferences can be interpreted either in a *mentalistic* or in a *behaviouristic* way. According to the mentalistic interpretation, 'preferences are understood as scientific refinements of the folk psychological concepts of desire and preference' (Heidl 2016, p.26), while the behaviouristic interpretation is that 'preferences are not mental entities but consistent patterns of choices' (Heidl 2016, p.27). Note that the behaviouristic interpretation of payoffs (such as interpreting payoffs as Von Neumann – Morgenstern utilities) does not require a theory of choice. Indeed, if players' choices respect certain consistency requirements, then the players choose *as if* they were maximising their expected payoff. The 'as if' formulation is explicit in the representation theorems of Von Neumann and Morgenstern (1947), Anscombe and Aumann (1963) or

Aumann and Drèze (2009): those theorems do not say anything about how players *actually* choose, but simply that RCT is a *possible* theory of choice, if payoffs are adequately defined. Once we have the (behaviouristic) payoffs, the players are supposed, *by definition of the payoffs themselves*, to maximise their expected payoffs. The issue of referring to a behaviouristic interpretation is twofold. First, game theorists can only describe behaviours, but can neither explain nor predict them (since we define preferences *ex post*, as a representation of a past choice). Second, this makes the theory immune to empirical criticism. So as to develop an empirically and scientifically satisfying theory of games, we should be able to define the preferences of the agents *independently* of their choices: it would then be possible to test how people choose in a game, for given (and observable *a priori*) payoffs.

In this paper, I will therefore refer to a mentalistic interpretation of preferences. A player's *payoff* (the representation of her preferences in a game) must be interpreted as in Sugden (2015), i.e. 'as normalised measures of the values of the relevant outcomes to [the player], *in terms of her own interests, as judged by her*'[3] (p.144, emphasis in original). A player who chooses the strategy that maximises her payoff will be called 'self-interested' or 'selfish'. I will however consider the possibility that people do not necessarily choose the strategy that maximises their payoff: such players will be called 'non-selfish'. Note that if players have genuine sympathetic concerns for other players (e.g. they are unhappy when payoffs are unequal in a PD, because they care about inequity for instance), then those concerns should be integrated in their payoff. This means that a PD is necessarily a game in which D is always an individual best reply, though we allow for the possibility that people do not necessarily play best reply strategies. With this definition of payoffs, we now need to consider different theories of choice, to explain *how* players choose. The two theories of choice I will consider now are RCT and TR.[4] In a PD,

[3] A caveat is necessary. The formulation 'as judged by her' seems to echo the 'as judged by themselves' criterion used in behavioural welfare economics, which Sugden however strongly rejects (Infante *et al*. (2016), Sugden 2017). The payoffs considered here do not represent the preferences the individual would have if she rationally assessed and revised her preferences – in which case we should also consider the possibility of rational strategic commitments (see Parfit 1984 for a detailed argument, and Winter *et al* (2016) or Lecouteux (2018) for related models) – but merely her subjective evaluation of the outcomes at a given moment in time.

[4] A similar model is developed by Courtois *et al* (2015), who assume that the players have the choice between two different 'behaviour rules'. The 'Nash rule' is essentially the same than RCT, and the 'Berge

this means for instance that players can *either* choose to be self-interested (and defect) *or* choose to do their part of the collectively good outcome (and cooperate). Individual preferences are thus observable *a priori*, and, depending on the situation, the players may either follow the recommendations of RCT or TR – other theories of choice could be considered, such as Morton's (2005) 'solution thinking' (see Guala, this issue).

2.2 Puzzling predictions of social preferences approaches

Explaining deviations from the predictions of the standard model requires either (i) another theory of preferences, and/or (ii) another theory of choice. I have argued in the previous section that, to offer a scientifically robust theory of games, we should interpret preferences mentalistically rather than behaviouristically. I now emphasise that changing the theory of choice rather than the theory of preferences offers a more parsimonious theory of unselfish behaviours.

Consider first the PD. A possible explanation of unselfish behaviours is that players display other-regarding preferences, in the sense that the payoff of others matter in their own evaluation of the choice outcome. If we assume that social preferences represent a genuine psychological trait,[5] and that RCT accurately describes how players *actually* choose, then it is indeed possible to observe the coexistence of cooperation and defection in a PD. However, if both players are (say) inequity averse in a PD, *then they should also be inequity averse in other games*. Their concern for fairness, as a psychological trait, should indeed remain stable across games. Consider now the two following games:

G1	C	D

rule' is close to the idea of TR, though the notion of Berge equilibrium represents a situation of 'reciprocal altruism' (Colman et al 2011) which slightly differs from the idea of mutual advantage advanced in TR. The refined version of Berge-Vaisman equilibrium could however offer a more satisfying account of mutual advantage (Crettez 2017).
Rabin (1993) also suggests a model of 'emotional reciprocity' in which players take into account within their decision process other players' intentions. Rabin however recognizes that his theory may lead to unsatisfactory results in sequential games (Rabin 1993, p.1296; see also Sugden 2011, pp.13-14).

[5] Another interpretation would be that social preferences represent the compliance to what the individual thinks is the social norm – in which case the preferences would be defined behaviouristically (players choose as if they were maximising their social preferences, though their choice is the result of a rule-following, rather than maximising, behaviour). See e.g. Binmore (2010).

C	($3;$3)	($0;$4)
D	($4;$0)	($2;$2)

G2	C	D
C	($3;$3)	($4;$0)
D	($0;$4)	($2;$2)

If both players only care about their own monetary gain, then G1 is a PD and both players should play D if they are payoff maximisers. The second game G2 corresponds to G1 with inverted monetary gains: if both players only care about their own monetary gains, then they should both play the dominant strategy C if they are payoff maximisers.
Suppose now that both players are averse to inequity, and that their payoff functions – as in Fehr and Schmidt (1999) – are:

$$\Pi_i(x;y) = \pi_i(x;y) - \alpha_i \max\big(\pi_j(x;y) - \pi_i(x;y); 0\big) - \beta_i \max\big(\pi_i(x;y) - \pi_j(x;y); 0\big)$$

With $\pi_i(x;y)$ the monetary gain of individual i, $0 \leq \beta_i < 1$ and $\alpha_i \geq \beta_i$. Suppose that $\alpha_1 = \alpha_2 = 1$ and $\beta_1 = \beta_2 = 2/3$. The payoffs for those two games are:

G1'	C	D
C	(3;3)	(-4;4/3)
D	(4/3;-4)	(2;2)

G2'	C	D
C	(3;3)	(4/3;-4)
D	(-4;4/3)	(2;2)

If both players are averse to inequity (their preferences are given by the games G1' and G2') and payoff maximisers, then (C;C) and (D;D) are Nash equilibria in both games. Inequity aversion seems to offer a relevant theory of unselfish behaviours in G1 (since it accounts for the empirical observation that some players cooperate while others defect), but it also implies a puzzling prediction in G2. Although C is a strictly dominant strategy in monetary gains and leads to a Pareto optimal profile, our two inequity-averse and payoff-maximising players could indeed also coordinate on the equilibrium (D;D). This illustration highlights that explaining cooperation in G1 as the combination of other-

regarding preferences and payoff maximisation is probably inadequate, because we should also expect some players to choose D in the game G2.[6]

If inequity aversion remains stable across games, then assuming the existence of other-regarding preferences is probably not a satisfying approach to explain why the players are likely to play C in both games, but D only in the first one. A first solution would be to assume that social preferences are not stable across games (which gives a behaviouristic interpretation to social preferences), or that preferences remain stable, and that the theory of choice may depend on the game situation. A theory of unselfish behaviours should therefore encapsulate the idea that the individuals do not always interact in the same way and, for given preferences, may present different *intentions*.[7]

The literature on psychological games (Geanakoplos *et al* (1989), Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006), Falk *et al* (2008)) aims at incorporating different intentions in game theory, by assuming that individual payoffs can depend on players' actions *and the beliefs* about those actions. Cooperation and coordination could therefore be a matter of beliefs about each other's actions, and not merely about other-regarding preferences. A difficulty with psychological games is that – just like with the behaviouristic interpretation of payoffs – the payoff functions that the players are maximising are not clearly interpretable (and not observable *a priori*). Furthermore, it may seem awkward to keep the assumption of payoff maximisation while the very idea of various intentions seems to imply that players, for given payoffs, do not always choose the same action. The position put forward in this paper is that intentions are not reducible to a parameter that could be integrated into payoff functions – otherwise we simply go back to the behaviouristic interpretation of preferences – but should represent distinct modes of reasoning, i.e. different theories of choice, for given preferences.

---

[6] I must recognise that I did not test experimentally this prediction, but I am confident that almost everyone would choose C in G2.

[7] This focus on intentions rather than preferences is supported by the experimental findings of McCabe *et al* (2003), who report data from trust games suggesting that intention-based models (psychological games) offer better predictions than outcome-based models (models with social preferences).

2.3 Collective agency and team reasoning

In standard game theory, the schema of practical reasoning (Gold and Sugden 2007a,b) of a *I-reasoner* – an individual whose theory of choice is RCT – can be described as follows:

(1) I must choose between 'C' and 'D'
(2) If I choose 'C', the outcome is $\Pi(C)$
(3) If I choose 'D', the outcome is $\Pi(D)$
(4) I want to achieve $\Pi(D)$ more than I want to achieve $\Pi(C)$
    ⇨ I should choose 'D'

I-reasoning implies that, for given beliefs about the strategy of other players, a rational individual should play her best reply, i.e. the strategy maximising her expected payoff. In a PD, I-reasoners should for instance always choose 'D'. Consider now the schema of a collective instrumental reasoning:

(1) We must choose between 'CC', 'CD', 'DC' and 'DD'
(2) If we choose 'CC', the outcome is $\left(\Pi_1(CC); \Pi_2(CC)\right)$
(3) If we choose 'DD', the outcome is $\left(\Pi_1(DD); \Pi_2(DD)\right)$
(4) If we choose 'CD', the outcome is $\left(\Pi_1(CD); \Pi_2(CD)\right)$
(5) If we choose 'DC', the outcome is $\left(\Pi_1(DC); \Pi_2(DC)\right)$
(6) We want to achieve $\left(\Pi_1(CC); \Pi_2(CC)\right)$ more than we want to achieve $\left(\Pi_1(DD); \Pi_2(DD)\right)$
    ⇨ We should not choose 'DD'

The conclusions of those two valid reasonings are in contradiction: while two I-rational players should play DD, two collectively rational players should not. According to Gold and Sugden (2007a,b), this must be because their premises are mutually inconsistent. In the former case, the unit of agency is 'I', whose motivation is achieving *my* preferred outcome. In the latter however, the unit of agency is 'we', whose motivation is achieving

our preferred outcome. Although 'I' prefer to defect and 'you' prefer to defect, 'we' prefer to cooperate: our collective intention to maximise our individual payoffs is therefore not reducible to the sum of our individual intentions to maximise our individual payoffs.

The possibility of collective agency means that individuals may conceive themselves not as 'individuals' but as 'members of a team'. The mode of reasoning is radically different: a *team-reasoner* is not pursuing her personal objectives, but the objectives of her team. From an individual perspective, the basic features of the reasoning of a team-reasoner may be described as follows:

(1) We are the members of a group S
(2) Each of us identifies with S
(3) We must choose between 'CC', 'CD', 'DC' and 'DD'
(4) We want to achieve $\left(\Pi_1(CC); \Pi_2(CC)\right)$ more than we want to achieve $\left(\Pi_1(DD); \Pi_2(DD)\right)$, $\left(\Pi_1(CD); \Pi_2(CD)\right)$, or $\left(\Pi_1(DC); \Pi_2(DC)\right)$
(5) Each of us wants to choose what we prefer
    ⇨ Each of us should choose 'C'

The novelty of this schema is that each of us is doing her part in the joint action resulting from our collective intention to satisfy our preferences. A team reasoner is pursuing the objectives of the team (which must be properly defined, see below for a discussion), and chooses her strategy as if she were the member of a coalition – in which an imaginary supervisor attributes to each individual her 'part' of the collective strategy profile:

> Choosing as a member of a team entails not only being motivated by the team objective, but also a different pattern of reasoning: an agent who 'team reasons' computes, and chooses her component in, a *profile* evaluated using the team's objective function (Bacharach, 1999, p.117, emphasis in original)

The difference with social preferences approaches is that TR is neither selfish nor altruistic: it represents individuals as reasoning together about the achievement of common goals (Sugden 2011). It is not the preferences of the individuals that must be

revised, but their theory of choice. TR may thus offer a more satisfying explanation of cooperation based on reciprocity rather than altruism: if P1 cooperates in a PD, it is probably not because she wants to maximize the payoff of P2, but because she expects P2 to cooperate too (so that they will be able to achieve *together* the 'good' outcome). Cooperation is not a matter of altruism (which is one-sided): it is the process of a group of individuals working together to promote a mutual advantage. Although I do not question the existence of altruistic motives or concerns for inequity aversion (which are included in the payoff function), the point here is that cooperation in a social dilemma is probably more a question of reciprocity and teamwork rather than of altruism.

As an illustration, consider the two games G1 and G2 discussed above:

G1	C	D
C	($3;$3)	($0;$4)
D	($4;$0)	($2;$2)

G2	C	D
C	($3;$3)	($4;$0)
D	($0;$4)	($2;$2)

Suppose that the players' payoffs are equal to their monetary gains (they ultimately care only about their monetary gain), and that it is commonly known that they are team reasoners (the case of *unreliable teams*, when the types of the players are not known, is analogous to Bayesian games[8]). Each player therefore perceives the games as the choice of 'us' (henceforth *N*) against nature. From the perspective of team reasoners, G1 and G2 are as follows:

G1''	CC	CD	DC	DD
N	(3 ;3)	(0 ;4)	(4 ;0)	(2 ;2)

G2''	CC	CD	DC	DD
N	(3 ;3)	(4 ;0)	(0;4)	(2 ;2)

[8] See Bacharach (1999) and Lecouteux (2015, chapter 6) for a discussion of the similarities and differences between 'unreliable team interactions' and Bayesian games.

Suppose also for convenience that the players in *N* mutually recognise the vector of payoffs (3;3) as preferable to (2;2), (0;4) and (4;0). We can easily justify that (3;3) is preferred to (2;2) by an argument of Pareto dominance, while the preference of (3;3) over (4;0) and (0;4) can be justified either by appealing to a utilitarian principle (as maximising the sum of monetary gains and therefore of payoffs) or a principle of fairness. In those two games, the optimal choice for *N* is CC. Knowing the optimal strategy profile, each player therefore plays her part of this profile, C.

TR thus offers a relatively simple explanation of cooperative behaviours in PD, but also – and unlike psychological games[9] – in sequential games like the trust or the Centipede games. The team *N* indeed chooses (alone) the optimal path of strategies, and the players of the team are committed to respect this path. The paradoxes associated to backward induction thus disappear (see Hollis (1998, p.137) analysis of the 'Enlightenment trail' for a similar argument).

Furthermore, TR offers a very natural and straightforward solution to the Hi-Lo game: while I-reasoners cannot be certain to coordinate on the payoff dominant profile, team reasoners can rationalise the selection of the payoff dominant equilibrium. In this game, 'we' (as the unit of agency) wants to obtain the Pareto-dominant equilibrium HH: as a member of the team 'we', I shall therefore play H. An important element in TR is indeed that 'players who think as a team do not need to form expectations about one's another's actions' (Sugden 1993, p.87). Team reasoners are able to avoid the infinite regress faced by I-reasoners: unlike within standard game theory in which players treat the actions of others as given, the expectations of the team reasoners are about mutual team membership and not actions (see Hédoin (this issue) for a discussion of the epistemic conditions of TR). If it is common knowledge that all the players are team reasoners,[10] then they do not need to form beliefs about the action of the others: they are indeed choosing the collective profile from the perspective of the team, and are then committed to play their part of this profile. Knowing that the others are team reasoners is a sufficient

[9] See Colman (2003) and Carpenters and Matthews (2003) for a discussion on this point.

[10] Note that assuming common knowledge of TR is not more demanding from an epistemic perspective than assuming common knowledge of rationality.

reason to ensure them that the others will effectively play their part of the collective profile.

By offering a single explanation of collective behaviours in both cooperation and coordination games, TR may offer a better foundation for a parsimonious theory of unselfish behaviours. Furthermore, a theory based on intentions with given payoffs can be tested experimentally, and do not require referring to a slippery notion of 'all-things-considered' preferences. Those two elements suggest that this kind of approach may be a scientifically preferable alternative to models based on RCT with behaviouristic preferences.

2.4 Common or collective interests?

According to RCT, a player *i* asks herself 'what do I want, and what should I do to achieve it?'. Team reasoners, on the other hand, ask themselves 'what do we want, and what should I do to help achieve it?'. Once the best profile for the team has been identified, each team member plays her part of this strategy profile. The definition of the team's best profile remains however an open question: although it seems relatively obvious in several common games (e.g. HH in Hi-Lo, CC in the PD above), this is generally not the case, for instance in non-symmetric games. Sugden and Bacharach offer two slightly different approaches to this issue and the definition of the 'common interest' of team reasoners.

In Bacharach's theory, the team's interests are represented by a group payoff function $U_S(x), S \subseteq N$, which is derived from the individual payoff functions $\Pi_i(x), i \in S$. The objective of the team is to choose the strategy profile $x_S \in X_S$ that maximises the group payoff function U_S (Bacharach 2006, pp. 87-88). TR is thus fundamentally similar to RCT, since teams (rather than individuals) should be instrumentally rational and payoff maximisers (see Sugden (2015) and Lahno and Lahno (this issue)). Bacharach suggests that the group payoff function should respect the principle of 'Paretianness': if $\Pi_i(x) \geq \Pi_i(x'), \forall i \in S$, then $U_S(x) \geq U_S(x')$. He then argues that 'in circumstances in which nothing is perceived by individual members about other individual members beyond the facts recorded in a bare game representation, principles of fairness such as those of Nash's

axiomatic bargaining theory will be embedded in $[U_S]$' (Bacharach 2006, p.88). Bacharach's theory of TR remains essentially faithful to RCT: a team is a kind of 'super-agent' who is individually rational with respect to its payoff U_S, and the members of the team are simply committed to play their part of what the 'supervisor' of the team identified as the best profile.

Sugden's approach to TR, on the other hand, does not require this kind of collective entity, and remains fundamentally individualistic:

> This way of thinking about the good of the team does not fit well with the idea of intentional cooperation for mutual benefit that I have suggested is at the heart of practices of trust. [...] intending that each player benefits is not the same thing as intending the benefit of the team of players, considered as a single entity. To put this another way, intending to promote the *common* interests of team members is not the same thing as intending to promote the *collective* interests of the team. The former intention is cooperative in a sense that the latter is not. (Sugden 2015, pp. 153-154).

Sugden (2011, pp. 14-17) argues that people may either be motivated by the satisfaction of their own interest or be motivated by the pursuit of a 'mutual advantage'. Individuals are 'self-interested' in the sense that they ultimately care about their own payoffs, but they do not choose a payoff-maximising profile: *each* player indeed intends to reach a profile that is good for *both* – this is typically the case in market transactions, in which the gain of each individual can be realised only if the exchange takes place, i.e. only if the other also gains from the transaction. The theory of preferences is therefore the same than in the standard model, but the theory of choice is not RCT any more: the procedure that each player applies to identify her choice is indeed distinct from payoff maximisation. In a preliminary sketch of a formal theory of TR, Sugden (2011, p.15) considers a game $G = \langle N, X, \Pi \rangle$, with a non-cooperative default profile $x_0 \in X$ (which has to be properly defined) and an alternative profile $x^* \in X$. If the following conditions hold for player *i*:

- *Mutual gain*: every individual benefits from playing x^* rather than x_0;
- *Fairness*: the benefits gained from x^* are distributed 'reasonably fairly';

- *Assurance*: *i* has reason to expect that every other player *j* will play her component x_i^* of the profile x^*;

Then *i* will play x_i^* with the intention of participating in fair cooperation, *if she is sufficiently motivated by mutual advantage*. There is therefore no actual 'team' with its own interests according to Sugden, but simply a group of individuals actuated by a specific motivation, the pursuit of a *mutual advantage*. Team reasoners are therefore trying to identify their *common interests* while Bacharach suggests that team reasoners are jointly maximising their *collective preferences*. While Bacharach's version of TR is that the unit 'we' intends to satisfy the objective of the unit 'us', Sugden suggests that the collective 'you and I' intends to satisfy the objectives of the unit 'you' and of the unit 'me'.

3. A brief literature review and open problems

The idea of TR and that individuals may use a distinct 'collective' mode of reasoning had initially been suggested by Hodgson (1967) in a discussion about the distinctive natures of rule and act utilitarianism. Hodgson's argument was then extended by Regan (1980) in his theory of 'cooperative utilitarianism'. The possibility of collective reasoning became central in the literature dealing with collective or joint intentions, with Tuomela and Miller (1988), Searle (1990), Gilbert (1992), or Bratman (1993). Collective intentions can indeed be interpreted as the *outcome* of TR, as in Gold and Sugden (2007a) or Pacherie (2013). Tuomela (2009) however criticises this view, on the basis that a conscious process of maximisation is not compatible with the existence of 'spontaneous' collective intentions. Gold (this issue) investigates this issue and clarifies the relationship between TR, intentions, and intentionality.

The relevance of TR for game theory was first highlighted by Sugden (1991), who then suggested a game theoretical analysis of unselfish behaviours based on TR (Sugden 1993, 2000, 2003). Bacharach (1995, 1997, 1999, 2006) developed a more formal approach using the standard tools of game theory, in which the players could –

unconsciously – switch between different *frames*, the 'I-frame' and 'we-frame'.[11] Many experiments suggest that TR may offer a good explanation of collective behaviours in coordination and cooperation games, though alternative theories are not always ruled out. Furthermore, open problems remain regarding (i) the definition of the objective of the team, and (ii) the rationality of endorsing TR as a mode of reasoning.

A key interest of the theory of TR is that it can explain coordination in matching games, while – despite their 'disarming simplicity and the often overwhelming intuitions we have about how it is rational to play them' (Bacharach and Bernasconi 1997, p.2) – standard game theory fails to provide satisfying approaches to solve them. Schelling advanced the idea that individuals successfully exploit apparently irrelevant features of the choice environment to determine their choices, and may coordinate on a 'salient' profile, that 'stood out from all others'. Various experiments on focal points have confirmed this intuition (e.g. Mehta *et al* 1994a,b, Bacharach and Bernasconi 1997, Bardsley *et al* 2010), and TR offers an interesting rationale for selecting focal points. Schelling's approach to coordination games is indeed that players intend to *solve* the game, and that a solution exists as soon as each player 'does exactly what the other expects him to do, knowing that the other is similarly trying to do exactly what is expected of him' (Schelling [1960] 1980, p.100). The idea of such 'meeting of minds' echoes Sugden's interpretation of TR as the participation to a mutually beneficial practice, and suggests that the mechanism driving coordination is that each player intends to participate in a joint action.[12] An alternative explanation that does not require a notion of collective

---

[11] Tuomela's (2007) distinction between an 'I-mode' reasoning and 'we-mode' reasoning offers a conceptual framework in which Bacharach's theory fits particularly well (see Hakli *et al*, 2010).

[12] A complementary explanation is based on the concept of positive 'social ties': a cohesive group (i.e. individuals sharing many social features of a high importance) may indeed be more likely to solve coordination problems – because they share common references for instance – and to achieve collectively beneficial outcomes (Attanasi *et al* 2014, 2016). This observation is broadly consistent with Bacharach's (2006) argument that players are more likely to identify as the members of a common team if they share a common social identity, and are thus more likely to be team reasoners.

agency is based on level-k theory (Stahl and Wilson 1994, Nagel 1995) and the closely related cognitive hierarchy theory (Camerer *et al* 2004), according to which coordination could be the result of best reply reasoning in responses to potential randomisations by 'naïve' players. Bardsley and Ule (2017) report data that are more consistent with TR, though Crawford *et al* (2008) and Faillo *et al* (2017) both suggest that TR and level-k may provide complementary explanations – and that a single theory is not sufficient to explain coordination. Guala (this issue) also offers a closely related explanation of coordination based on Morton's (2005) idea of 'solution thinking' and simulation theory.[13]

Lahno and Lahno (this issue) investigates whether TR offers a good explanation of coordination in a series of Hi-Lo games in which the behaviour of one of the players is partially randomised. One of their main findings is that the 'opportunistic' interpretation of TR (i.e. that the team – as a superagent – is instrumentally rational given its collective payoff) is not supported by their results, because players tend to stick to accustomed behavioural patterns. This suggests that the definition of the team's objective is not reducible to individual payoff functions (as in Bacharach's framework), and that a more satisfying account of TR should consider the influence of custom and habit on behaviours, through the formation of social norms.

In addition to coordination games, TR may also offer an explanation of collective behaviours in games where the interests of the players are not perfectly aligned. Various experiments indeed suggest that TR may explain cooperative behaviours in social dilemmas (Colman *et al* 2008, Guala *et al* 2009, Butler 2012), the traveller's dilemma (Becchetti *et al* 2009), or Centipede games (Pulford *et al* 2017). This means that we need a theory to determine more precisely the objective of the team, which is relatively trivial only in coordination games, when the interests of the players are identical. Few works however developed general approaches to this problem. Following Bacharach's intention to refer to principles of bargaining solution, Lecouteux (2015) models the choice of the collective preferences as a bargaining game between the players of the team – which can either actually take place during a phase of cheap talk, or is only hypothetical, in line with

[13] See Guala (2016, chapters 7-8), and Larrouy and Lecouteux (2017) for a game theoretical analysis of simulation theory.

the theory of virtual bargaining (Misyak and Chater 2014, Misyak *et al* 2014). Karpus and Radzvillas (2018) suggest on the other hand defining a measure of mutual advantage (in line with Sugden's account of TR) as a relative distance between the gain of the player from cooperation and a reference point. Stirling and Tummolini (this issue) offer a slightly different approach to the problem, thanks to the tools of *conditional game theory* (Stirling 2012; see Ross 2014 for a review, and Hofmeyr and Ross 2016 for a discussion on conditional games and TR). The basic idea of conditional game theory is that the preferences of the players are 'conditional' rather than 'categorical', and depend on the preferences of the other players: it is therefore the social network within which the players are embedded which determines the preferences upon which the players act. In particular, any alteration in the network (such as the arrival of a new player, or a modification of the topology of the network – representing the social influence among players) may trigger more or less cooperative behaviours.

Once we have determined the objective of the team, the complementary issue is the *rationality* of TR, i.e. whether it is rational for an individual to make choices on the basis of TR rather than another mode of reasoning. Bacharach intended to develop a fully-fledged game theoretical framework explaining (i) *how* individuals team reason ('given that someone team reasons […] to what choice does this lead her?', Bacharach 1999, p.142) and (ii) *why* individuals team reason. He unfortunately unexpectedly died before being able to complete his work – his 2006 book was published posthumously, edited by Gold and Sugden. Several explanations of the possibility of TR were advanced by Bacharach: he first suggested that the probability of TR depends positively on 'certain quantitative features of the payoff structure, such as "scope of co-operation" and "harmony of interest"' (Bacharach 1999, p.144). Another explanation is the 'Interdependence Hypothesis' (Bacharach 2006), according to which group identification could result from a perception of interdependence between two agents – see Hindriks (2012) for a critical discussion. The idea of game harmony is developed further by Tan and Zizzo (Zizzo and Tan 2007, 2009, Tan and Zizzo 2008), and Smerilli (2012) offers a model to account for the 'vacillation' between the different frames based on the interdependence hypothesis. Bacharach's main explanation was however evolutionary,

by appealing to multilevel selection theory. Lempert (this issue) reviews the different evolutionary explanations of TR, and highlights some open lines of research.

An important idea in Bacharach's model is that – while TR could be individually costly in social dilemmas compared to a selfish behaviour – it may be individually preferable for individuals to team reason in coordination games such as Hi-Lo, because team reasoners are more efficient in coordinating. This is for instance a key mechanism driving the evolutionary stability of TR according to Amadae and Lempert (2015). Paternotte (this issue) however argues that – contrary to the common claim that team reasoners are necessarily better off than I-reasoners in coordination games – the relative efficiency of TR compared to I-reasoning is relatively narrow when teams are unreliable. Indeed, team reasoners will be able to coordinate if and only if there is a sufficiently high probability that the other players are team reasoners: but if most of the players are team reasoners, then I-reasoners will be able to anticipate the choice of the rest of the population, and will successfully coordinate with the other players. This may seriously jeopardise the evolutionary viability of TR, since it turns out that the evolutionary advantage of team reasoners relatively to I-reasoners in coordination games completely disappears: although their presence in the population tends to create 'more collective order' (Hakli *et al* 2010, p.306), this is beneficial *both* for team reasoners and I-reasoners.

Hédoin (this issue) explores the question of the rationality of TR, and argues that the endorsement of TR in specific contexts can be interpreted as a commitment that can be rationally assessed from an agent-subjective perspective – by referring to a distinction between 'preferences' and 'values' which is broadly consistent with the distinction between 'theory of preferences' and 'theory of choices' introduced above. An important issue is the criterion that should be used to compare the relative efficiency of different modes of reasoning, in particular in Bacharach's framework in which TR implies a payoff transformation. This has important implications for normative economics, which is traditionally based on the satisfaction of individual (stable and context-independent) preferences: considering different theories of choices (or values) indeed gives a central role to the agent, who is not reducible to her welfare any more.

4. Conclusion

At first sight, TR may look like a simplistic – and even naïve – explanation of collective action phenomena. Assuming that individuals can cooperate without being coerced indeed seems to trivialise the central problem of non-cooperative game theory, while game theorists spend decades trying to offer a single coherent theory of human interactions founded on the view of the individual as a self-interested, strategic rational actor. However, it is worth noting that this commitment to rational choice theory is probably not the result of sound epistemological considerations, but rather of the victory of 'rational choice liberalism' over Marxism during the Cold War (Amadae 2003, 2016). The progressive adoption of game theory, together with RCT, in academic circles and public policy significantly affected the development of social sciences, and our perceptions of social interactions and collective decisions. We moved from the Enlightenment idea of the individuals as 'citizens of the world which they construct on liberal principles' (Hollis 1998, p.162), subscribing to a communitarian idea about persons (Rousseau's 'People') while demanding the mutual respect of their individual freedom, to a narrower view of persons as isolated rational actors choosing strategically against each other. There is no 'dilemma' in a PD when the individuals play *with each other*, while it indeed leads to a collectively undesirable outcome when individuals play *against each other*.

By analysing social relations as the joint pursuit of mutual advantages rather than the strategic interactions of isolated welfare-maximisers, TR 'account for the relational nature of humankind' (Smerilli 2012, p.540), and put at the centre of human relations the classical liberal 'no harm' principle, according to which the pursuit of self-interest should not disadvantage others. This idea is central in Sugden's defence of market transactions as instances of mutually beneficial cooperation between individuals. Bruni and Sugden (2008, 2013) indeed argue that participants in a market transaction need not be self-interested, but rather that they intend to bring a mutual benefit through exchange. Both parties benefit from the transaction when the price is strictly between the willingness to receive of the seller and the willingness to pay of the buyer, and they both know that the gain they can obtain depends on the participation of the other to the transaction.

There is nothing naïve in recognising that persons may *actually* think of themselves as members of a community, and be actuated by the group's common goals.

Endorsing this richer view of persons may also offer alternative solutions to collective action problems, such as the management of common-pool resources (Ostrom *et al* 1994, pp.322-328), since appealing to the collective identity of a group may trigger more cooperative behaviours (Nagatsu 2015).

References

Amadae, S. M. (2003). *Rationalizing capitalist democracy: The cold war origins of rational choice liberalism*. University of Chicago Press.

Amadae, S. M. (2016). *Prisoners of reason: game theory and neoliberal political economy*. Cambridge University Press.

Amadae, S. M., & Lempert, D. (2015). The long-term viability of team reasoning. *Journal of Economic Methodology*, *22*(4), 462-478.

Anscombe, F. J., & Aumann, R. J. (1963). A definition of subjective probability. *The annals of mathematical statistics*, *34*(1), 199-205.

Attanasi, G., Hopfensitz, A., Lorini, E., & Moisan, F. (2014). The effects of social ties on coordination: conceptual foundations for an empirical analysis. *Phenomenology and the cognitive sciences*, *13*(1), 47-73.

Attanasi, G., Hopfensitz, A., Lorini, E., & Moisan, F. (2016). Social connectedness improves co-ordination on individually costly, efficient outcomes. *European Economic Review*, *90*, 86-106.

Attanasi G. & Nagel, R. (2008). A survey of psychological games: theoretical findings and experimental evidence, in: A. Innocenti & P. Sbriglia (Eds.), *Games, Rationality and Behavior. Essays on Behavioral Game Theory and Experiments*. Houndmills: Palgrave McMillan, 204-232.

Aumann, R. J., & Dreze, J. H. (2009). Assessing strategic risk. *American Economic Journal: Microeconomics*, *1*(1), 1-16.

Bacharach, M. (1995). Co-operating without Communicating. Working Paper, Institute of Economics and Statistics, University of Oxford.

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

Bacharach, M. (1999). Interactive team reasoning: A contribution to the theory of co-operation. *Research in economics*, *53*(2), 117-147.

Bacharach, M. (2006). *Beyond individual choice: teams and frames in game theory*. Princeton University Press.

Bacharach, M., & Bernasconi, M. (1997). The variable frame theory of focal points: An experimental study. *Games and Economic Behavior*, *19*(1), 1-45.

Bardsley, N., Mehta, J., Starmer, C., & Sugden, R. (2010). Explaining focal points: cognitive hierarchy theory versus team reasoning. *The Economic Journal*, *120*(543), 40-79.

Bardsley, N., & Ule, A. (2017). Focal points revisited: Team reasoning, the principle of insufficient reason and cognitive hierarchy theory. *Journal of Economic Behavior & Organization*, *133*, 74-86.

Becchetti, L., Degli Antoni, G., & Faillo, M. (2009). *Common reason to believe and framing effect in the team reasoning theory: an experimental approach* Econometica Working Papers wp15, Econometica.

Binmore, K. (1994). *Game Theory and the Social Contract*, Vol. I. *Playing Fair*. Cambridge, Mass.: MIT Press.

Binmore, K. G. (2007). *Playing for real*. New York, NY: Oxford University Press.

Binmore, K. G. (2009). *Rational decisions*. Princeton, NJ: Princeton University Press.

Binmore, K. (2010). Social norms or social preferences?. *Mind & Society*, *9*(2), 139-157.

Bratman, M. E. (1993). Shared intention. *Ethics*, *104*(1), 97-113.

Bruni, L., & Sugden, R. (2008). Fraternity: why the market need not be a morally free zone. *Economics & Philosophy*, *24*(1), 35-64.

Bruni, L., & Sugden, R. (2013). Reclaiming virtue ethics for economics. *Journal of Economic Perspectives*, *27*(4), 141-64.

Butler, D. J. (2012). A choice for 'me' or for 'us'? Using we-reasoning to predict cooperation and coordination in games. *Theory and Decision*, *73*(1), 53-76.

Camerer, C. F., Ho, T. H., & Chong, J. K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, *119*(3), 861-898.

Carpenter, J. P., & Matthews, P. H. (2003). Beliefs, intentions, and evolution: Old versus new psychological game theory. *Behavioral and Brain Sciences*, *26*(2), 158-159.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, *117*(3), 817-869.

Colman, A. M. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and brain sciences*, *26*(2), 139-153.

Colman, A. M., Pulford, B. D., & Rose, J. (2008). Collective rationality in interactive decisions: Evidence for team reasoning. *Acta psychologica*, *128*(2), 387-397.

Colman, A. M., Körner, T. W., Musy, O., & Tazdaït, T. (2011). Mutual support in games: Some properties of Berge equilibria. *Journal of Mathematical Psychology*, *55*(2), 166-175.

Courtois, P., Nessah, R., & Tazdaït, T. (2015). How to play games? Nash versus Berge behaviour rules. *Economics & Philosophy*, *31*(1), 123-139.

Crawford, V. P., Gneezy, U., & Rottenstreich, Y. (2008). The power of focal points is limited: even minute payoff asymmetry may yield large coordination failures. *American Economic Review*, *98*(4), 1443-58.

Crettez, B. (2017). On Sugden's "mutually beneficial practice" and Berge equilibrium. *International Review of Economics*, *64*(4), 357-366.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and economic behavior*, *47*(2), 268-298.

Edgeworth, F. Y. (1881). *Mathematical psychics: An essay on the application of mathematics to the moral sciences*. Kegan Paul.

Faillo, M., Smerilli, A., & Sugden, R. (2017). Bounded best-response and collective-optimality reasoning in coordination games. *Journal of Economic Behavior & Organization*, *140*, 317-335.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and economic behavior*, *54*(2), 293-315.

Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness—Intentions matter. *Games and Economic Behavior*, *62*(1), 287-303.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817-868.

Geanakoplos, J., Pearce, D., & Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and economic Behavior*, *1*(1), 60-79.

Gilbert, M. (1992). *On social facts*. Princeton University Press.

Gold, N. (2018). Team Reasoning and Spontaneous Collective Intentions. *Revue d'Économie Politique*.

Gold, N., & Sugden, R. (2007a). Collective intentions and team agency. *The Journal of Philosophy*, *104*(3), 109-137.

Gold, N. & Sugden, R. (2007b). Theories of team agency. In: *Rationality and commitment*, ed. F. Peter & H. B. Schmid. Oxford University Press

Guala, F. (2016). *Understanding Institutions: The Science and Philosophy of Living Together*. Princeton University Press.

Guala, F. (2018). Coordination, Team Reasoning, and Solution Thinking. *Revue d'Économie Politique*.

Guala, F., Mittone, L., & Ploner, M. (2013). Group membership, team preferences, and expectations. *Journal of Economic Behavior & Organization*, *86*, 183-190.

Hakli, R., Miller, K., & Tuomela, R. (2010). Two kinds of we-reasoning. *Economics & Philosophy*, *26*(3), 291-320.

Hédoin, C. (2018). On the Rationality of Team Reasoning and Some of its Normative Implications. *Revue d'Économie Politique*.

Heidl, S. (2016). *Philosophical problems of behavioural economics*. Routledge.

Hindriks, F. (2012). Team reasoning and group identification. *Rationality and Society*, *24*(2), 198-220.

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

Hofmeyr, A., & Ross, D. (2016). Team Agency and Conditional Games. *University of Cape Town (mimeo)*.

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

Infante, G., Lecouteux, G., & Sugden, R. (2016). Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology*, *23*(1), 1-25.

Karpus, J., & Radzvilas, M. (2018). Team reasoning and a measure of mutual advantage in games. *Economics & Philosophy*, *34*(1), 1-30.

Lahno, A., & Lahno, B. (2018). Team Reasoning as a Guide to Coordination. *Revue d'Économie Politique*.

Larrouy, L., & Lecouteux, G. (2017). Mindreading and endogenous beliefs in games. *Journal of Economic Methodology*, 24(3): 318-343.

Lecouteux, G. (2015). Reconciling Normative and Behavioural Economics. PhD thesis, École Polytechnique.

Lecouteux, G. (2018). Micro-microfoundations: Public Policies with Endogenous Preferences. GREDEG Working paper.

Lehtinen, A. (2011). The Revealed-Preference Interpretation of Payoffs in Game Theory. *Homo Oeconomicus*, *28*(3).

Lempert, D. (2018). On Evolutionary Game Theory and Team Reasoning. *Revue d'Économie Politique*.

McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization*, *52*(2), 267-275.

Mehta, J., Starmer, C., & Sugden, R. (1994a). Focal points in pure coordination games: An experimental investigation. *Theory and Decision*, *36*(2), 163-185.

Mehta, J., Starmer, C., & Sugden, R. (1994b). The nature of salience: An experimental investigation of pure coordination games. *The American Economic Review*, *84*(3), 658-673.

Misyak, J. B., & Chater, N. (2014). Virtual bargaining: a theory of social decision-making. *Phil. Trans. R. Soc. B*, *369*(1655), 20130487.

Misyak, J. B., Melkonyan, T., Zeitoun, H., & Chater, N. (2014). Unwritten rules: virtual bargaining underpins social interaction, culture, and society. *Trends in cognitive sciences*, *18*(10), 512-519.

Morton, A. (2005). *The importance of being understood: Folk psychology as ethics*. Routledge.

Nagatsu, M. (2015). Social nudges: their mechanisms and justification. *Review of Philosophy and Psychology*, *6*(3), 481-494.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review*, *85*(5), 1313-1326.

Ostrom, E., Gardner, R., & Walker, J. (1994). *Rules, games, and common-pool resources*. University of Michigan Press.

Parfit, D. (1984). *Reasons and persons*. OUP Oxford.

Pacherie, E. (2013). Intentional joint agency: shared intention lite. *Synthese*, *190*(10), 1817-1839.

Patternote, C. (2018). The Efficiency of Team Reasoning. *Revue d'Économie Politique*.

Pulford, B. D., Colman, A. M., Lawrence, C. L., & Krockow, E. M. (2017). Reasons for cooperating in repeated interactions: Social value orientations, fuzzy traces, reciprocity, and activity bias. *Decision*, *4*(2), 102.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American economic review*, 1281-1302.

Regan, D. (1980). *Utilitarianism and Cooperation*. Oxford: Clarendon Press.

Ross, D. (2014). Theory of conditional games. *Journal of Economic Methodology*, 21, 193–198.

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and society*, *7*(1), 58-92.

Schelling Thomas, C. (1980 [1960]). *The strategy of conflict*. Harvard University Press.

Searle, John (1990): "Collective Intentions and Actions". In: Philip Cohen, Jerry Morgan and Martha Pollack (Eds.). *Intentions in Communication*. Cambridge, MA: MIT Press, p. 401–415.

Smerilli, A. (2012). We-thinking and vacillation between frames: filling a gap in Bacharach's theory. *Theory and decision*, *73*(4), 539-560.

Sobel, J. (2005). Interdependent preferences and reciprocity. *Journal of economic literature*, *43*(2), 392-436.

Stahl, D. O., & Wilson, P. W. (1994). Experimental evidence on players' models of other players. *Journal of economic behavior & organization*, *25*(3), 309-327.

Stapleton, M., & Froese, T. (2015). Is collective agency a coherent idea? Considerations from the enactive theory of agency. In *Collective agency and cooperation in natural and artificial systems* (pp. 219-236). Springer, Cham.

Stirling, W. C. (2012). *Theory of conditional games*. Cambridge University Press.

Stirling, W.C., & Tummolini, L. (2018). Coordinated Reasoning and Augmented Individualism. *Revue d'Économie Politique*.

Sugden, R. (1991). Rational choice: a survey of contributions from economics and philosophy. *The economic journal*, *101*(407), 751-785.

Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfish behavior. *Social philosophy and policy*, *10*(1), 69-89.

Sugden, R. (2000). Team preferences. *Economics & Philosophy*, *16*(2), 175-204.

Sugden, R. (2003). The logic of team reasoning. *Philosophical explorations*, *6*(3), 165-181.

Sugden, R. (2011). Mutual advantage, conventions and team reasoning. *International Review of Economics*, *58*(1), 9-20.

Sugden, R. (2015). Team reasoning and intentional cooperation for mutual benefit. *Journal of Social Ontology*, *1*(1), 143-166.

Sugden, R. (2017). Do people really want to be nudged towards healthy lifestyles?. *International Review of Economics*, *64*(2), 113-123.

Tan, J. H., & Zizzo, D. J. (2008). Groups, cooperation and conflict in games. *The Journal of Socio-Economics*, *37*(1), 1-17.

Tuomela, R. (2007). *The philosophy of sociality: The shared point of view*. Oxford University Press.

Tuomela, R. (2009). Collective intentions and game theory. *The Journal of Philosophy*, *106*(5), 292-300.

Tuomela, R., & Miller, K. (1988). We-intentions. *Philosophical Studies*, *53*(3), 367-389.

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior*. 2nd edition, Princeton university press.

Winter, E., Méndez-Naya, L., & García-Jurado, I. (2016). Mental equilibrium and strategic emotions. *Management Science*, *63*(5), 1302-1317.

Zizzo, D. J., & Tan, J. H. (2007). Perceived harmony, similarity and cooperation in 2× 2 games: an experimental study. *Journal of Economic Psychology*, *28*(3), 365-386.

Zizzo, D. J., & Tan, J. H. (2011). Game harmony: a behavioral approach to predicting cooperation in games. *American Behavioral Scientist*, *55*(8), 987-1013.

DOCUMENTS DE TRAVAIL GREDEG PARUS EN 2018
GREDEG Working Papers Released in 2018

2018-01 Lionel Nesta, Elena Verdolini & Francesco Vona
Threshold Policy Effects and Directed Technical Change in Energy Innovation

2018-02 Michela Chessa & Patrick Loiseau
Incentivizing Efficiency in Local Public Good Games and Applications to the Quantification of Personal Data in Networks

2018-03 Jean-Luc Gaffard
Monnaie, crédit et inflation : l'analyse de Le Bourva revisitée

2018-04 Nicolas Brisset & Raphaël Fèvre
François Perroux, entre mystique et politique

2018-05 Duc Thi Luu, Mauro Napoletano, Paolo Barucca & Stefano Battiston
Collateral Unchained: Rehypothecation Networks, Concentration and Systemic Effects

2018-06 Jean-Pierre Allégret, Mohamed Tahar Benkhodja & Tovonony Razafindrabe
Monetary Policy, Oil Stabilization Fund and the Dutch Disease

2018-07 Pierre-André Buigues & Frédéric Marty
Politiques publiques et aides d'Etat aux entreprises : typologie des stratégies des Etats Membres de l'Union Européenne

2018-08 Jean-Luc Gaffard
Le débat de politique monétaire revisité

2018-09 Benjamin Montmartin, Marcos Herrera & Nadine Massard
The Impact of the French Policy Mix on Business R&D: How Geography Matters

2018-10 Adrian Penalver, Nobuyuki Hanaki, Eizo Akiyama, Yukihiko Funaki & Ryuichiro Ishikawa
A Quantitative Easing Experiment

2018-11 Lionel Nesta & Stefano Schiavo
International Competition and Rent Sharing in French Manufacturing

2018-12 Melchisedek Joslem Ngambou Djatche
Re-Exploring the Nexus between Monetary Policy and Banks' Risk-Taking

2018-13 Dongshuang Hou, Aymeric Lardon, Panfei Sun & Theo Driessen
Compromise for the Per Capita Complaint: An Optimization Characterization of Two Equalitarian Values

2018-14 Gérard Mondello & Evens Salies
The Unilateral Accident Model under a Constrained Cournot-Nash Duopoly

2018-15 Stéphane Gonzalez & Aymeric Lardon
Axiomatic Foundations of a Unifying Concept of the Core of Games in Effectiveness Form

2018-16 Claire Baldin & Ludovic Ragni
François Perroux : Echange pur contre échange composite - Controverses et enjeux de justice

2018-17 Guilhem Lecouteux
What Does 'We' Want? Team Reasoning, Game Theory, and Unselfish Behaviours