The emergence of Co-operative Governance:

the role of conformity to impartially agreed principles of justice♦

Lorenzo Sacconi
University of Trento, Department of Economics and Management
lorenzo.sacconi@unitn.it

Ermanno Tortia
University of Trento, Department of Economics and Management
ermanno.tortia@unitn.it

Abstract

We use Psychological Game Theory (PGT), according to the *conformity to norms preference model*, to analyze aspects of the process of formation and institutional viability in later stages of development of governance in co-operative enterprises. We study how well-known social dilemmas and failures in co-operative governance can be overcome when involved actors conform to simple rules of procedural and distributive fairness. PGT is applied to three basic games – stag-hunt, trust and ultimatum game - transformed in their psychological counterpart (endowed with a different equilibrium set. The Nash Psychological Equilibria, PNE) to analyze three different stages: FIRST, the emergence of the egalitarian cooperative association; SECOND, managerial authority is constituted and each cooperator has to decide to trust or not to trust managerial governance: LAST, the management can unilaterally allocate the surplus amongst cooperative members, which would generate the risk of opportunistic appropriation.

Key words: cooperation; governance; psychological game theory; fairness; reciprocity; stag-hunt game; trust game; ultimatum game.

♦ The authors wish to thank and are indebted to Fabio Milana, who intensively work on the theme of this paper in his Master Dissertation (completed in 2017).

The Emergence of Co-operative Governance.

A psychological game theory approach♦

Extended abstract

Co-operative enterprises have associative nature requiring positive explanation of their process of formation and institutional viability in later stages of development of their governance. The state of the art in economic theory, especially in the new-institutionalist approach, show several shortcomings resulting in inability to single out the distinctiveness of cooperative governance, and the possible origins of its break-down. We use Psychological Game Theory (PGT), according to the *conformity to norms preference model*, to analyse such critical aspects, and how well-known social dilemmas and failures in co-operative governance can be overcome when involved actors conform to simple rules of procedural and distributive fairness. PGT is applied to three basic games – stag-hunt, trust and ultimatum game - transformed in their psychological counterpart (endowed with a different equilibrium set. The Nash Psychological Equilibria, PNE) to analyse three different stages: first, the emergence of the egalitarian cooperative association; second, managerial authority is constituted and each co-operator has to decide to trust or not to trust managerial governance: last, the management can unilaterally allocate the surplus amongst cooperative members, which would generate the risk of opportunistic appropriation. These games are all studied under the hypothesis that: (a) co-operators are allowed to interact in a pre-play cheap talk stage wherein they agree on a norm of fair distribution under the veil of ignorance ; and (b) they may develop expectations and derive utility from reciprocal conformity to the agreed norm. In the PNE of the games, co-operators are able to join an efficient and fair cooperative association (Stag Hunt), prevent abuse of authority (Trust Game) and obtain a fair and efficient allocation of the surplus (Ultimatum Game). The somewhat demanding assumptions concerning the payoff functions required by these results and the compatibility between our results and existing experimental evidence especially concerning the ultimatum game are discussed.

Key words: cooperation; governance; psychological game theory; fairness; reciprocity; stag-hunt game; trust game; ultimatum game.

♦ The authors wish to thank and are indebted to Fabio Milana, who intensively work on the theme of this paper in his Master Dissertation (completed in 2017).

Introduction

This paper has been conceived to overcome existing shortcoming in the explanation of the emergence and development of co-operative governance, as especially found in membership based, mutual benefit organizations such as co-operative enterprises (Birchall, 2010). We start from existing explanations in new-institutionalism, especially the approaches by Elinor Ostron (1990) and Henry Hansmann (1996). We then proceed to highlight their strengths and weaknesses, and their explanatory gaps, to introduce our own approach, whose institutional analysis is based on Psychological Game Theory (PGT).

In the Hansmann (1988, 1996) approach, the feasibility of co-operative governance is depends on the possibility to reduce organizational costs. The strength of co-operative governance is usually recognized in relatively low monitoring costs, due to efficient horizontal control and to the reduction of agency costs, especially when the membership is characterized by homogeneous features and preferences (Hansmann, 1996, 1999). On the other hand, the same approach finds in high coordination and collective choice costs the main its limitation, especially in presence of a heterogeneous membership unless membership.

The main shortcoming of this kind of explanation rest with the fact that the dimension aggregate organizational costs does not justify nor explain institutions as equilibria. In other words the emergence and sustainability of institutional equilibria of co-operative governance is not explained, but assumed. Such explanation is, instead, necessary is progress is to be achieved in the understanding of the process of formation of co-operatives, and on the limitation of this pattern thereof. Several observers identified in the low rate of creation of new co-operatives one of the main limitations in the process of diffusion of this type of enterprise. More generally, the spread of co-operative forms of governance in different enterprise typologies (investor owned and other hybrid forms), can be limited by insufficient understanding of its process institutional emergence. One initial question, therefore, addresses the contractual and governance failures that prevent the creation of cooperatives, also when they can be efficient entrepreneurial ventures. A second shortcoming of existing new-institutionalist approaches is that they do not clearly distinguish between intrinsically costly governance solutions, and the failures of otherwise efficient governance solutions: better understanding and better design of governance rules can alleviate the latter category of costs, but not the former. Finally, the prevailing new institutionalist approaches have paid insufficient attention to the co-evolution of how institutional models with behavioral and preference patterns.

The Ostrom (1990) approach focuses on the governing rules of common-pool natural resources. When dealing with organizations, it concentrates on those organisations that involved in the governance of the commons, and that contribute to overcome the tragedy of the commons though their involvement into wider communal governance structures. It is fundamentally an empirically based, inductive approach, in which governance is understood as complex set of rules, constraints, and sanctions aligning individual to collective objectives. This approach stresses the co-evolution of

cooperative institutions and the corresponding behavioural models and attitudinal patterns, introducing a role for the (intrinsic) motivations of economic actors. This kind of motivations would favour community oriented action, the emergence of trust, and decision making processes based on involvement and deliberation (Sacchetti and Tortia, 2016; Sacchetti, Christoforou and Mosca, 2018). Complementary to this approach, several experimental results show that motivations to cooperate are effectively developed through "cooperative/democratic" agreements reached in ex ante cheap talk (REFERENCES….). Finally, this approach can be nicely reconnected with the social contract literature in which social institutions and structures arise (without the sword) as result of collective agreements under the veil of ignorance (Sacconi, …. ¸other ….).

The eight Ostrom's design principles for the governance of the commons[1] represent a powerful achievement, but have been hardly at all applied to the theory of the production firm (Tortia, 2018), to which our contribution is addressed. While this approach has produced an immense amount of empirical tests, case studies and field studies, and experimental results, its underlying analytical explanatory model has been not fully developed to date. New contributions can widen and deepen it. As a final limitation of this approach, it can be stated that, as a norm, governance models of the commons are most linked found and studied in small communities, while extensions of the model to wider communities and areas require further explanatory effort.

Existing new institutionalist approaches walked a long way in clarifying the conditions under which co-operative governance is most likely to be efficient (e.g. market failures, asset and human capital specificity, role of preferences and homogeneity/heterogeneity of the participants in collective action, costs of decision making processes, horizontal monitoring etc…; cfr. Hansmann, 1999). However, insufficient attention was paid to date to the process of emergence of cooperative governance, and to its criticalities, which may designate the failure of several attempts. Especially, co-operative governance has not been explained in terms of endogenously emerging equilibrium outcomes, if not in the restricted case of market failure. In Hansmann, insufficient attention is paid to governance design by the same actors involved in governance (e.g. members in cooperatives). There has been scant explanation of how co-operators solve governance problems, especially when founding a cooperative, when legitimizing a governance structure, and while living day by day within a cooperative organization wherein managerial slack may occur. These are the issues to which our contribution is addressed. Hansmann himself recently pleaded for intensified research activity on the issue of co-operative governance: "the principal question regarding enterprise ownership is not whether the firms in any given industry will be owned by their patrons or not, but simply which class(es) of patrons will be made the owners. … the answer to this question … also depends heavily

[1] The eight design principles are: 1. Define clear group boundaries. 2. Match rules governing use of common goods to local needs and conditions. 3. Ensure that those affected by the rules can participate in modifying the rules. 4. Make sure the rule-making rights of community members are respected by outside authorities. 5. Develop a system, carried out by community members, for monitoring members' behavior. 6. Use graduated sanctions for rule violators. 7. Provide accessible, low-cost means for dispute resolution. 8. Build responsibility for governing the common resource in nested tiers from the lowest level up to the entire interconnected system.

on the governance structures that firms can and do adopt, which in turn depends on the surrounding legal and institutional environment, including organizational (corporation) law, capital market … development and regulation, and labour law. The full potential role of … cooperatives should become much clearer when scholarship regarding their governance and financing becomes both broader and deeper." (2013, p. 8).

Starting from these premises, our research questions in this paper address the explanation of: (i) the emergence of voluntary cooperative associations and governance legitimization: (ii) the sustainability (in equilibrium) of cooperatives in terms of motivations to invest inter-temporally and to be included in the membership also in the presence of asymmetric distribution of decision making power; (iii) how co-operators may achieve fair distribution of outcomes in the presence of asymmetric distribution of decision making power.

The process of cooperative governance formation and functioning

The first objective of our paper is to study co-operative governance formation, as exemplified by the processes of creation of a co-operative enterprise. We deal with the very basic collective choice to form a fair cooperative association as an alternative to remaining in a state of individualistic lack of cooperation, that is in a state in which all individuals work and produce independently, and to the submission to unfair hierarchies, that is a situation in which the involved individuals participate in production under hierarchical control and without ant decision-making right.

At this stage of our analysis, not only the founding agreement, but also the stability of compliance with it matters. The acceptance and well-functioning of an established governance structure implies that delegation of managerial authority may be necessary to prevent co-operators' free-riding and to favour internal coordination. Co-operative governance entails also the inherent risk of creating room for abuse of authority and hence the trust problem of legitimization of the governance structure emerges. The second step in co-operative governance formation concerns trust in the ability of the co-operative to invest resources effectively, and positive reciprocation (fairness) in the sharing of the obtained surplus. It must overcome both highlighted problems.

Once an authority has been legitimized and hence installed, the risk of abuse may re-emerge in day by day management since, periodically, the surplus of the co-operators' joint effort can be shared among them. A strong enough governance structure endowed with authority (i.e. that cannot be overthrown by a single decision) may ask the co-operator to accept a distribution that deprives him/her from access to the surplus, while allowing the management to expropriate substantial value (for example the annual assembly is asked to accept a balance sheet without any distribution of the surplus). The third step concerns distributive equity and sustainability of governance. This part of the analysis must show how degenerative results are prevented.

Application of the framework and historical reconstruction

The main application of this work concerns the emergence and sustainability of the governance structure of co-operative enterprises. Co-operatives are membership based entrepreneurial organizations, which are created through collective action of members. Co-operative governance arises as solution to collective action problems. The collective element, as based on the presence of a collection of non-investor members, has direct and unavoidable bearing on the creation of co-operative governance, in terms of collective mechanisms of control, participation, inclusion and deliberation. On the other hand, collective action and its governance are themselves liable to dangers of opportunism and break-down due to free-riding and abuse of authority in presence of asymmetric information and contract incompleteness.

The main examples targeted in this paper are the case of a worker cooperative in which members need to find an organizational equilibrium in which (delegated) decisions are taken in a fair and trustworthy manner. Members seek eventually to participate in the surplus of the firm (not only wages but also improvement in the quality of labour, work context and relations). Sharing of the surplus can take place in terms of dividend (also rebates or patronage refunds) distribution, but not necessarily so. In fact, an essential function of governance in worker co-operative is to provide the organization with an autonomous capacity of investment of resources for the development of the firm activity, and for employment preservation and creation, in a situation in which capital is, by definition, scarce. However, typical instances of managerial opportunism and slack include such scrupulous policies not to distribute resources, which are then diverted to increasing managerial benefits and self-dealing.

Simple historical reconstruction shows that informal (at times complex and lengthy) ex-ante agreements between the founder-members of co-operatives have been shown to have fundamental importance in the best known historical cases. Rochdale equitable pioneers (1844, Manchester): decision by a group of spinning workers in the textile industry to create the first sustainable and self-organized (consumer) co-operative of the modern age. The creation of the first (worker) co-operative in the Basque region: it took 8 years of discussion, education and formation stage (1944-1952) between mentor and 5 workers, who eventually created the first co-operative (the Mondragon group counts nowadays about 80.000 employees, most of whom are workers).

A second relevant example concerns a social enterprise (SE, e.g. a social co-operative in the Italian institutional set-up), with multi-stakeholder membership. Distribution in this case does not mean paying the workers high wages or end of the year dividends (also rebates or patronage refunds), but distributing periodically the surplus to finance projects that improve the stakeholders' welfare. As before, managers prevent distribution of the surplus to stakeholders in order to capitalize the SE,

while at the same time appropriating substantial parts of the surplus through assignation to managerial benefits.

In this case, historical reconstruction can be substantiated by projects of multi-stakeholder social enterprises, which always arise out of complex processes of interaction between and within stakeholder groups. This amounts to a pre-commitment stage, in which stakeholder reach informal agreements based on mutual expectation. Social co-operatives in Italy (since the law on the social co-operative in 1991) most of the times emerge out of close interaction between workers, clients, beneficiaries, volunteers and public institutions).

We now turn to the methodological approach used to analyse the governance problems that we just discussed.

Psychological Game Theory with conformity preferences

PGT modifies individual preferences representation through the introduction of an augmented utility function, in which a non self-seeking component is added to the purely self-seeking one, leading the individual to attach positive utility to fair outcomes. Activation of such preferences is conditional on mutual expectations of the correspondent behaviour by the other players in the game (i.e. if I believe that you will conform and I believe that you believe that I will conform to an agreed rule, then I have the desire to conform to the rule itself). Different individual attitudes and preferences imply a different pay-off matrix, leading to new equilibrium choices.

Basic intuition: conformist preference and sketch of the theory

In the presence of conformist preferences, players are faced with a non-cooperative game where only "socially unsatisfactory" Nash equilibria are feasible. In the pre-play communication stage, players agree on a principles (a norm) for solving the ensuing non-cooperative game. Through an "ex ante" (ideal or conjectural) bargaining game under the veil of ignorance, players agree on a specified principle of justice. These agreements are not binding, so that they can be understood as "cheap talk". However, players attach "motivational force" (a desire) to conformity to the principle (a disposition or an attitude to comply). The effectiveness of the disposition to conform is conditional on the other players' expected behaviour and reciprocity. The idea of *conditional conformity* refers to how much player A contributes to full conformity , given A's belief over player B's action. *Reciprocal expected conformity* refers to how much player A believes that player B contributes to full conformity, given A's prediction of B's belief over A' action. Conformity is obtained under the assumption that mutual (and consistent) expectations are formed that other agents will conform, in turn, to the principle.

The idea of inclusion of beliefs among the arguments of the utility function calls for an extension of the standard concept of Nash equilibrium to the new concept of the *Psychological Nash Equilibrium* (PNE). In psychological games, reciprocal beliefs of any order enter the utility function of players. This, intuitively, means that if we all reciprocally believe that we will conform to a given mode of behaviour (i.e. a rule or a certain strategy vector) then an additional positive component enters our payoff function beyond the utility of the predicted personal outcome. Hence, a basic assumption in the formation of PNE is that, when we are in equilibrium, the ongoing beliefs do predict the strategies that are being played, and such beliefs system affects preferences entering the utility function.

Previous experiments working on PNE suggest that alignment of first and second order beliefs and compliance systematically depend on the ex-ante agreement on principles (especially egalitarian ones). This result is not entailed by the theory, but can be considered welcome for its wide application.

Conformist preferences imply apparent inconsistency with mutually expected rationality (expectations are not consistent with utility maximization). This inconsistency in not to be interpreted as "normative fallacy" (from an "ought" - the commitment - it follows an "is" - an expectation), but more simply as (fallible) normative reasoning behind the veil of ignorance, which most of the times leads to an agreement, which is a joint commitment. Such a commitment provides the basis for an intention to act, which may translate into a preference. An agent will not act on a desire-based intention but on a commitment-based intention that engenders a reason-to-act-based preference. This joint commitment is rationally valid by default, i.e. as far as we believe to act as a single body. In fact, the commitment-based explanation constitutes a model for understanding other agents' behaviour. Well known cognitive constraints on reasoning (framing theory and reasoning based on mental models), explain why this model rules out other predictions of other agents' behaviour that could be, in principle, possible. As far as no contradictory evidence unfolds – i.e. by default – subjects expect that the typical agent, having agreed on a rule, will act accordingly. Default conformity can be projected to the simulation/prediction (Alvin Goldman, … ref. …) of other agents' behaviour, and this completes the picture of the emergence of the conditional desire to comply.

Since the agent expects mutual conformity, the commitment-based intention is selected as the one effectively determining her choice, and hence de facto her desire to behave. This not only is consistent with the 'sense of justice' idea, but also explains why so many subjects behave consistently with this idea in experimental settings.

The formal model

The fairness-principle *T* represents formally the agreed ideal (the solution of the game under the "veil of ignorance"),

$$T(\sigma) = N(U_1, \ldots U_N) = \prod_{i=1}^{N} (U_i - c_i) \tag{1}$$

The criterion must be consistent with the idea of a cooperative agreement. We use the Nash Bargaining Solution (NBS), i.e. the Nash social welfare function *T.* We can compute the values of function *T* at each state, and define a complete ordering of states of affairs according to their value for T.

In the augmented utility function, the parameter λ represents the motivational force of the agent's psychological disposition to act on the motive of reciprocal conformity with an agreed norm. This parameter appears in conjunction with an overall index F of conditional and expected reciprocal conformity for each player in each state of the game. This index operates as a weight (taking values between 0 and 1) on the exogenous parameter λ deciding whether λ will actually affect or not (and, if so, to what extent) the player's payoffs.

In the general form of the overall utility function, V_i, we have the linear combination of U_i is the material utility for states σ (described as consequences), with an Ideal component. In this second component, λ_i is an **exogenous psychological** parameter (a **disposition**) that expresses how important the ideal component is within the motivational system of player I; *T* is a fairness principles (Nash Bargaining Solution) defined over states σ; *F* is the composite index of conformity to the principle T expressing both the agent's conditional conformity and other individuals' reciprocal conformity to *T:*[2]

$$V_i(\sigma) = U_i(\sigma) + \lambda_i F[T(\sigma)] \tag{2}$$

The overall utility function in explicit form is:

$$V_i(\sigma_i, b_i^1, b_i^2) = U_i(\sigma_i, b_i^1) + \lambda_i \left[1 + f_i(\sigma_i, b_i^1)\right]\left[1 + \tilde{f}_j(b_i^2, b_i^1)\right] \tag{3}$$

If player *i* expects that player *j* is responsible for the maximal value of T (given what *j* expects about *i*'s behaviour), player *i* is also responsible for the maximal value of T given what he expects about *j*'s behaviour. Then the motivational weight of conformity λ will enter his/her utility function and it will show all the force of the disposition to conform to agreed norms. If λ it is large enough to counterbalance material payoffs, then complying with the principle will yield psychological utility additional to the material payoff.

More specifically, λ is an exogenous psychological or psychosocial parameter that enters in the payoff functions. It may represent the weight of the psychological disposition to conform which is typical to a

[2] In the Ideal component of the utility function, if player i completely conforms and expect that player j conforms too, then the value of its ideal utility is: $\lambda \times 1 \times 1 = \lambda$. If player i does not conform and doesn't expect complete conformity on the part of player j, the ideal utility is: (1-x) (1-y) $\lambda < \lambda$. If conformity is nil at least for one player, the ideal utility is: (1-1)(1-y) $\lambda = 0$.

culture or an evolutionary stage of the species. Empirical studies on the λ level in various cultural contexts could be conducted (REFERENCES?).

We also propose an endogenous explanation, based on the idea that λ depends on *deliberation* (taking in account different arguments for decision). Given an agreement, we have an intentional reason to act, even if there are other reasons to act (e.g. impellent needs). No reason mechanically determines whether or not to undertake the ex-post behavior. John Searle (REF….?) would say that there is a gap representing the true freedom of choice, free will. Within this gap we deliberate the weight of any argument (i.e. λ) according to which deliberation will carry us to a conclusion (OTHER REFERENCES IN THIS STREM?).

In our model, as in other similar models, there is a threshold of λ above which psychological equilibria exist that support conformity to the ex-ante agreement. Under the threshold psychological equilibria may not exist, or they may be uninteresting (i.e. do not support conformity). This happens when the psychological propensity to conform to the fairness norm is too weak, and self-seeking Nash equilibria are dominant.

In our model we can calculate the λ value that is necessary for a psychological equilibrium coherent with the emergence and sustainability of co-operative governance. This requires substantial conformity with the ex-ante agreement over cooperative forms of governance.[3] If, instead, λ does not allow the psychological utility of conforming to the agreement to overcome the material payoff of other outcomes, the self-seeking Nash equilibrium prevails and co-operative governance is either not achievable, or it breaks-down.

Three games to represent the emergence and stability of co-operative governance

We now proceed to briefly introduce the three games, whose psychological game theoretical analysis is then spelled of in depth in the following sections.

The Stag-Hunt game

The stag-hunt game represents the problem of founding a stable and fair initial cooperative association by exiting a non-cooperative status quo that could seem less risky given the high uncertainty of cooperation. The cooperative equilibrium, in which individuals are able to obtain the surplus deriving from joint production (the production function is super additive) in opposed to the "risk-dominant equilibrium", in which cooperation does not take place and the status quo prevail.

We introduced a modified version of the stag-hunt game, in which asymmetrical forms of stag hunting (efficient but unfair) and the relative Nash equilibria are possible. We do so to highlight the

[3] Further justification may require various sociological and historical, evolutionary, but also theory of intentionality.

possibility of obtaining unfair cooperative pact, in which one player ends up in a dominant and one player in the dominated position. In this way, we start analyzing the problem of opportunism and abuse of authority in co-operative associations. Cheap talk in the original position (under the veil of ignorance) creates a Constitutive Pact ("*pactum unionis*") wherefrom conformist preferences do follow. This gives the problem the form of a solvable stag-hunt.

The Trust Game

The Trust Game represents the "problem of trust", which is at stake in the decision of each cooperator to enter or exit the cooperative when a cooperative authority has been appointed, as it may work to the mutual advantage. However, it also entails the risk of abuse by those who are in the authority position. In our interpretation, a "*pactum subjectionis*" allows for the formation of conformist preferences that overcome the incentive to abuse whereby the governance structure results are possible in the psychological equilibrium even if this is not usually the case in the standard trust game.

The extended (sequential) form of the trust game can also be understood in an intertemporal, inter-generational manner: the first generation of members (the trustors) hand over resources to managers to be invested in their own favor, but also in favor of future generation of members. Trustee have to decide whether to really invest in favor of prospective members, or to abuse and appropriate such resources in the absence of control and consent by "future" beneficiaries. This possibility corresponds to reinvestment of surpluses into "indivisible reserves" of capital, which cannot be shared or otherwise liquidated (e.g. through sale of the firm) among incumbent members, and benefit non only incumbent members, but also future ones.

The ultimatum game

In co-operative enterprises distribution of the surplus is reported to be, on average, much more egalitarian than in profit enterprises. Fair distribution of the surplus does not conceal the existence of distributive issues and problems, which do arise, especially in presence of stakeholder groups deprived of decision making power (non-member stakeholders). Distributive issues can arise also between members and decision makers. The right to vote and to elect representatives in the board of directors of the organization can be thought as a mechanism allowing members to have a say in distributive issues and to reject unfair distribution through the rejection of the annual balance sheet. This possibility, however, is likely to bear relevant costs with it, in terms of instability of the governance structure and substantial governance costs (decision making costs, cost of confrontation etc…).

The ultimatum game represents the case of a governance structure that has been established and hence has the authority to offer any level of distribution of the surplus to members, but also the discretion to distribute it in an unfair way, favoring private consumption and status of decision-makers. In such a situation, members would receive a small share of the surplus. Like in the ultimatum

game, they only have the tiny power to reject the offer and to block the decision, but at a great risk and cost to themselves, since they would share the negative consequence of unfair decisions. In our interpretation, while responders (members) would reject the "rational" but very nasty offer of the standard case, given a constitutive contract on a fair principle of distribution, conformist preferences can induce decision makers to offer substantially better conditions to associates than in the standard case.

The modified stag-hunt game

We introduce a modified game matrix in which the players have three options: "genuinely cooperate and eventually submit to a cooperative authority" (hunt the stag), "exercise as much power as possible in an organization" (so as to organize stag hunting, but also obtaining as large a share of the plunder as possible), "stay out from any cooperative enterprise and seek an individual goal" (hunt the hare).

Our modified Stag/ Hunt hence is a coordination game with three Nash equilibria: two asymmetric equilibria (asymmetric stag hunting) that are weakly Pareto dominant; one suboptimal but risk dominant equilibrium (Hare, Hare); one symmetrical and Pareto dominant outcome which is not an equilibrium.

Table 1. Modified stag hunt game, with asymmetric hunt-the-stag equilibria.

		Player B		
		C_1^B	C_2^B	L_3^B
Player A	C_1^A	6,6	**4,7**	0,3
	C_2^A	**7,4**	2,2	0,3
	L_3^A	3,0	3,0	**3,3**

Notes: Three Nash equilibria in pure strategies; risk of abuse of authority (as in the hawk-dove game) in the asymmetric Nash equilibria (7,4) (4,7); stay out strategies are equilibrium strategies; the mutually beneficial Pareto optimal outcome (6,6) is not a Nash equilibrium, but it can be a psychological equilibrium.

Players need to overcome both the risk of abuse under an asymmetrical equilibrium (as in the hawk-dove game), and the risk of non-compliance, which pushes toward the individualist (Hare, Hare) equilibrium. They resort to a co-operative *pactum unionis*: in a pre-play communication stage, players agree on the constitution of an egalitarian cooperative association that allows, if the agreement is mutually kept, to "hunt the hare" sharing equally the outcome. Following the *pactum unionis*, the psychological conformist model of payers' preferences enters the scene by radically changing the equilibrium set of the game.

Table 2. Consistency with the ex-ante agreed principle of cooperation (T = NBS)

		PLAYER B		
		C_1^B	C_2^B	L_3^B
PLAYER A	C_1^A	T = (6-3)×(6-3) = 9	T = (7-3)× (4-3) = 4	T= (0-3)× (3-3) = 0
	C_2^A	T = (4-3)×(7-3) = 4	T = (2-3)× (2-3) = 1	T =(0-3)× (3-3) = 0
	L_3^A	T = (3-3)×(0-3) = 0	T = (3-3)×(0-3) = 0	T = (3-3)×(3-3) = 0

Notes: The value of T at each outcome represents the basic consistency of that outcome with the ex-ante agreed principle of cooperation (T = NBS); "Hunting the hare" is considered as the non-cooperative status quo (as it is risk dominant, it represents a self-assurance behaviour); T is calculated by taking payoffs net of the status quo.

Table 3. Conformity indexes and psychological pay-offs

		PLAYER B		
		C_1^B	C_2^B	L_3^B
PLAYER A	C_1^A	6+(λ); 6+(λ)	4+[$(2/3)$ (λ)]; 7+[$(2/3)$ (λ)]	0, 3
	C_2^A	7+[$(2/3)$ (λ)]; 4+[$(2/3)$ (λ)]	2+[$(1/16)$ (λ)]; 2+ [$(1/16)$ (λ)]	0, 3
	L_3^A	3, 0	3, 0	3+(λ); 3+(λ)

Notes: Conformity to the principle T is maximal for C_1^A givenC_1^B (and vice versa). Hence conformity index is F = 1, and the ideal utility entering the payoff function is *λ for both.* Playing the aggressive hierarchical strategy C_2^A given a cooperative submissive strategy (and vice versa), reduces significantly conformity, i.e. index F = 2/3, hence ideal utility entering payoff is 2/3*λ*. However, playing L_3^A against L_3^B (and vice versa) also locally maximizes the expected T since, given the individualist strategy of the other side, the player has no way to augment the T value by resorting to any other strategy, hence in this case the index is F = 1 and the ideal utility is *λ for both.*

Table 4. Example λ=5

		PLAYER B		
		C_1^B	C_2^B	L_3^B
PLAYER A	C_1^A	**11; 11**	7.33; 10.33	0, 3
	C_2^A	10.33; 7.33	2.3; 2.3	0, 3
	L_3^A	3, 0	3, 0	**8; 8**

Notes: a threshold value of λ set at 5 allows the emergence of the co-operative and symmetrical psychological Nash equilibrium.

Results in the Stag-Hunt game and creation of the co-operative association

The symmetric Pareto optimal outcome re-emerges as a psychological Nash equilibrium, whereas previously, in the self-seeking version of the game, it was not an equilibrium. The resulting psychological game is a "typical" Stag-Hunt with two equilibria (one Pareto optimal, the second risk dominant). However, it should be noticed that a psychological equilibrium exists under the assumption of converging expectations over a given mode of behavior. Hence, if the symmetric stag-hunt equilibrium exists, it is also the one selected by the players' expectations. The question still to be answered concerns why expecting that an ex-ante agreement can activate proper expectations.

The modified Stag Hunt is a coordination problem wherein there are multiple equilibria and the most probable solution is the suboptimal one. Founding the cooperative association gives a (partial solution) to such a problem. It eliminates two asymmetrical equilibria and gives rise to an optimal and fair equilibrium. As far as it "exists", it is also the selected one - i.e. it solves the sub-optimality problem posed by the risk dominant equilibrium (Hare, Hare).

The Trust game

In the trust game (TG), one player, the Trustor, has to decide whether to enter the organization as a member and carry out a specific investment, or to stay out. The second player, the Trustee, has decision making power. In case the sender (trustor) carries out the investment, he may decide to abuse. Only one Nash equilibrium exists in the basic case: the trustor stays out since the trustee would surely abuse. To make sense of the TG in our psychological game we have to assume that before playing the TG, players have ex ante agreed to delegate authority to one of them, who will take the position of the co-op administrator or manager. The delegated authority may then act in mutual benefit or abuse the cooperator's trust, once the authority relationship is entered. The trustor has the first move; its decision amounts to a legitimization of the trustee authority position.

We modify the basic version of the TG and add the idea that the trustor can decide whether to carry our minimal generic investments (for example generic and cheap investments in human capital) that are just sufficient for the organization to survive in a competitive environment, or highly specific investments (e.g. costly and time consuming investments in human capital and professional growth) which can generate higher then minimal value added and allow the organization to gain market shares (cfr. Williamson, 1975, 1985 on perfunctory and consummate co-operation). This modification is aimed at making the game more adherent to reality, since members in co-operatives do indeed decide the degree of specific investments they are willing to carry out (e.g. low investments corresponding to short term labor contracts characterized by perfunctory degree of cooperation and engagement in the production process vs high investments corresponding to long term labor relation

and consummate co-operation with and engagement in the aims and growth perspectives of the organization).

Table 5. Modified trust game in normal form: material payoffs and self-seeking result

		TRUSTEE [B]	
		a	$\neg a$
TRUSTOR [A]	e_*	0 , 9	5 , 6
	e^*	0, 15	6, 9
	$\neg e$	**1 , 1**	1 , 1

Notes: Modified version of the TG: the Trustor, when trusting the trustee, decides also to invest, he has two investment levels: e_* = Low investment (Min.); e^* = High investment (Max.). The only equilibria in material payoffs remains "abuse, stay out"

The two players are assumed to have ex-ante agreed that the trustee will pursue a fair distribution of the maximum surplus (max NBS), conceding a bonus for the manager. Then they assess any outcome in terms of the value of T that realizes at each outcome. T is calculated by considering the surplus net of the status quo, where status quo is the outcome where no authority is legitimized and the trustor does not enter the organization and does not carry out any specific investment.

Table 6. The value of the ex-ante agreed principle T at each end-state of the game

		TRUSTEE [B]	
		a	$\neg a$
TRUSTOR [A]	e_*	$T = (0-1)\times(9-1) = -8$	$T = (5-1)\times(6-1) = 20$
	e^*	$T = (0-1)\times(15-1) = -14$	$T = (6-1)\times(9-1) = 40$
	$\neg e$	$T = (1-1)\times(1-1) = 0$	$T = (1-1)\times(1-1) = 0$

Notes: The players assess the value of T that realizes at each outcome. T calculated as surplus net of the status quo.

Conformity of strategy combinations

If the trustor invests in an optimal way in order to deliver consummate co-operation, and the trustee does not abuse his/her decision making power, the value of the fairness norm T value is maximized. In this case a conformity index equal to one is obtained, and ideal payoff is maximal (corresponds to λ) for both players. Even if the trustee is not abusing, the trustor can nonetheless decide to reduce T by playing a perfunctory strategy, thus obtaining an intermedium value of the conformity index (F =1/2, and ideal payoff 1/2 λ). On the contrary, if the trustee abuses, he minimizes T for whatever choice of entrance by the Trustor. In this case, no ideal utility is obtained. However, if the trustor responds by staying out of the cooperative pact to prevent abuse, the trustee cannot do any better for improving the level of T. In this degenerative case, the trustor maximized T by staying out, since playing e would reduce T. The conformity index is again maximal and and the ideal utility is λ for both players.

Table 7. Conformity indexes and psychological pay-offs

		TRUSTEE [B]	
		$\boldsymbol{a}$	$\neg\, \boldsymbol{a}$
TRUSTOR [A]	$\boldsymbol{e}_*$	0, 9	5 + [(1/2) (λ)] , 6 +[(1/2) (λ)]
	$\boldsymbol{e}^*$	0, 15	6 + (λ), 9 + (λ)
	$\neg\, \boldsymbol{e}$	1 + (λ) , 1 + (λ)	1 , 1

Notes: If the Trustor plays e^* given $\neg a$ by the Trustee (and vice versa) the T value is maximized, hence the conformity index is F =1 and the ideal payoff is λ for both. The Trustor can also reduce T by playing e_*. In this case F =1/2, and ideal payoff 1/2 λ. If the Trustee plays a, T is minimized for all choices of the Trustor. No ideal utility is obtained, unless Trustor responds $\neg\, e$ to a. T cannot be improved (Max = Min). The Trustor is maximizing T given a since playing e would reduce T (F =1 and ideal utility equal to λ for both).

To exemplify, we consider a value of λ equal to 7, which is the threshold at which the psychological Nash equilibrium is obtained. The pair "**entering, not abusing**" become a psychological equilibrium, which corresponds to full acceptance and legitimization of the authority and of the governance structure, as far as this is not abusive. At the same time, "**staying out, abuse**" is itself a psychological equilibrium corresponding to the case in which the Trustor does not invest and acts to prevent abuse (real or

presumed) by the Trustee. Conformity to the principle is preserved by the precautionary decision of the Trustor, who does not legitimize the manager's authority.

Table 8. Example: λ=7

		TRUSTEE [B]	
		a	$\neg a$
TRUSTOR [A]	e_*	0 , 9	8.5 , 9.5
	e^*	0, 15	**13, 16**
	$\neg e$	**9, 9**	1 , 1

The Ultimatum game

The Ultimatum Game (UG) game concerns the distribution of the benefits of cooperation in the presence of asymmetric distribution of decision making power. Our interpretation of the ultimatum game has to do with the ongoing operation of the governance structure, in which case the division of the surplus can be more of less fair, depending on the offer of the first mover (the manager) to the respondent (the members of the co-operative). The first mover may be tempted by offering an unfair distribution of the surplus, in which members are penalized by excessive managerial appropriation. In the self-seeking case, the manager can nonetheless decide to offer an amount higher than the minimal (any positive epsilon) out of the fear that the responder rejects an unfair offer. This is obtained when the first mover is unsure whether the second mover is a purely self-seeking individual (in such a case he/she will accept any offer higher than zero), or is instead he/she attaches a positive weigh to the fairness of the partition of the surplus (in which case he/she will reject unfair offers). In the psychological Nash equilibrium, instead, both players agree ex-ante on the fairness norm T, and act accordingly out of the belief that the other player will do the same, and out of their propensity to conform to the norm.

In Table 9 there is risk of abuse by the decision maker (first mover), while the responder (member or stakeholder) can only decide whether to accept the offer of the first mover or not. In case the offer is rejected the respondent would lock any distribution out, damaging both himself/herself and the first mover. In our simplified version of the UG there can be several offers, some fairer and some less fair: high offer (fairest); median (quite fair); low (unfair); very low (exclusive appropriation by the first

mover, apart from a tiny amount epsilon). The “rational” self-seeking responder would accept all of them, but in the standard game the only “rational” offer of the first mover is the most unfair (10-ε; ε).

Table 9. Ultimatum Game. Discrete offers and pay-off matrix

		RESPONDER (MEMBER OR STAKEHOLDER)	
		a	$\neg a$
FIRST MOVER (MANAGER)	$v.l.$	10-ε; ε	0 ; 0
	l	9; 1	0 ; 0
	m	7; 3	0 ; 0
	h	5; 5	0 ; 0

We assume that the constitution of the firm provides for fair participation of any member to the cooperative surplus. Hence, each outcome of the unilateral distribution is assessed according to the value taken in it by the fairness norm T (corresponding to the Nash Bargaining Solution). The status quo required to calculate each player’s participation in the surplus is settled to be zero (the worst offer by the manager).

Table 10. The value of the ex-ante agreed principle T (NBS)

		RESPOSNDER (MEMBER OR STAKEHOLDER)	
		a	$\neg a$
FIRST MOVER (MANAGER)	$v.l.$	$T = (10-0) \times (0-0) = 0$	$T = (0-0) \times (0-0) = 0$
	l	$T = (9-0) \times (1-0) = 9$	$T = (0-0) \times (0-0) = 0$
	m	$T = (7-0) \times (3-0) = 21$	$T = (0-0) \times (0-0) = 0$
	h	$T = (5-0) \times (5-0) = 25$	$T = (0-0) \times (0-0) = 0$

Conformity of strategy combinations

If the first mover makes the fairest offer (5,5) the responder accepts, and the value of T value is maximal. In fact, had the first mover offered less, given the responder acceptance, he would have reduced the reduced the value of T. On the other hand, had the responder rejected, the value of T would have been nullified. This outcome corresponds to F = 1 and the ideal payoff is λ for both players. For any positive offer by the first mover, the responder strategy that maximizes T is always acceptance. However, any positive offer lower than the highest, given the responder acceptance, makes T less than maximal. In our payoff matrix F equal to 0.36 for offer (9,1); 0,84 for the offer (7,3). The ideal utility is, respectively, 0.36λ and 0.84λ.

When the first mover offer is (10-ε; ε), the responder rejection makes T as large as possible and this is true also for this offer of the first player (given rejection by the second player). Hence T is locally maximal, and F =1 and λ enter both players' payoffs. This would not be the case if the responder accepted the (10,0) offer, since the choice of the first mover then would minimize T (with T = 0) , with respect to the highest offer. In such a case, F = 0, so that no ideal utility enter the payoffs. These results are spelled out in Table 11 (we omit epsilon for sake of simplicity).

Table 11. Conformity indexes and psychological pay-offs

		RESPONDER (MEMBER OR STAKEHOLDER)	
		$\boldsymbol{a}$	$\neg\, \boldsymbol{a}$
FIRST MOVER (MANAGER)	$\boldsymbol{v.l.}$	10; 0	0 + (λ) ; 0 + (λ)
	$\boldsymbol{l}$	9 + [(0.36)(λ)]; 1 + [(0.36)(λ)]	0 ; 0
	$\boldsymbol{m}$	7 + [(0.84)(λ)]; 3 + [(0.84)(λ)]	0 ; 0
	$\boldsymbol{h}$	5 + (λ); 5 + (λ)	0 ; 0

Table 12. Example: λ=13

		Responder (member or stakeholder)	
		a	$\neg a$
First mover (manager)	$v.l.$	10; 0	**13 ; 13**
	l	13.68; 5.68	0 ; 0
	m	17.92; 13.92	0 ; 0
	h	**18; 18**	0 ; 0

Our model assumes that players refer to the bargaining agreement they would underwrite in an hypothetical cooperative bargaining session, thus allowing them to assess the level of conformity and the level of ideal utility associated to each outcome. Psychological equilibria coincide with perfect compliance strategies only when players show substantial psychosocial propensity to comply with a cooperative ex-ante agreement. Given that the responder will accept any offer, the maximum NBS value is for the offer (5.5), which means that the T function is at a maximum for the symmetric bid, T MAX = (5.5). In order to obtain that the fairest offer (5.5) is also the psychological equilibrium it is necessary to increase disproportionately lambda, setting it at its maximum possible value relative to the value of the material payoffs. In our example λ needs to be as high as 13 in order for the perfect compliance to be obtained. This outcome may be unrealistic. If the maximum value of λ is constrained equal to the maximum achievable payoff in the game, in our example 10, perfect compliance in our example cannot be achieved.

More realistically, conformity preferences should make room for "normal" values of the parameter λ, which are able to make sense of the current deviation in the experimental ultimatum game from the Nash extreme solution (10,0). Indeed, a psychological equilibrium different from (10,0) can be obtained for reasonably but not surprisingly high values of λ (significantly lower than 13). In fact, perfect conformity to the ex-ante agreement is not necessary on balance, because an average compliance is enough to remunerate psychologically both players in addition to the material benefit.

For example, a typical offer can be (7,3) (RECALL 7X3=21 is much closer to 25 - maximum value 5x5 - than to minimum (10,0) => 10x0 = 0). By this offer (7,3) the first mover does not deviate too much from the T principle (conformity is 0.84). Even if λ is lower than 13, the first mover best response to the responder acceptance is (7,3) while the responder continues to accept the agreement.

The typical level of concession found in the ultimatum games in experimental sessions is close to 30 per cent of the total available. Hence, conformist preferences, through mere application of the calculus, predicts standard experimental behaviours for a reasonable range of λ values. Reasonable

interval means that the values of λ must be at least fairly high to overcome the incentive not to propose the unfair outcome (9.1). However, they cannot be too high, for example not higher than the maximum material payoff available (10). For values of λ in this range the psychological component would enter the utility of both players motivating them to implement and accept such a level of distribution.

A second interesting result concerns the rejection strategy. The material outcome (0,0) is the second psychological equilibrium of the game. In this equilibrium the value of λ enters entirely in the utility function of the players, with psychological payoff equal to (λ, λ). The theory predicts that when the first mover does not grant anything to the responder (Nash strategy), or grants only a tiny ε, the responder rejects rationally making T as large as possible. Rejection by the responder implies that also the offer by the first mover becomes maximizing, which takes place if the responder accepted. Rejecting the nasty ultimatum is a «rational play», which intuitively support a fairer offer.

Conclusions and further research

Our model allows us to rebuild and reinterpret the existing experimental results according to the model of conformist preferences for specified values of λ. Beyond other theoretical development, further applied research could go in the direction of running an experiment with or without ex ante agreement, to see if results change. If results do not change, then the theory would not predict behaviour with explicit agreement. It is to be tested if it can be hypothesized that an explicit agreement can have a (limited or even marginal?) impact on the outcomes. An explicit agreement may increase endogenously the value of λ until the agreement overrides the material payoffs. This hypothesis appears as a reasonable explanation and deserves experimental testing.

Finally, applications of the model to real world contractual structures may include the employment relation and industrial relations interpreted as ultimatum and trust games.

APPENDIX: The formal model

A **psychological Nash equilibrium** for a n-person normal form psychological game G is a pair such that:

$$\left(\hat{b}, \hat{\sigma}\right) \in B \times \Sigma$$

i) $\hat{b}$ is a system of beliefs such that

$$\hat{b} = \beta\left(\hat{\sigma}\right)$$

ii) for all

$$\sigma_i \in \Sigma_i$$

$\hat{\sigma}$ is a strategy vector such that

$$V_i\left(\hat{b}_i, \left(\sigma_i, \hat{\sigma}_{-i}\right)\right) \leq V_i\left(\hat{b}_i, \hat{\sigma}\right) \tag{A.1}$$

- Condition (i) restrains the beliefs to be **coherent** with the equilibrium strategy (**beliefs predicting the strategy profile**)

Condition (ii) is a simple restatement of the **standard Nash equilibrium condition**: for each player the equilibrium strategy must confer a payoff not smaller than what attained by any other feasible strategy, *given* the opponents' strategies *and* beliefs.

T represents formally the ideal (the solution of the game under the "veil of ignorance"), and must be consistent with the idea of a cooperative agreement. The Nash Bargaining Solution (NBS), i.e. Nash social welfare function N is:

$$T\left(\sigma\right) = N\left(U_1, \ldots U_N\right) = \prod_{i=1}^{N}\left(U_i - c_i\right) \tag{1}$$

Drawing on formula (1) we can compute the values of the function at each state, and define a complete ordering of states of affairs according to their corresponding value of T. The next step in developing the formal model requires the introduction of an index of conditional conformity, which measures the extent to which (given the other agents' expected actions) the first player approximates the maximum value of T. The meaning of this index relates to the extent to which the first player is personally responsible for a fair distribution, given what (she expects that) the other player will do. Next an index of reciprocal conformity is introduced signaling the extent to which the other player is expected to maximize T, given what she (is expected to) expects from the first player's behavior. This

index specifies the (expected reciprocal) responsibility of the other player for the realization of a fair allocation of the surplus, given what she (is believed to) believes.

Formally, player i personal index of conditional deviation from full conformity to T (varying from 0 to -1 and representing the deviation from the ideal due to i's choice given his expectation):

$$f_i\left(\sigma_i, b_i^1\right) = \frac{T\left(\sigma_i, b_i^1\right) - T^{MAX}\left(b_i^1\right)}{T^{MAX}\left(b_i^1\right) - T^{MIN}\left(b_i^1\right)} \tag{A.2}$$

Where, b_i^1 represents the belief of player i over player j's action; $T^{MAX}\left(b_i^1\right)$ and $T^{MIN}\left(b_i^1\right)$ respectively represent the maximum and the minimum attainable by the function T given i's belief over j's strategy.

The estimation function of player j index of reciprocal deviation from full conformity to T varies from 0 to -1 and represents the expected deviation from the ideal outcome due to j's choice:

$$\tilde{f}_j\left(b_i^1, b_i^2\right) = \frac{T\left(b_i^1, b_i^2\right) - T^{MAX}\left(b_i^2\right)}{T^{MAX}\left(b_i^2\right) - T^{MIN}\left(b_i^2\right)} \tag{A.3}$$

where b_i^2 represents player i's second order belief over the belief of player j over the choice of player i.

An exogenous parameter λ is introduced, representing the motivational force of the agent's psychological disposition to act on the motive of reciprocal conformity with an agreed norm. The previous steps coalesce in defining an overall index F of conditional and expected reciprocal conformity for each player in each state of the game. This index operates as a weight (between 0 and 1) on the exogenous parameter λ deciding whether λ will actually affect (if so, to what extent) or not the player's payoffs.

All the spelled out conditions allow us to introduce the general form of the overall utility function:

$$V_i(\sigma) = U_i(\sigma) + \lambda_i F\left[T(\sigma)\right] \tag{2}$$

where *Ui* is the material utility for states ⍰ (described as consequences); λ_i is an exogenous psychological parameter (a disposition) that expresses how important the ideal component is within the motivational system of player i; T (NBS) is the fairness principles defined over states ⍰; F is the composite index of conformity to the principle T expressing both the agent's conditional conformity and other individuals' reciprocal conformity to T.

The overall utility function Vi is the linear combination of the two components (material and ideal)

$$V_i\left(\sigma_i, b_i^1, b_i^2\right) = U_i\left(\sigma_i, b_i^1\right) + \lambda_i\left[1 + f_i\left(\sigma_i, b_i^1\right)\right]\left[1 + \tilde{f}_j\left(b_i^2, b_i^1\right)\right] \tag{3}$$

This function spells out both the material component (*Ui*) and the psychological component. If i completely conforms and expects that j conforms too, then the value of her ideal utility is λ ($\lambda \times 1 \times 1 = \lambda$). If i does not conform and doesn't expect complete conformity on the part of j, we get [(1-x) (1-y) $\lambda < \lambda$]. Finally, when conformity is nil at least for one player, then [(1-1)(1-y) λ =0].