

Bernard Enjolras
Institute for Social Research
Munthesgt 31
N-0260 Oslo
Norway
E-mail : ben@samfunnsforskning.no

Collective action, self-interest and norms

Abstract

The conceptualization of social behavior based on norms constitutes a challenge for social sciences. The rational choice understanding of normative behavior consists in including the benefits of complying with a norm in the utility function of the individual. This formulation of normative behavior does not allow to account for behaviors for which complying with a norm dominates in any case the individual's interest.

This paper puts forward a model where both norms and self-interest enter into the proximate explanations of action and discusses how such a behavioral model affects both the nature and the outcomes of social dilemmas. I first propose a formulation of normative behavior based on dissonance theory. Secondly I introduce a model of individual behavior that allows individuals to be both self-interested and committed to a norm. I finally discuss the consequences of such a behavioral model on the nature and outcomes of social dilemmas using first the framework of a complete information two persons game and then that of an incomplete information two persons game.

Introduction

Social dilemmas may be defined as “situations in which individual rationality leads to collective irrationality” (Kollock, 1998: 183). Among the many possible solutions to social dilemmas the existence of a norm of cooperation is often considered (Hardin, 1982; Elster, 1989b; Von Borgstede, 2002). However the conceptualization of social behaviour based on norms constitutes a challenge for social sciences. As claimed by Wrong (1961: 183), social science “originates in the asking of general questions about man and society”. To the Hobbesian question of how men become tractable to social control, social science answers either with an over-socialized conception of man (Wrong, 1961) where the individual's behaviour is the result of internalized norms and expectations of the others, or by an under-socialized conception of man for which the individual's behaviour is the result of self interested choice.

Social norms are usually characterized in the sociological and psychological literature as required behaviour without calculated purpose or concern for the consequences (Sherif, 1966; Scott, 1971, Opp, 1979; Elster, 1989a, 1989b). Individual behaviour in conformity with social norms is often contrasted with behaviour based on instrumental rationality. Indeed it is common in sociology to distinguish between rational action and normative action. As stated by J. Elster (1989b:99):

Rational action is concerned with outcomes. Rationality says: if you want to achieve Y, do X. By contrast I define social norms by the fact that they are not outcome-oriented. The simplest social norms are of the type: Do X, or don't do X. More complex norms say: if you do Y then do X, or if others do Y then do X. More complex norms still say: Do X if it would be good if everyone did X. Rationality is essentially conditional and future oriented. Social norms are either unconditional or, if conditional, are not future oriented. For norms to be social, they must be shared by other people and partly sustained by their approval and disapproval.

Compliance with norms is considered to be the result of two interrelated mechanisms: (i) approval of compliance and disapproval of non-compliance by the persons (others) that share the norm; (ii) internalization to the extent to which individuals through a socialization process develop an intrinsic system of sanctions that sustains and reinforces the norm (Scott, 1971).

It is possible to identify two main accounts of normative behaviour: *normative behaviour as internalized behaviour* and *normative behaviour as rational choice*.

The thesis of the normative regulation of behaviour can be stated as follows (Opp, 1979: 776): “if norms are institutionalized and internalized, then the behaviour which is in accordance with the norms occurs”. This understanding of normative behaviour is subject to three types of objection: (i) norms do not provoke behaviour automatically; (ii) there may be other determinants (beside the norm) of behaviour than the norm; (iii) the behaviour may be explained by other variables than the norm.

The rational choice understanding of normative behaviour consists of including the benefits of complying with the norm in the utility function of the individual. Since for rational choice theorists it is rational to conform to norms, compliance is reduced to utility maximization. It is indeed possible to introduce, as an argument of the utility function, the net benefit of conforming to the norm.

$$U = U(N)$$

where N , the net benefit derived from conforming to the norm, is the benefit of conforming, minus the expected costs of not conforming to the norm.

This formulation of normative behaviour does not allow to account for behaviors for which complying to the norm dominate in any case the individual's interest as for example in the case of duels where the code of honor is more important than the survival (Elster, 1990). Another objection to this understanding was formulated by Elster) (1999: 302-303):

We may consider the case of guilt. Assume that a person is tempted to steal a book from the library. If he feels guilty about doing it, he may abstain. If he steals the book and then feel guilty he may return the book to the library. On the assumption that guilt is to be modeled as a cost, both the abstention from stealing and the return of the book would be explained by a simple cost-benefit analysis. This approach has the great advantage that it allows us to account for the undeniable existence of a trade-off between moral emotions and self interest. (...)

Independently of its predictive adequacy, I submit that this model is conceptually flawed. If guilt was nothing but an anticipated or experienced cost, an agent whose guilt deters him from stealing or retaining the book should be willing to buy a guilt-erasing pill if it was sufficiently cheap. I submit that no person who is capable of being deterred by guilt would buy the pill. In fact he would feel guilty about buying it. (...)

We need therefore a model that can account for the trade-off between norm and interest and yet does not imply that a reluctant agent would buy the guilt-erasing pill. In other words we need a model that allows the individual to be at the same time: (i) committed or not to a norm; and (ii) self-interested.

This paper puts forward such a model (where both norms and self interest enter into the proximate explanations of action) and discusses how such a behavioral model affects both the nature and the outcomes of social dilemmas.

I first propose a formulation of normative behaviour based on dissonance theory. Secondly I introduce a model of individual behaviour that allows individuals to be both self-interested and committed to a norm. Finally I discuss the consequences of such a behavioral model on the nature and outcomes of social dilemmas using first the framework of a complete

information two persons game, and secondly that of an incomplete information two-persons game.

Dissonance and normative action

The main difficult question concerning social norms is to determine how these emerge and are sustained. In terms of game theory norms are considered as equilibria of strategic interaction (Lewis, 1969; Ullmann-Margalit, 1977). Other explanations for the emergence and sustenance of social norms are the existence of a “metanorm” (Axelrod, 1986) that leads people to punish the defectors and the “instrumentality proposition” (Hechter & Opp, 2001) that considers that norms emerge if their emergence is in the interest of a collective of people. However these explanations are not fully satisfying. As pointed out by Elster (1989a: 102) “convention equilibria are guided by outcome in a substantive sense, not just in the formal sense that people want to avoid disapproval”. The existence of a metanorm may be a necessary condition for explaining the sustenance of a norm but does explain why a norm emerges in the first place. The instrumentalist approach presents two problems (Feld, 2002): (i) that of determining which interests and benefits determine the norm; and (ii) that of determining the causal relation between given interests and a given norm (there may be several possible norms serving group interests equally well).

Instrumental approaches to social institutions have three shortcomings:

- (i) Actors are not only interested in material benefits: symbolic and psycho-social factors are important motivation for action. Legitimacy and power constitute symbolic reasons for action; esteem and differentiation constitute psychosocial reasons for action. The mechanisms underlying these motives for action are not necessarily those of cost-benefit analysis.
- (ii) Actors are not necessarily self-interested; they can behave following altruistic motives or normative motives (commitment).
- (iii) The results and intentions of interactions, the institutions, are not necessarily Pareto-improving. Keeping the status-quo over the asymmetrical distribution of resources (material and symbolic) as well as the differentiation of status may be both the intention and the outcome of institutions.

This paper will not attempt to give a theoretical understanding of norms emergence. Here, I am interested in conceptualizing how individuals may act in a context where norms are given. However, the way one conceptualizes the individual’s action in a normative context is not independent of the conception one has of the emergence and social role of norms. I will therefore make a brief presentation of my own understanding of the emergence of norms. Following Jon Elster I will consider that norms present three characteristics (Elster, 1989a: 104; 1989b: 99):

- (i) When people obey norms, they want to avoid disapproval – ranging from raised eyebrows to social ostracism – of other people.
- (ii) Norms do not need external sanctions to be effective. When norms are internalized, they are followed even when violation would be unobserved and not exposed to sanctions.
- (iii) They are also sustained by the feelings of embarrassment, anxiety, guilt and shame that a person suffers at the prospect of violating them. A person obeying a norm may also be propelled by positive emotions, like anger and indignation.

Given these characteristics I submit that the role of social norms is that of regulating social relations and interactions between individuals, and satisfy psychological as well as social needs. Norms emerge and function because they satisfy at the same time psychological and social needs. On the social level norms regulate social interaction and ensure predictability of behaviour when individuals interact (each part of a social interaction knows how to behave according to the situation and the relevant norm, and knows how others will behave), and at the same time constitute a mean of expression for social identity or social status. Predictability also fulfills a psychological function that of ontological security. In addition on the psychological level they satisfy the needs of self-esteem and self-security. Norms coordinate social expectation not through cognition but through emotion. The emotional mechanisms involved in sustaining social norms are dissonance, self esteem and self security.

A possible objection to this understanding of norms (Opp, 1979: 781) is that if norms emerge to reduce insecurity then any arbitrary norm could emerge. It is therefore necessary to explain why some types of norms emerge. Norms exist because they make social interactions more predictable and therefore easier by enhancing self-security in social interaction. But at the same time they are a means for expressing group belonging and social status or identity (code of honor, etiquette...). Self-esteem (as psychological mechanism) and social status (as social mechanism) are related: social status is a component of self-esteem. My esteem of myself (self-esteem) is not independent of others' judgement or assessment; my identity is basically social: I need confirmation by others of who I am. In socialization I internalize social norms that become part of my identity. Norms emerge therefore as a result of a psychological and social process, and to this extent are group (or culture) specific or context dependent. They are a solution to real or potential conflict endangering social interactions in a group and they are means for distributing, expressing and maintaining social status and identities within a collectivity. To this extent they also reflect the social structure and stratification of a group or collectivity. They are, at the same time, determined by the prevailing social structure and the existing social relations and used in order to structure these social relations. They do not serve necessarily collective or particular interests (although they might) but contribute to maintain and regulate social interactions within a group.

If one considers that norms emerge in order to *regulate* social interaction, are based on psycho-social needs, and that a social norm defines what is (socially) legitimate to be done in a given social interaction, the existence of a norm may be explained by the outcome intended to be obtained. But the instrumental reason or intended outcome may be either to improve the efficiency of the interaction or to maintain the asymmetric distribution of symbolic and material resources between actors.

In other words, the regulation of behaviour in social interactions as well as the realization of the intended social outcome of a norm (efficiency or status-quo maintenance) is made possible and sustained because norms satisfy psycho-social needs. However, the outcomes of norms (according to the content of the normative rule) may be Pareto-improving i.e. improving the efficiency of the interaction, or status-quo maintaining, i.e. contributing to maintaining the asymmetric distribution of symbolic and material resources.

A possible typology of norms will be based on the type of interaction and will distinguish between norms of:

- Coordination: that defines rules for coordination and cooperation;
- Exchange: that defines rules for exchange both symbolic and material;
- Distribution: that defines rules for distribution of symbolic and material resources and rules of justice;

- Differentiation: that contributes to maintain the differentiation in terms of symbolic and material resources (status, privileges...).

When I do not conform to those norms a process of dissonance generates feelings of shame and guilt. There is dissonance when a distance is created between what I should do (obey a given norm) given my social status (and my conception of myself), and what I do, i.e. what I am communicating about myself to others (not obeying this norm). The distance between my subjective perception of how *I* am perceived by others (social status or social identity) and my own claim about the person I am (self-esteem) creates dissonance.

As put by Elster (1999: 304):

An individual who is subject to several motivations that point in different directions will feel an unpleasant feeling of tension. When on balance he favors one action, he will try to reduce the tension by looking for cognitions that support it; when he favors another, he will look for cognitions that stack the balance of argument in favor of that action. Thus the timing of the switch in behaviour will be path-dependent. Dissonance theory is more realistic than the cost benefit model in that it views individuals as making hard choices on the basis of reasons rather than on the basis of introspections about how they feel.

In the next section I will propose a model that conceptualizes agents' actions as determined at the same time by interest and norms. Individuals are seen as (i) maximizing the utility derived from a given action, and (ii) maximizing the consonance between – on the one hand their social identity (i.e. how they are perceived by others) and – on the other, the norms that are valid given the social context of the action.

According to Leon Festinger (1957) the existence of dissonance, being psychologically uncomfortable, motivates the person to try to reduce the dissonance and achieve consonance. Dissonance is defined (Festinger, 1957:3) as:

the existence of non-fitting relations among cognition (...) by cognition I mean any knowledge, opinion or belief about the environment, about oneself or about one's behaviour. Cognitive dissonance can be seen as an antecedent condition which leads to activity oriented toward dissonance reduction just as hunger leads to activity oriented toward hunger reduction.

Formally, X and Y are dissonant if not-X follows from Y. The magnitude of dissonance (Festinger, 1957:16) is “a function of the importance of the elements. The more these elements are important to or valued by a person, the greater will be the magnitude of a dissonant relation between them”.

Dissonance with respect to a norm will occur when an agent's action, given his conception of himself (self-esteem) does not fit with the given the social expectations (norms) that are attached to the agent's self-conception (social identity). In other words normative dissonance will occur when the agent feels that what he is doing does not fit with what he is expected to do given the social context and given the person he wants to be socially.

The level of consonance of an action with respect to a norm may be represented by a consonance function. Psychologists agree on the importance of internalization of rules or norms for behaviour. Sociologists stress the role of group pressure in the process of norm conformity. When a person has internalised a norm and when the norm is enacted by the group the person is a member; the prospect of an action which is not in conformity with the norm will create dissonance between what the person should do and what he or she is planning to do. This dissonance will produce feelings of anxiety, guilt and shame. The

individual is assumed to reduce this dissonance by choosing the action which is the most consonant with the norm.¹

We consider an individual faced with possible actions in some set X , his action set. The individual is assumed to have preferences on the actions in the action set X in terms of the level of consonance of each possible action with respect to a given norm. Actions that are more consonant with respect to a norm are preferred to actions that are less consonant with respect to this norm. It is assumed that the preferences in terms of the level of consonance with respect to a norm satisfy the standard properties of the consumer theory: completeness, reflexivity, transitivity, continuity and strong monotonicity. Given these properties a numerical representation for those preferences in terms of the level of consonance of an action with respect to a norm exists and is given by any function C with domain X and range the real line such that $x \succ y$ if and only if $C(x) > C(y)$.

So far I have considered that norms are given and internalized by actors. As indicated by Coleman (1990) and Koford and Miller (1991), two types of norms may be distinguished: those enforced by external sanctions and those based on internal sanction. The model presented here considers only the second type .

One way to do it consists of defining the normative consonance of an action i by individual j , with a norm n is the following:

$$c_{ijn} = a_{ijn} - \frac{1}{k} \sum_{-j=1}^k a_{i-jn}$$

The consonance of an action i , by individual j , with a norm n is the difference between the “subjective approval” (Holländer, 1990) of the action by individual j , minus the average approval of this action by others, $-j$.

In this formulation, in addition to the internalization of the norm, the individual is assumed to be sensitive to social pressure. The dissonance is then the difference between (i) the subjective valuation by the individual of a norm associated with a given action and (ii) the average subjective valuation of this norm associated with this action by the group of reference of the individual. However, social pressure does not need to be the result of a deliberate action of individuals constituting a group. Social pressure may be as well internalized as the “Generalized Other” (Mead, 1967). In this case the consonance with the norm may be represented by $c_{in} = a_{in} - \bar{a}_n$, where \bar{a}_n represents the internalized “Generalized Other”. The consonance of an action i , with a norm n is then the difference between the “subjective approval” of the action by the individual, minus the internalized approval of the “Generalized Other”. Using the terminology of Mead (1967), a_{in} will represent the “I”, that is the subjective part of the self, whereas \bar{a}_n will represent the “Me”, that is the internalized attitude of the others. If $a_{in} = \bar{a}_n$, the action will be consonant with the norm. If $a_{in} < \bar{a}_n$, negative dissonance will characterize the action, i.e. the action is anticipated to provoke social disapproval because of the deviation from the norm. If $a_{in} > \bar{a}_n$ positive dissonance characterizes the action, i.e. the action is anticipated to provoke social disapproval because of excessive conformity with the norm.

¹ We will see that choosing another action is not the only way of reducing dissonance.

A model of self-interested and normative choice

Formally any action A_i will be undertaken if $A_i > 0$

$$A_i = p_i S(s_i) + (1 - p_i) C(c_{in})$$

where:

s_i = satisfaction for action A_i .

c_{in} = consonance of action A_i

p_i = subjective weight or subjective probability of the impact of the action A_i on satisfaction and consonance.

The utility function S and the consonance function C express different dimensions or attributes of the choice of an action. It is possible to use one dimension (for example utility) to define a quantitative unit in order to express “equal-utility intervals” along the consonance dimension (Baron, 2000: 341). In order to guarantee that the two dimensions can be summed, it is here assumed that utility and dissonance constitute two “psychologically independent attributes”, i.e. attributes that the individual sees as independent (Baron, 2000: 343). In other words, the level of consonance of a given with respect to a norm action does not affect the level of utility of this action and reciprocally.

How shall the subjective weights be understood? The subjective weights are first personal judgment of the likelihood of an event as is the case in expected-utility theory (Savage, 1954). Second, the subjective weights may be understood as lexical rules (Baron, 1994: 346). Lexical rules give a list of the order in which different issues are considered. Lexical rules reflect high *subjective weight* for one dimension of a trade-off as opposed to another (Baron, 1994: 347).

The subjective weights may change over time as the individual tries to reduce dissonance by reducing the value of the dissonance in their valuation of an action. Preference change (Elster, 1983) may be explained as a change in the values of the subjective weights since it may reduce the value of the dissonance. People need reasons in order to do or not to do something: when the dissonance is too high, additional reasons or new reasons may affect the subjective weights and therefore the utility of the action.

To what extent does the consonance function differ from other formulations of the effect of norms on behaviour in terms of loss of utility such as those proposed by Akerlof (1980), Lindbeck et al. (1996), and Kübler (2001) among others? In other words to what extent does the formulation proposed here, $C(c_{in})$, differ from the formulation which is usually found in the literature: $U(n)$?

First, note that it is possible to include the loss of utility of not conforming to the norm within the utility function. One may behave strategically in relation to norms by evaluating the loss of utility of not conforming to a norm. Second, dissonance is not an anticipated or experienced cost. It is not compensated with benefits even if the net benefits of a given action are high, its dissonance with a given norm will not be reduced. Consider once again the example cited by Elster (1999:302-303) of a person tempted to steal a book from the library. A cost-benefit analysis may lead this person to steal the book if, for example, the risk of being caught is very small. However, most individuals will not steal the book, even when the costs are small. The normative dissonance that deters most of us from stealing a book is not a result of a rational calculation but that of a psychological mechanism that commands us not to steal.

Conversely, in the case of the norm of revenge (Elster, 1990), in spite of the high cost (the person may lose his life) the action is carried out because the dissonance of not acting is too high. One may argue that dissonance is nothing else but a psychological cost and consonance psychological benefits. If this was the case, different types of costs and rewards (monetary and psychological) should compensate each other. The difference between costs and dissonance may be exemplified by the following story: I am on my way to an important appointment (for example a job interview) when a person in front of falls and breaks a leg. My choice is between (i) helping this person and coming late to my appointment, and (ii) not helping this person, ignoring her, and arriving for my appointment on time. A simple cost-benefit analysis (including the psychological costs, i.e. the bad feeling of not helping somebody in need) will lead me to choose the second option. The costs of helping will be high (may be several years actualized wage) and the benefits low (some gratitude and a good feeling of myself). However I will probably help this person, not because the psychological benefits of helping and the costs of not helping will be high, but because I will feel that I have to do so and because not helping this person does not concur with my conception of being a human being. The difference between cost and dissonance may be thought in terms of what A. Sen (1977: 329) calls commitment, to the extent that dissonance as commitment “drives a wedge between personal choice and personal welfare”.

Thirdly, the two formulations may lead to the same result (mathematically) if one assumes that complying with a norm is source of utility. But it is necessary to precise what does mean the utility function in order to differentiate the two formulations. The utility function may be defined as a formal rule of behaviour based on the valuation of the level of utility of an action. This rule of behaviour reflects a psychological process by which the actor values the benefits (satisfaction) and costs (dissatisfaction) of an action in order to make the decision to carry out a given action. In contrast, the consonance function may be defined as a formal rule of behaviour based on the valuation of the level of consonance of an action with a *prescriptive* norm. It reflects a different type of psychological process by which the actor values the conformity of an action with a prescriptive norm. In addition the utility function and the consonance function differ regarding the underlying hypothesis concerning the behaviour of the actor. Whereas the utility function presupposes a *rational and self-interested* actor, the consonance function presupposes a *moral* actor. The two formulations $C(c_{in})$ and $U(n)$ are equivalent if the moral satisfaction of complying with a norm (which is different from the gain the actor can realize from strategically complying with the norm) is considered as a source of utility. But this idea contradicts the hypothesis of a rational self-interested actor. A rational self-interested actor will not get any satisfaction from morality *in itself* unless he acquires a gain which does not come from morality *in itself* but from his interest, i.e. from the advantages inherent to his action. In other words, complying with a norm for the sake of the norm will not procure any satisfaction for a rational self-interested actor, whereas a moral actor will value the norm for its own sake. To comply with a norm because I can gain advantage from it or because I feel an injunction to do so are not the same thing, even if both in a general sense (that reminds the pleasure principle of psychoanalysis) may be said to procure satisfaction. In addition the motivation that at the outset of the action is not of the same nature.

Differentiating moral action from self-interested action appears as a necessity for the social sciences in order to account for social phenomena. The rational choice model has failed (i) to take into account the fact that social actors are moved both by interest and morality; (ii) to distinguish within the utility framework the difference of nature between satisfaction procured by the realization of interest and that which is inherent to the realization of morality, and (iii)

to take into account the trade-off between interest and morality that characterizes social action.

Returning to our model, if one consider the discrete case where the satisfaction provided by an action s_i and the consonance generated by this action c_m takes two values High and Low, the possible outcomes are given by the following matrix:

	Consonance High	Consonance Low
Satisfaction High	HH	LH
Satisfaction Low	HL	LL

In the continuous case two main outcomes have to be distinguished:

Case (i) positive sum action

$$p_i s_i \geq 0 \text{ and } (1-p_i)c_m \geq 0$$

The norm reinforces or increases the satisfaction: this case can explain some paradoxes such as vote participation; even if the satisfaction (interest) is low the consonance with the norm increases the total utility.

Some norms such as reciprocity enhance both self-esteem and self-interest.

Case (ii) negative sum action

$$p_i s_i \geq 0 \text{ and } (1-p_i)c_m \leq 0 \quad (1)$$

or

$$p_i s_i \leq 0 \text{ and } (1-p_i)c_m \geq 0 \quad (2)$$

The outcome of the action depends on the values of the pay-off s_i and c_m and on the value of the subjective weights p_i .

In the first situation, if $|(1-p_i)c_m| > p_i s_i$ then the norm will determine the action, the value of the action being negative no action will be taken. Conversely, if $|(1-p_i)c_m| < p_i s_i$ interest will determine the action, the value of the action being positive the action will be taken.

In the second situation, if $|p_i s_i| > (1-p_i)c_m$ then interest will determine the action and the action will not be undertaken. Conversely, if $|p_i s_i| < (1-p_i)c_m$, the norm will determine the action undertaken.

A third situation is given by $p_i s_i \leq 0$ and $(1-p_i)c_m \leq 0$, where no action will be taken.

Collective action and cooperation norm

In his simplest statement of the logic of collective action Olson (1965) presents the equation of costs (C), gross benefits (V_i) to the individual i and net benefits (B_i) to the individual from i 's own contribution to a group's collective good: $B_i = V_i - C$. In the absence of selective incentive individuals will contribute to the collective good if $B_i > 0$. In Olson's formulation collective action will take place if the net benefit for each individual is positive.

But the problem of collective action presents another feature, and may be restated as a prisoner dilemma (Hardin, 1992), for even when individuals may achieve a positive net benefit from collective action, no matter what other individuals do, a given individual is better off not contributing. Not contributing (free-riding) is the best strategy irrespective of whether others contribute or not.

However, a reasonable reading of the experimental economic literature on voluntary contribution mechanisms and social dilemmas leads one to conclude that the major findings to date are:

1. In one-shot trials and initial stages of finitely repeated trials, subjects generally provide contributions halfway between Pareto-efficient level and free-riding,
2. Contributions decline with repetition, and
3. Face to face communication improves the rate of contribution. (Leydard, 1995: 121)

Introducing a normative motivation as determinant of the decision to contribute to the collective good modifies the formulation of the problem and may explain why experiments conclude in a presence of a relatively high level of contribution and why face to face communication increases the rate of contribution. The use of norms to mitigate free riding in the provision of public goods is discussed by Sen (1977 : 332) in these terms: "What is at issue is whether people (...) always give a gains-maximizing answer (...). The presence of non-gains-maximizing answers (...) immediately brings in commitment as part of the behaviour." Von Borgstede (2002) identifies two types of norms which can be relevant in presence of social dilemmas: norm of commitment and norm of equity. The commitment norm implies that we expect others to act in the way they have committed themselves to act, whereas the equity norm prescribes that individuals should be rewarded in proportion to their contribution. The formation of social norms for voluntary contribution of a public good has been analysed within the gains-maximizing framework by Holländer (1990) and Rege (2001), showing that such a norm can emerge from agents' strategic interactions.

In this case the decision to contribute depends on the net benefits of the contribution and on the degree of consonance with a social norm that prescribes not to free ride:

$$A_i = p_i S(B_i) + (1 - p_i) C(c_{in})$$

Using Olson's formulation, in the absence of selective incentive, the existence of a norm by generating dissonance may motivate the individual to contribute to the collective action, in spite of the net benefit of contribution being negative.

The existence of a social norm may also contribute to overcome the collective action dilemma as formulated by Hardin. The “individual vs. collective game” considered by Hardin (1982: 26) may be reformulated by introducing the consonance effect of a cooperation norm.

A complete information game

Consider the following game where two players decide to the public good, and contributing is a 0-1 decision. Each player derives a benefit if at least one of them provides the public good and no benefit if neither of them does. In the following matrix the row entries are the payoffs for individual i , and the column entries are the payoffs for individual j . The payoffs are given by the decision function: $A_i = p_i S(s_j) + (1 - p_i) C(c_{in})$.

In order to simplify I consider that $p_i S(s_j) = 1 - t_i$ for both individuals where t_i is the cost of contributing. I neglect $(1 - p_i)$ since in this formulation it reduces to a constant. The dissonance of individual's i action in relation to the norm of cooperation is given by a_i , with $a_i = 0$ if individual i subjectively values the norm of cooperation and does not contribute, and $a_i = 1$ if individual i subjectively values the norm of cooperation and does contribute. Conversely, $a_i = 1$ if individual i does not subjectively value the norm of cooperation and does not contribute, and $a_i = 0$ if individual i does not subjectively value the norm of cooperation and does contribute. In the first place I consider the case where the type of the individuals (conforming to the norm or not conforming) is common knowledge.

<i>Individual: i</i>	<i>j</i>	
	Contributes	Does not contribute
Contributes	$1 - t_i + a_i, 1 - t_j + a_j$	$1 - t_i + a_i, 1 + a_j$
Does not contribute	$1 + a_i, 1 - t_j + a_j$	0,0

When the action is consonant with the norm of cooperation the value of the consonance is added to the net benefit of the action whereas it is subtracted when the action is dissonant with the norm.

Compared to the Prisoner's dilemma game, where the dominant strategy is not contributing for both players, this formulation offers a richer range of outcomes to social dilemmas. The interactions yield four possible outcomes:

- (i) i and j conform to the cooperation norm.

The matrix is that of an assurance game, but unlike the assurance game (where defection can be a dominant strategy if one player believes that the other player will defect) there is only one equilibrium – that of cooperation, since both players are committed to cooperate.

<i>Individual: i</i>	<i>j</i>	
	Contributes	Does not contribute
Contributes	$1-t_i+1, 1-t_j+1$	$1-t_i+1, 1$
Does not contribute	$1, 1-t_j+a_j$	0,0

(ii) *i* and *j* do not conform to the cooperation norm. The matrix is that of a prisoner dilemma game.

<i>Individual: i</i>	<i>j</i>	
	Contributes	Does not contribute
Contributes	$1-t_i, 1-t_j$	$1-t_i, 2$
Does not contribute	$2, 1-t_j$	0,0

(iii) *i* does not conform to the cooperation norm and *j* does.

<i>Individual: i</i>	<i>j</i>	
	Contributes	Does not contribute
Contributes	$1-t_i, 1-t_j+1$	$1-t_i, 1$
Does not contribute	$2, 1-t_j+1$	0,0

In this case, *i* does not contribute whereas *j* does contribute.

(iv) *i* conforms to the cooperation norm and *j* does not.

In this case the matrix is the reverse of the preceding one, and *i* contributes whereas *j* does not.

Taking into account the effect of normative motivated behaviour contributes to increase the complexity of the situations characterized by social dilemmas. The game presented above shows that the nature and outcome of a social dilemma depends on the type of players interacting (i.e. whether they are sensitive or not to the dissonance of not conforming to a cooperation norm). The outcomes of social dilemmas are determined by the proportion of players who conform to the norm. If everybody follows the norm the obvious outcome is cooperation; if nobody follows the norm the situation is that of a prisoner dilemma where everybody free rides. If one of the players follows the norm the public good is provided.

An incomplete information game

What is the effect of introducing a cooperation norm on social dilemmas when information is incomplete, i.e. when some players do not know the pay-offs of the other and when the possibility that the players may not be self-interested, (i.e. are conforming to a norm) is considered?

Consider the following public good game:

<i>Individual: i</i>	<i>j</i>	
	Contributes	Does not contribute
Contributes	$1 - t_i + a_i(\theta_i),$ $1 - t_j + a_j(\theta_j)$	$1 - t_i + a_i(\theta_i),$ $1 - t_j - a_j(\theta_j)$
Does not contribute	$1 - t_i - a_i(\theta_i),$ $1 - t_j + a_j(\theta_j)$	0,0

Where $a_i(\theta_i) = 1$ if θ_i , i.e. i 's type is to conform to the cooperation norm, and $a_i(\theta_i) = 0$ if $\bar{\theta}_i$, i.e. i 's type is to not conform to the cooperation norm. t_i and t_j , the costs of contributing, are supposed to be the same for individuals i and j : $t_i = t_j$ and are common knowledge.

If p_i is the equilibrium probability that player i contributes: $p_i = \Pr(s_i^*(a_i(\theta_i))) = 1$.

Player j 's optimal behaviour is :

$$s_j^* = 1 \text{ if } 1 - t_j + a_j(\theta_j) > 1 - a_j(\theta_j)(1 - p_i)$$

$$s_j^* = 0 \text{ if } 1 - t_j + a_j(\theta_j) < 1 - a_j(\theta_j)(1 - p_i)$$

Player j will contribute if his pay-off is more than his benefit from the public good, times the probability that player i does not contribute.

Conversely, for player i

$$s_i^* = 1 \text{ if } 1 - t_i + a_i(\theta_i) > 1 - a_i(\theta_i)(1 - p_j)$$

$$s_i^* = 0 \text{ if } 1 - t_i + a_i(\theta_i) < 1 - a_i(\theta_i)(1 - p_j)$$

If $a_i(\theta_i) = 0$:

$$s_i^* = 1 \text{ if } 1 - t_i > 1 \cdot (1 - p_j)$$

$$s_i^* = 0 \text{ if } 1 - t_i < 1 \cdot (1 - p_j)$$

If $a_i(\theta_i) = 1$:

$$s_i^* = 1 \text{ if } 2 - t_i > 0$$

$$s_i^* = 0 \text{ if } 2 - t_i < 0$$

In other words, a player whose type is of conforming to the norm will contribute whatever the probability that the other player does not contribute, as long as there is a positive pay-off for contributing (i.e. the cost does not exceed the benefit). But the interesting feature of this game is that a player whose type is of not conforming to the norm *will* contribute if his subjective probability that the other player will not contribute (i.e. the other player's type is of not conforming to the norm) is high enough, and his cost is low enough.

Self-interested individuals who take into account the fact that some individuals are committed to contribute whereas others will free ride, may end up contributing even if they are not themselves committed to contribute. The prisoner dilemma structure of the public good game is due to the fact that it is common knowledge that both players are self-interested. Considering the possibility that the players may be committed (i.e. have a norm-based behaviour) and that the players ignore the type of the other player, changes the structure of the game.

The most important conclusion of this analysis is that strategic interactions in a “normative” social setting, i.e. where individuals take into account the fact that other individuals may be committed to a norm, do not necessarily lead to the same outcomes as interactions where individuals anticipate that other individuals will act according to their self-interest.

The inclusion of a norm as determinant of individuals’ behaviour has two consequences:

- (i) Individuals who are committed to a norm may act in contradiction to their self-interest, i.e. even if the costs of an action exceed the benefits;
- (ii) Self-interested individuals (who do not value a norm) may act in a different way when they take into account the fact that other individuals are not necessarily self-interested as they may do when facing self-interested individuals. The reason is not that they include as a determinant of their decision the “costs” generated by social disapproval (here social disapproval is a cost, since strictly self-interested individuals do not experience dissonance by not conforming to a norm), but that they include in their strategic calculation the possibility that other individuals may not be self-interested.

In order to assess the respective effects of social disapproval and of being confronted with possibly not self-interested individuals on the outcomes of a public good game, consider the following game:

<i>Individual: i</i>	<i>j</i>	
	Contributes	Does not contribute
Contributes	$1 - t_i + a_i(\theta_i + \theta_j),$ $1 - t_j + a_j(\theta_i + \theta_j)$	$1 - t_i + a_i(\theta_i + \theta_j),$ $1 - t_j - a_j(\theta_i - \theta_j)$
Does not contribute	$1 - t_i - a_i(\theta_i - \theta_j),$ $1 - t_j + a_j(\theta_i + \theta_j)$	0,0

Where $a_i(\theta_i) = 1$ if $\theta_i = 1$, i.e. i 's type is to conform to the cooperation norm and $a_i(\theta_i) = 0$ if $\theta_i = 0$, i.e. i 's type is to not conform to the cooperation norm. The pay-offs $a(\theta_i - \theta_j)$ and $a(\theta_i + \theta_j)$ are then dependent on the type of individual and on the social approval (i.e. approval by the other player according to his type) of his action: defection or contribution.

As was the case in the preceding formulation of the game, committed individuals ($\theta = 1$), will contribute in any case. The optimal strategy for strictly self-interested individuals ($\theta = 0$) will be then given by:

$$s_i^* = 1 \text{ if } 1 - t_i + a_i(\theta_i + p_j \theta_j) > 1 - a_i(\theta_i - p_j \theta_j)(1 - p_j)$$

$$s_i^* = 0 \text{ if } 1 - t_i + a_i(\theta_i + p \theta) < 1 - a_i(\theta_i - p \theta)(1 - p)$$

where p_j is the equilibrium probability that player j is of type $\theta_j = 1$.

It is easy to see that the additional costs generated by social disapproval as well as the additional benefits of social approval have no effect on the decision to contribute. A strictly self-interested individual will contribute if the probability that the other player is not committed to the cooperation norm is high enough and if the costs for contributing are low enough.

Summary and conclusion

The behavioural model proposed in this paper has permitted the demonstration of a well known fact: that the prevalence of a cooperative norm within a group may explain that – in spite of the presence of a social dilemma that would arise if all individuals were self interested – cooperation does occur. The dissonance created by not conforming to the norm of cooperation may lead individuals to contribute to the collective good.

Taking into account the possibility that individuals may be committed to a norm leads to a reformulation of social dilemmas. First, non-cooperation constitutes one possible outcome of interactions where cooperation is beneficial for the collectivity. Second, non-cooperation occurs as the result of interactions between strictly self-interested individuals. Third, even if individuals are self-interested, non-cooperation does not necessarily occur when the possibility that individuals may be committed to a norm is taken into account, and when individuals ignore whether they interact with a self-interested or a norm committed individual.

References

- Akerlof, G., 1980, "A theory of social custom, of which unemployment may be one consequence", *Quarterly Journal of Economics*, 94, 49-75.
- Axelrod, R., 1986, "An evolutionary approach to norms", *American Political Science Review*, 80:4, 1095-1111.
- Baron, J., 1994, *Thinking and deciding* (second edition), Cambridge: Cambridge University Press.
- Baron, J., 2000, *Thinking and deciding* (third edition), Cambridge: Cambridge University Press.
- Coleman, J.S., 1990, *Foundations of social theory*, Cambridge MA: The Belknap Press of Harvard University Press.
- Elster, J., 1983, *Sour Grapes*, Cambridge: Cambridge University Press.
- Elster, J. 1989a , "Social norms and economic theory", *Journal of Economic perspective*, 3:4, 99-117.
- Elster, J., 1989b, *The cement of society*, Cambridge: Cambridge University Press.
- Elster, J., 1990, "Norms of revenge", *Ethics*, 100:4, 862-885.
- Elster, J., 1999, *Alchimies of the mind*, Cambridge : Cambridge University Press.
- Feld, S.L., 2002, "On the emergence of social norms, review essay", *Contemporary sociology*, 31:6, 638-640.
- Fehr, E., & Gächter, S., 1998, "Reciprocity and economics: the economic implications of homo reciprocans", *European Economic Review*, 42 : 845-859.
- Festinger, L., 1957, *A theory of cognitive dissonance*, Stanford: Stanford University Press.
- Gouldner, A.W., 1960, "The norm of reciprocity: a preliminary statement", *American Sociological Review*, 25 : 2, 161-178.
- Granovetter, M., 1978, "Threshold models of collective behaviour", *American Journal of Sociology*, 83:6, 1420-1443.
- Hardin, R., 1992, *Collective action*, Baltimore: The Johns Hopkins University Press.
- Hechter, M. & Opp, K-D., (ed.), 2001, *Social norms*, New York, Russel Sage Foundation.
- Holländer, H., 1990, "A social exchange approach to voluntary cooperation", *American Economic Review*, 80:5, 1157-1167.

- Koford, K.J., & Miller, J.B., (ed.), 1991, *Social norms and economic institutions*, Ann Arbor, The University of Michigan Press.
- Kollock, P., 1998, "Social dilemmas: the anatomy of cooperation", *Annual Review of Sociology*, 24: 183-214.
- Kübler, D., 2001, "On the regulation of social norms", *Journal of Law, Economics and Organization*, 17:2, 449-476.
- Leibenstein, H., 1950, "Bandwagon, snob and Veblen effects in the theory of consumers' demand", *Quarterly Journal of Economics*, 64:2, 183-207.
- Leydard, J.O., 1995, "Public goods: a survey of experimental research" in Kagel, J.H. & Roth, A.E., (eds.) *The handbook of experimental economics*, Princeton: Princeton University Press.
- Lewis, D.K., 1969, *Convention*, Cambridge MA: Harvard University Press.
- Lindbeck, A., Nyberg, S., Weibull, J.W., 1996, "Social norms, the welfare state and voting", Seminar papers, Institute for International Economic Studies, Stockholm University.
- Mead, G.H., 1967, *Mind, self and society*, Chicago, The University of Chicago Press.
- Olson, M., 1965, *The logic of collective action*, Cambridge MA: Harvard University Press.
- Opp, K-D., 1979, "The emergence and effects of social norms, a confrontation of some hypotheses of sociology and economics", *Kytlós*, 32:4, 775-801.
- Rege, M., 2001, *Five essays on the role of status, norms and reputation in social and economic interaction*, Thesis, Economics Department, University of Oslo.
- Scott, J.F., 1971, *Internalization of norms*, Englewood Cliffs: Prentice-Hall.
- Sen, A.K., 1977, "Rational fools: a critique of the behavioral foundations of economic theory", *Philosophy and Public Affairs*, 6:4, 317-344.
- Sherif, M., 1966, *The psychology of social norms*, New York, Harper Torchbooks.
- Ullmann-Margalit, E., 1977, *The emergence of norms*, Oxford: Clarendon Press.
- Veblen, T., [1894] 1943, "The economic theory of woman's dress", in L. Ardzrooni (ed.), *Essays in our changing order*, New York, Viking Press.
- Von Borgstede, C., 2002, *The impact of norms in social dilemmas*, Thesis, Department of Psychology, Göteborg University.
- Wrong, D.H., 1961, "The oversocialized conception of man in modern sociology", *American Sociological Review*, 26:2, 183-193.