The Evolution of Theories of Collective Action

Ernesto Reuben

Tinbergen Institute

reuben@tinbergen.nl

ABSTRACT. Explaining collective action is still one of the most challenging problems for the social sciences. The purpose of this paper is to describe how the main ideas on the subject have evolved over the years and additionally, to illustrate how our understanding of the problem has grown. It starts by describing perhaps the most influential model in the literature, namely Olson's model of collective action. Later on, it uses the same model to relate six essentially different approaches to solve the collective action problem; for each, highlighting their contributions to our general understanding as well as their main drawbacks. I conclude by stressing the fact that we still do not have a satisfactory theory of collective action. However, recent work is described which uses cognitively and emotionally bounded agents and promises to deliver significant improvements.

April 2003

1. Introduction

Unfortunately, even though in our lives we constantly engage in collective action, that is, everyday we produce and consume collective goods, our understanding of how people organize to produce such goods is far from complete. On the surface, this might seem trivial; after all, we have been working in groups to provide such goods from the dawn of humankind. Therefore, one could reasonably ask, with so many problems yet to solve, should we invest time and effort investigating an everyday phenomenon that seems to have worked so well for so long? The answer to such question is unquestionably yes. Although, understanding collective action is certainly interesting because of its successes, it is essential because of its failures. It is true that we have been providing ourselves with collective goods for a long time. However, there have also been uncountable instances when we have failed to do so. Understanding why is one the highest priorities of the social sciences (Ostrom, 1998).

As can be expected, such an important subject has received considerable attention both inside and outside economics, and therefore, the literature related to it is quite extensive. It would require much more than one paper or even a book to review all the work that has been written on this field of study. Therefore, it is not intended in the following pages to cover all these material. Instead, the purpose of this paper is to describe how some of the main ideas on the subject have evolved over the years and additionally, to illustrate how our understanding of collective action has grown.

The paper consists of four more sections. In section two, I describe the most influential model in the literature, namely Olson's model of collective good provision. With this model, it is possible to understand most of the research on collective action as different attempts to solve the main idea highlighted by it, specifically, the free rider problem. Exactly this is done in section three, where the literature is divided into six essentially different approaches to Olson's argument. In section four, the suggestions for future research are presented, with special emphasis on models that incorporate bounded rationality and bounded reasoning. Lastly, section five concludes.

2. The Problem of Collective Good Provision

As it turns out, in spite of our daily collective behavior, we have a clearer understanding of why individuals fail to act collectively than why they succeed. In his much quoted book *The Logic of Collective Action*, Mancur Olson (1965) clearly explains that, when the decision to provide the collective good is analyzed from the individual point of view, there is a high incentive to free-ride on the efforts of the others and to provide a sub-optimal level of the good oneself. Of course, if everybody acts in such a way, very little if any of the collective good would be supplied to the group as a whole. Thus, in his book Olson argues that unless a group has very specific characteristics, the provision of the collective good is doomed to fail. This stark conclusion clearly explains why many identifiable groups fail to organize themselves despite the obvious benefits of doing so.[1] Moreover, it also highlights the fact that successful groups must somehow overcome the free-rider problem. Understanding how this is done has been the focus of a vast literature, which has given us numerous insights but has still not provided a complete and satisfactory explanation to collective good provision.

In the remaining part of this section, we will first go through Olson's formal model. Its simplicity and logic can be useful to understand the difficulties faced by groups confronting a collective action problem. Additionally, it also helps us distinguish the different characteristics of the problem that other authors have concentrated their research on. Secondly, to conclude we will take a quick look as to how Olson's model can also be seen as an N-person prisoners' dilemma, as was pointed out by Hardin (1971 and 1982).

2.1 Olson's model

Suppose a group consists of $N = \{1, \ldots, n\}$ individuals who can all produce a non-negative amount of a collective good. More specifically, every individual $i \in N$ produces an amount $\gamma_i \geq 0$. Thus, the total amount produced of the collective good Γ is given by the sum of individual contributions, $\Gamma = \sum_i \gamma_i$. The utility gained by an individual i from the consumption of the collective good is captured by $v_i(\Gamma)$ where $v_i' > 0$ and $v_i'' \leq 0$. Although, the utility of the individual depends on the total amount produced of the good, each individual can value the consumption of the collective good differently.[2] The utility gained by the whole group, $V(\Gamma)$, is given by summing the utilities of the individuals, [3]

---

[1] Oppressed citizens in an authoritarian country, consumers faced with protectionist tariffs, and small stake shareholders in a sale of a firm, are just a few of such examples.

[2] Individuals may value differently the consumption of the collective good for numerous reasons. It could be simply a matter of size; for example, an individual that owes a large plot of land would benefit much more from a one percent reduction in property taxes than an individual that owes a small plot of land. However, it can also be that people perceive differently the importance of the collective good; for example, some neighbors might appreciate more than others the added security of a policeman walking around the neighborhood.

[3] In fact $V(\Gamma)$ can be seen as the social welfare function of the group. In this case, a specific type for the social welfare function is chosen, namely, a function that gives the same weight to the utilities of all its members. However, a different type of welfare function can be used without changing the main results of the model. The existence of such welfare functions has been questioned extensively in the social welfare literature. Nevertheless, addressing this issue is not this paper's propose, for discussions on this subject see Sen (2002).

$V(\Gamma) = \sum_i v_i(\Gamma)$. The cost or utility loss of each individual for the production of the good is given by the cost function $C_i(\gamma_i)$ where $C_i' > 0$ and $C_i'' > 0$.[4] Finally, in order to formalize the maximization problem faced by each individual, we should note that an individual produces a positive amount of the collective good only if the utility she gains by doing so exceeds the utility she looses, that is, if $v_i(\Gamma) > C_i(\gamma_i)$. Thus, each individual in the group faces the following maximization problem:

$$\max_{\gamma_i} v_i(\Gamma) - C_i(\gamma_i) = v_i\left(\gamma_i + \sum_{i\neq j}\gamma_j\right) - C(\gamma_i)$$
$$s.t.\ v_i(\Gamma) > C_i(\gamma_i)$$

In the simplest case, in which individuals take the actions of others as constant, that is when $d\gamma_j / d\gamma_i = 0$ for all i and $j \neq i$, the best response γ_i^* of each individual $i \in N$ is given implicitly by,[5]

$$\begin{cases} dv_i(\gamma_i^* + \sum_{i\neq j}\gamma_j) \big/ d\gamma_i = dC_i(\gamma_i^*) \big/ d\gamma_i & if \quad v_i(\gamma_i^* + \sum_{i\neq j}\gamma_j) > C_i(\gamma_i^*) \\ \gamma_i^* = 0 & if \quad v_i(\gamma_i^* + \sum_{i\neq j}\gamma_j) \leq C_i(\gamma_i^*) \end{cases}$$

The solution to the problem is obtained when all individuals use their best response. That is, for every i,

$$\begin{cases} dv_i(\Gamma^*) \big/ d\gamma_i = dC_i(\gamma_i^*) \big/ d\gamma_i & if \quad v_i(\Gamma^*) > C_i(\gamma_i^*) \\ \gamma_i^* = 0 & if \quad v_i(\Gamma^*) \leq C_i(\gamma_i^*) \end{cases} \text{ and } \Gamma^* = \sum_i \gamma_i^*$$

Although this model is extremely simple, it already gives us important insights into the collective action problem. As expected, the individual will produce the collective good up to the point where her marginal gain equals her marginal loss. Unfortunately, the individual does not take into account the effect of her production on the utility of other group members and thus produces a sub-optimal amount of the good (from the group's perspective). To see this more clearly, consider the maximization problem the group would face if it acted as a unitary actor. In this case the maximization problem would be,

---

[4] This cost function is slightly different from the one used by Olson. He used a cost function of the form $C_i(\gamma_i, \Gamma_{-i})$ where $\partial C_i / \partial \gamma_i > 0$ and $\partial C_i / \partial \Gamma_{-i} > 0$. Nevertheless, the use of the simpler cost function above does not alter the insights given by the model.

[5] We also need the second order conditions to hold, that is $v_i'' - C_i'' < 0$. Since by assumption $v_i'' \leq 0$ and $C_i'' > 0$ this condition always holds.

$$\max_{\gamma_1,\ldots,\gamma_n} V(\Gamma) - \sum_i C_i(\gamma_i)$$
$$s.t.\ V(\Gamma) > \sum_i C_i(\gamma_i)$$

Consequently, the optimal amount produced by every individual i would be given implicitly by,

$$\begin{cases} \sum_j \partial v_j(\Gamma^{**}) \big/ \partial \gamma_i = \partial C_i(\gamma_i^{**}) \big/ \partial \gamma_i & if \quad V(\Gamma^{**}) > \sum_i C_i(\gamma_i^{**}) \\ \gamma_i^{**} = 0 & if \quad V(\Gamma^{**}) \leq \sum_i C_i(\gamma_i^{**}) \end{cases} \text{ and } \Gamma^{**} = \sum_i \gamma_i^{**}$$

Where γ_i^{**} is the amount of the collective good produced by i which maximizes $V(\Gamma)$. Since $dv_i(\Gamma)/d\gamma_i < \sum_j \partial v_j(\Gamma)/\partial\gamma_i\ \ \forall\ \Gamma > 0,\ i \in N$, and $C_i'' > 0\ \forall\ i \in N$, it follows that, γ_i^{**} must be strictly bigger than γ_i^{*} for all i, and consequently Γ^{**} is strictly bigger than Γ^{*}.[6] In other words, there is an undersupply of the collective good.

These results might not seem particularly breathtaking; after all, undersupply is a basic result in any public good. What is striking, especially at the time when Olson published his book, is that he draws attention to the public good character of collective goods. Before Olson's book, most people thought of public goods as things such as air pollution and public parks; problems that might be solved through the intervention of the government or by a clearer definition of property rights. However, Olson points out that the free-rider problem is present in many other cases. In fact, it is pervasive in society since the above analysis can be applied to almost every group one can think of. From employees working in teams in a firm to nationwide elections, the free-rider problem is much more common than it was initially thought. Olson did not put it quite this way. He acknowledges that this need not apply to small groups in which communication, sanctions and care for others play important roles. However, when it comes to large groups the free-rider problem is bound to have a significant effect.

As it was hinted earlier, the other important result in Olson's model is the effect of group size on the provision of the collective good. The general argument is that large groups will have more problems providing themselves with an optimal amount of the collective good than small groups. This can be seen easily from the results above. It is rather obvious to note that if we add more individuals to the group, $\sum_j \partial v_j(\Gamma^{**})/\partial\gamma_i$ grows while $dv_i(\Gamma^{*})/d\gamma_i$ remains constant. Therefore, the difference between the optimal amount of the collective

[6] The only exception is when costs are so high that even for the group as a whole there is no reason to produce the collective good, that is $\Gamma^{**} = \Gamma^{*} = 0$.

good for the group and the amount of the good provided by individual i increases. In other words, as the group grows larger the bigger is the undersupply of the collective good.

In his book, Olson differentiates between three types of groups: privileged, intermediate, and latent. A privileged group can be thought of as a group in which one of the members receives a disproportionately high utility from the consumption of the collective good, and thus, she would be willing to produce a big share of the good herself. Hence, the group would produce an almost optimal amount of the collective good without any need of organization. The intermediate group lacks such a group member; however, it is still possible for a small number of individuals to coordinate their actions and produce an almost optimal amount of the good. The key to the existence of an intermediate group is that the group is not big enough so that the actions of an individual can considerably affect the utility of the other members. If this is the case then a combination of strategic interaction and institutions like monitoring, sanctioning, and rewarding might be enough to facilitate an adequate provision of the collective good. Latent groups on the other hand are too big for their members to have a considerable effect on each other's utility. In this case, the provision of the collective good is doomed to fail. First because as we have seen, larger groups are bound to supply less of the collective good, and second because not being able to affect the payoffs of others makes many institutions unfeasible. In fact, in the presence of high costs, the collective good might not be provided at all. To see this, recall that if $v_i(\Gamma^*) < C_i(\gamma_i^*)$ then individual i will not produce the collective good. Hence, we can have a situation in which the previous inequality holds, and at the same time for $v_i(\Gamma^{**}) > C_i(\gamma_i^{**})$ to be true as well for every $i \in N$. In other words, nobody produces the collective good even though everybody could benefit from doing so. A case in which fixed costs are relatively high for every individual could constitute such an example.

2.2 The N-person prisoner's dilemma

One of the most useful characterizations of Olson's model is the N-person prisoner's dilemma. The prisoners' dilemma, perhaps the most famous of all games, has been studied extensively by many authors. Hence, thinking of this model as an N-person prisoner's dilemma allows us to use the results from that literature to better understand collective action. Furthermore, it gives an intuitive explanation of the situation faced by the group.

To start, lets highlight the characteristics of an N-person prisoner's dilemma. Such a game is characterized by $N = \{1, \ldots, n\}$ players, each of whom can take an action $A \in \{c, d\}$, they can either cooperate of defect. Each player has a payoff for each of the possible outcomes of the game, that is $\pi_i(A_1, \ldots, A_n) \in \Re^{2 \times n}$. The payoffs of the game are such that, for all players, the strategy to always defect strictly dominates all other strategies, and hence, all players defecting is the unique Nash equilibrium. Consequently, each player receives the payoff $\pi_i(d_1, \ldots, d_n)$. As it turns out, for all players $\pi_i(d_1, \ldots, d_n) < \pi_i(c_1, \ldots, c_n)$. Therefore, although everybody would be better off if all cooperated, all players defect and end up with a lower payoff.

As Hardin demonstrated, in some instances, Olson's model has the same structure as the N-person prisoner's dilemma. To illustrate we will examine one such case, specifically, the case in which the utility gained by an individual from the consumption of the collective good is linear, that is, $v_i(\Gamma) = v_i(\gamma_1 + \ldots + \gamma_n) = v_i\gamma_1 + \ldots + v_i\gamma_n$ for all $i \in N$. Furthermore, we assume that, if an individual is the only one that produces a positive amount of the collective good, then the costs of doing so would exceed the benefits, that is, for any $i \in N$, if $v_i(\Gamma) = v_i\gamma_i$ then $v_i(\Gamma) < C(\gamma_i)$ for all $\gamma_i > 0$. Moreover, if the individual produces zero units of the collective good then costs are equal to zero, $C(0) = 0$. Note that this describes well the situation faced by both intermediate and latent groups; in both types of groups, no single individual sees it as profitable to produce a significant amount of the collective good if nobody else does it. Now, lets look at the payoff of individual i when others produce a positive amount of the collective good. If i decides to produce an amount $\gamma_i > 0$, her payoff would equal $v_i\gamma_i + \sum_{j \neq i} v_i\gamma_j - C(\gamma_i)$. Similarly, if i decides to produce zero units of the collective good, her payoff would equal $\sum_{j \neq i} v_i\gamma_j$. Since, by assumption $v_i\gamma_i < C(\gamma_i)$, this implies $\sum_{j \neq i} v_i\gamma_j > v_i\gamma_i + \sum_{j \neq i} v_i\gamma_j - C(\gamma_i)$ for all possible values of γ_j. Hence, we can conclude that individual i prefers $\gamma_i = 0$ to any positive amount $\gamma_i > 0$ irrespective of the amount produced by other group members. In other words, individual i has a dominant strategy to defect, that is, produce nothing. Since this applies to all individuals $i \in N$, in the resulting outcome nobody will produce a positive amount and the collective good will not be provided at all. Finally, if we make the very reasonable assumption that there exists a set of values $\{\gamma_1, \ldots, \gamma_n\}$ for which $v_i\gamma_1 + \ldots + v_i\gamma_n > C(\gamma_i)$ for all $i \in N$,[7] we get exactly the same structure of the N-person prisoners' dilemma. Everybody would be better off if all produced a positive amount, however all individuals choose to produce zero and everybody ends with a lower payoff.

---

[7] Note that fulfilling this assumption in large groups should be relatively easy. Since $v_i\gamma_1 + \ldots + v_i\gamma_n$ grows as the group becomes larger while $C(\gamma_i)$ remains constant, for any value of γ_i one could find a group large enough so that $v_i\gamma_1 + \ldots + v_i\gamma_n > C(\gamma_i)$ holds.

3. The Collective Action Literature

The free-rider problem faced by individuals when acting collectively provides a simple and intuitive explanation to why so many people fail to organize themselves even though they would certainly benefit from doing so. However, recognizing this fact simply brings out another question: how do the existing groups we observe in our everyday lives overcome this free-rider problem? It is certainly not true that all organized groups are of the privileged type. Moreover, there is plenty of evidence that the provision of collective goods is, in many cases, above the Nash prediction of the model previously described.[8] Consequently, some kind mechanism that helps people cooperate must exist. Since Olson published his book, there has been a vast amount of papers written in an attempt to investigate exactly this issue.

The literature related to collective action is quite extensive, and therefore, it cannot be covered all in this survey. Nevertheless, with a few examples and by examining what part of Olson's model different authors choose to concentrate on, we can obtain a better picture of the state of research in this area. In this section, we will go through six quite distinct ways of tackling the collective action problem. Each approach, concentrates on different elements of Olson's argument in order to provide us with conceivable explanation to the provision of collective goods.

Here is a brief summary of each:

- Overcoming the lack of excludability through the use of selective incentives: In this approach, the free rider problem is overcome providing a selective incentive, from which exclusion is possible, along with the collective good, from which exclusion is not.
- Further analysis of the strategic interaction of intermediate groups: In his book, Olson concentrates on latent groups and neglects to analyze the strategic interaction present in intermediate groups. This approach concentrates on studying this interaction. It presumes the Nash equilibrium might be a good predictor of collective behavior in spite of the free-rider problem.
- Changing the game that best describes collective action: The literature that covers this topic questions the basic structure of the model presented by Olson. Solutions to the free-rider problem are found by altering the model so that there exist equilibria in which there is a high amount of collective good provision.

---

[8] There are many examples many of which can be found in Davis and Holt (1993), Ledyard (1995), Offerman (1997), and van Winden (2002).

- Changing the utility function of individuals, allowing for social preferences: This approach changes the assumption that individuals posses entirely selfish preferences. Collective action is explained by altering the utility function of individuals in multiple ways in order to make provision of the collective good more attractive.
- Agent types and evolutionary models: This approach attempts to test under what conditions is collective action a good survival strategy. Multiple agent types (strategies) are used. The free rider problem is overcome by the emergence of agent types that prefer cooperation.
- Networks, information transmission, and bandwagon effects: The authors in this approach assume away the free-rider problem present in Olson's model and concentrate instead in the coordination problem that arises when multiple individuals have to act collectively. Since communication is crucial for coordination, what can be communicated and how communication flows becomes crucial.

3.1 Provision of selective incentives along with the collective good

One of the important characteristics of collective goods is that individuals cannot be excluded from consuming the collective good even though they might not contribute to its production. Precisely for this reason the free-rider problem emerges and underprovision occurs. Hence, it is natural to think that a way of overcoming the free-rider problem is by incorporating excludability into the public good. To do so, one can use selective incentives that, unlike the collective good, can be applied to individual members of the group.

One such incentive was suggested by Olson himself in *The Logic of Collective Action*, namely, providing a private good along with the collective good. Since private goods are excludable and therefore do not suffer from a free-rider problem they can be used to make the provision of the public good more attractive and avoid free riding. To see this we can use again the special case in which $v_i(\Gamma)$ is linear. As we have seen previously, if an individual i cooperates by producing an amount $\gamma_i > 0$ of the collective good, she receives a utility equal to $v_i\gamma_i + \sum_{j\neq i} v_i\gamma_j - C(\gamma_i)$, while if she defects, by choosing $\gamma_i = 0$, her utility would equal $\sum_{j\neq i} v_i\gamma_j$. Since $v_i\gamma_i < C(\gamma_i)$ defection pays more than cooperation. Now lets assume that although individuals cannot be excluded from the consumption of the collective good, their contributions can still be observed and thus free riders can be identified. In this case, a private good can be used to eliminate the incentive of individuals to free ride. For example, lets say the utility gained by an individual $i \in N$ from the consumption of the private good is given by $u_i(\alpha(\gamma_i, \gamma^*), x_i, p) = \alpha(\gamma_i, \gamma^*)(u_i(x_i) - px_i)$ where x_i is the amount of the private good consumed by i, p is the price paid for x_i, and $\alpha(.)$ is the indicator function, specifically $\alpha(\gamma_i, \gamma^*) = 0$ if γ_i

$< \gamma^*$ and $\alpha(\gamma_i, \gamma^*) = 1$ if $\gamma_i \geq \gamma^*$. In other words, we assume that an individual can consume the private good only if she produces a minimum amount $\gamma^* > 0$ of the collective good.[9] Now, if the individual decides to cooperate and chooses $\gamma_i \geq \gamma^*$ her utility would be equal to $v_i\gamma_i + \sum_{j\neq i} v_i\gamma_j - C(\gamma_i) + u_i(x_i) - px_i$, and since the utility of defection remains unchanged, if the utility gained from the consumption of the private good is big enough, an individual can get a higher utility by cooperating than by defecting. Namely, if $v_i\gamma^* + u_i(x_i) > C(\gamma^*) + px_i$ for some $x_i > 0$, then the individual i will choose to participate in the production of the public good. Thus, this setup implies that the provision of a private good can allow even a latent group, where strategic interaction is not possible, to provide itself with an adequate amount of the collective good. So for example, even though a union cannot avoid benefiting non-union workers when it negotiates higher wages it can provide its members with an excludable good such as low-interest credit so that non-members have an incentive to join and pay the union fees.

Private good provision is not the only selective incentive that has been suggested. The use of sanctions is another important incentive (or more accurately a disincentive) that has been considerably studied.[10] The modeling of sanctions into Olson's setup is slightly more complicated. However, one can get a clear picture of the intuition behind it by taking the very simple case in which the cost of sanctioning is shared equally among group members.[11] In this case, the utility an individual i receives by defecting is given by $\sum_{j\neq i} v_i\gamma_j - s(\gamma^*) - C(s)$, where $s(\gamma^*)$ is the sanction an individual receives when producing an amount $\gamma_i < \gamma^*$ and $C(s)$ is the unitary cost of the sanctioning mechanism.[12] The utility of cooperating is equal to $v_i\gamma_i + \sum_{j\neq i} v_i\gamma_j - C(\gamma_i) - C(s)$. An individual will cooperate and produce $\gamma_i = \gamma^*$ if $v_i\gamma^* > C(\gamma^*) - s(\gamma^*)$, and this way, if $s(\gamma^*)$ is high enough, cooperation can be enforced. However, as we will see later, the use of sanctions is itself a collective good and therefore suffers from the free rider problem. Consequently, we are left with an incomplete

---

[9] Of course this is a very simple case in which there is a threshold value γ^* after which consumption of the private good is possible. One can also use a utility function $u_i(x^{max}(\gamma_i), x_i, p)$ where $x^{max} \geq x_i$, in which the maximum amount of the private good that an individual can consume, x^{max}, depends on the amount produced of the collective good.

[10] Other types of selective incentives have been suggested. A couple of examples would be: the use of side payments (Schofield, 1975) and encouragement (Coleman, 1988). Also, more generally Wilson (1973) identifies three broad types of selective incentives.

[11] In the more realistic case in which such a sanctioning mechanism does not exist, one has to model who bears the cost of sanctioning. See for example Hardin (1982) or Heckathorn (1988).

[12] Similarly to the private good provision model, one can also use a sanctioning schedule $s(\gamma_i)$ which establishes a sanction to every contribution γ_i.

explanation. Nevertheless, even though there are theoretical complications, empirical evidence suggests that people are indeed willing to use costly sanctions, and consequently, it is a successful way of promoting cooperation.[13] One clear example of the use of sanctions is the aggressive response that striking workers exhibit towards coworkers who decide to cross the picket line. In order to understand such cases, more research on the implementation of sanctions in definitely merited.

Although the selective incentives approach provides a plausible solution to free riding, it assumes very specific conditions are met, and thus, it leaves many aspects unexplained. In the case of a positive incentive such as the provision of a private good, the group has first to find a good that appeals enough to all group members so that it can be effectively used to bribe individuals into cooperating, such a good might not exist. Furthermore, even if such a good does exist, then why do private firms not provide it? One would think that, a firm that is not saddled by the additional costs related to the production of the collective good would be more efficient producing the private good and thus could offer it at a lower price.[14] Therefore, unless the group has an unexplained advantage over normal firms or is the sole producer of the private good, such a selective incentive does not provide an adequate solution (Udéhn, 1993). Similarly, in the case of a negative incentive, such as sanctions, one can reasonably ask, why would a group member who wants to defect accept the implementation of a sanctioning mechanism? If exiting the group is possible then it sounds more logical to free ride and then exit in order to avoid the sanction. In other words, for sanctions to work leaving the group must be impossible. A few explanations of why people might not leave have been suggested, for instance by assuming loyalty to the group (Hirschman, 1970) or some kind of dependence (Hechter, 1983). However, they have not received much attention since they do not explain how can group loyalty or dependence develop in such a setting. Another important limitation of this approach is that it is restricted to the cases in which contributions to the collective good are observable. This assumption is particularly strong when it is applied to latent groups in which sheer size would make observing individual actions very costly if not impossible (Hechter, 1984).

Finally, and perhaps most importantly, the biggest problem of the selective incentive approach is that it solves the initial free riding problem but it creates a second-order

[13] For experimental evidence see reference Fehr and Gächter (2000).

[14] In the example used above, non-union members would be able to get credit with an even lower interest rate from a bank and thus have no incentive to join the union. Of course, this ignores the lower cost a union might face screening credit applications. A union might have more accurate information and thus face a smaller adverse selection problem.

collective action problem. In the private good case, the group must have organized itself somehow to start the provision of the private good and to agree on the rules for its provision. Such an organization is itself a collective good and thus faces a similar free riding problem. It might be the case that an organization is formed to produce the private good and then it decides to facilitate the provision of the public good. However, this kind of explanation relegates collective action to almost a side effect. Individuals would be mainly concerned with acquiring the private good and they participate in collective action simply because doing so allows them to consume the private good. Although this side effect case can certainly occur, it by no means describes most of the observed collective action. In the case of sanctions, the bearing of the cost of sanctioning is also a collective good. If sanctioning leads to cooperation then an individual can certainly benefit if she decides to sanction defectors. However, she would benefit even more if other group members were the ones that sanction the defectors. Consequently, although selective incentives can solve a collective action problem, they do so by creating another (Oliver, 1980). If the selective incentives were used again to solve the second-order collective action problem, it would simply create a third-order problem followed then by a forth-order problem and so on in an infinite regress. Without clarifying how these higher order problems are solved, selective incentives cannot satisfactorily explain why collective action occurs in so many cases in our societies.[15]

3.2 Further analysis of the strategic interaction of intermediate groups

When Olson wrote his book in 1965, game theory and the analysis of non-cooperative games was not commonly used by economists. Hence, lacking the proper tool for the analysis of strategic interaction, Olson concentrated on the study of collective action by latent groups. However, with the development and popularization of game theory, further analysis of the strategic interaction present in intermediate groups became a fertile area for research. In general, this approach acknowledges that in very large groups there will be a considerable undersupply of the collective good. Nevertheless, the focus of this research is not to see how to overcome this problem but to analyze what is the predicted level of provision and how it is affected by different variables.[16]

---

[15] Among others, Heckathorn (1989) and more recently Fehr and Gächter (2002) argue individuals are intrinsically willing to punish. However, the reason why this is so is still not clear.

[16] As we will see later there is an effort to explain high contributions in large groups but this is done by reducing in some way the number of people with whom a single individual interacts. For example, by dividing the large group into smaller ones.

Much of the analysis in this area can be taken directly from the public goods literature. Since collective goods, as conceived by Olson, share a very similar setup to public goods, all the research done on public goods could be used to understand collective action. It would be beyond the purpose of this paper to review all the public goods literature.[17] Therefore, just a couple of examples of the main results will be illustrated here.

One important result, illustrated by Bergstom, Blume and Varian (1986), is the effect of income redistribution on the equilibrium amount of the collective good. They find that a small redistribution of income among contributing individuals will not affect the total provision of the collective good. However, if the redistribution is large enough so that the set of contributors changes, then the amount produced will indeed vary. To see this we should recall that the amount an individual contributes to the collective good is given by the set of equations:

$$\begin{cases} dv_i(\Gamma^*)/d\gamma_i = dC_i(\gamma_i^*)/d\gamma_i & \text{if} \quad v_i(\Gamma^*) > C_i(\gamma_i^*) \\ \gamma_i^* = 0 & \text{if} \quad v_i(\Gamma^*) \le C_i(\gamma_i^*) \end{cases}$$

We should note two important characteristics of this set of conditions. First, note that for individuals that provide positive contributions, in equilibrium the amount they decide to contribute depends on the marginal utility received from the collective good, which in turn depends on the total amount contributed. One can think of the individual as having a preferred total amount of the collective good and adjusting her contribution in order to reach this amount. The second point to note is that the cost function $C_i(\gamma_i)$ can be seen as simply the opportunity cost of consuming the collective good. In other words, it is the utility lost because of forgone consumption of private goods due to money spent on the contribution to the collective good. In addition, if the utility of consuming other goods exhibits decreasing marginal returns, then it is easy to see that, when an individual *i* has a high income, $C_i(\gamma_i)$ will be lower for all $\gamma_i > 0$ than when she has a low income. In other words, for a given γ_i, if an individual has a high income she consumes a large amount of other goods, and because of decreasing marginal returns, she looses less utility for any increase in her level of contribution than when she is poor.[18] Thus, changes in income can be seen as changes in the

---

[17] For the interested reader, excellent introductions to the public goods literature can be found in Baumol and Oates (1988), and Laffont (1988). Furthermore, see Sandler and Tschirhart (1980) for a survey in the related theory of clubs.

[18] This assumes that the saving rate does not change with wealth and thus, for instance, for a fixed contribution to the collective good, a 10% increase in wealth would mean a 10% increase in consumption of other goods.

cost function $C_i(\gamma_i)$. Combining these two observations Bergstom, Blume and Varian show that a small redistribution of wealth among contributing individuals will not affect the set of individuals for whom $v_i(\Gamma^*) > C_i(\gamma_i^*)$ holds. In this case, even though income changes, contributors will still find optimal the same amount of the collective good Γ^* and therefore, will adjust their contributions in order to attain it. An interpretation of this result indicates that, if the government taxes contributors in order to provide some of the collective good itself, then contributors will simply adjust their contribution by the taxed amount and Γ^* will not change. This result is usually stated as government intervention "crowds out" private provision of collective goods. On the other hand if the redistribution of wealth is big enough so that $v_i(\Gamma^*) > C_i(\gamma_i^*)$ no longer holds for at least one contributor then her contribution will drop to zero and the equilibrium amount Γ^* will vary.[19]

Following this line of thought one can easily see another important characteristic of collective goods, that is, the exploitation of the big by the small. This result has been emphasized by Olson (1965) himself as well as other authors (see Stigler, 1974, and Oliver, Marwell and Teixeira, 1985). To see more precisely what is meant by this, remember that an individual's decision to contribute a positive amount versus contributing nothing depends on the inequality $v_i(\Gamma^*) > C_i(\gamma_i^*)$. Hence, by observing what type of individuals are more likely to satisfy it, one can get an idea of the characteristics of the contributors. To start, recalling that individuals with higher incomes have a lower $C_i(\gamma_i)$ one can presume that the rich are more likely to satisfy the above inequality than the poor. Similarly, it is rather obvious to note that individuals that value more each unit of the collective good,[20] meaning they posses a high $v_i(\Gamma)$, are also more likely to satisfy the inequality. Thus, we can infer that contributors in general will be bigger, that is, will have higher incomes and valuations, than non-contributors. Since the smaller non-contributors receive the benefits of the collective good without having to pay for it, they benefit from the presence of the contributors. If it were not for the presence of the big, then the small might not only have to contribute to the collective good themselves but they would also enjoy a much lower total production of the good (if any, since the collective good might not be produced at all). Although not tested extensively these results have been observed empirically (see Stigler, 1974).

Besides public goods, variations of Olson's model have been analyzed in order to explain other more complicated cases of collective action. One important variation that fits this

[19] The total amount of the collective good might increase or decrease depending on who gets more and who gets less income. For details see Bergstom, Blume and Varian (1986).

[20] High valuations can be associated to a bigger size, see footnote 2 for further discussion.

description is the case in which you have more than one group competing with each other for the collective good. In this case, an interesting relation exists between group size and the probability of winning the good. As we have seen, large groups suffer from a bigger free riding problem, however, since they have more members, they still might have a larger probability of winning the collective good.[21] This relation has been extensively studied in the literature on rational voter turnout. In this case, winning an election is considered a collective good and the majority rule determines who wins it. Very complete analysis of this voting model can be found in Ledyard (1984) and in Palfrey and Rosenthal (1985). In short, the general conclusion of these models is that, in the presence of some small voting costs, in most circumstances voter turnout for large groups will be considerably low. In fact, it can be so low that in some cases majorities can lose elections to much smaller groups (Haan and Kooreman, 2000). Such cases have indeed presented themselves in modern democracies,[22] giving some validity to the model's argument. However, the difference between actual turnout in national elections and the turnout predicted by the model is so big that we cannot accept the model as true in its present form.

However, there are other settings, such as mass political demonstrations, in which a similar model performs much better. Unlike in national elections, relatively few people participate in mass demonstrations. Thus, a model that predicts low turnout might be much more applicable. Lohmann (1993 and 2000) proposes such a model. The author uses a similar setup to the voting model to demonstrate how rational agents can decide to participate in mass demonstrations in order to transmit information to a policy maker. In her model, there is still free riding and thus, a low turnout. Nevertheless, the rational policymaker takes this into account and is able to perceive some of the information possessed by the public. Lohmann points out how the total number of people demonstrating is not important. What is important is that the number of people demonstrating is considerably above the expected number of demonstrators. Thus, if the policymaker expects a small number of people protesting, then a few additional individuals might make a lot of difference and therefore free riding does not represent such a big problem.

---

[21] For example, Esteban and Ray (2001) prove that large groups are more effective than small groups only if marginal costs rise sharply or if the collective good has a high jointness of supply (that is, a group members can consume the good without reducing the consumption of other group members).

[22] Two recent examples are the Irish rejection of the Nice treaty in a first referendum in 2001 (Economist, 2001a) and the winning by republicans of the mayor's seat in New Your City (Economist, 2001b). In both cases, polls indicated that the larger group constituted of people that in the first case favored accepting the treaty and in the other supported the democrat party.

To conclude, we can affirm that game theoretic studies of variations of Olson's model have provided us with a much better understating of collective action. Many of the predictions of these analyses have been confirmed in numerous experimental settings. For example, people do contribute less to collective goods if their marginal costs rise (Issac, Walker and Thomas, 1984), changes in income are indeed positively correlated to contributions (Chan et al., 1996), and larger groups do exhibit lower relative turnout than smaller groups (Großer and Kugler, 2002). However, there is still the uncomfortable fact that they predict much lower levels of collective action than are observed empirically.[23] Moreover, one must not forget that, although these theories are very general, when an author builds a model he/she usually has some subgroup of collective goods in mind. Hence, some care is needed when interpreting the results presented in a paper. To illustrate such a case, consider the much used income effect from the public goods literature, where it is commonly assumed that as people become richer their demand for public goods increases. This is a reasonable assumption if one has in mind public goods such as unpolluted air or public parks. However, if one considers a much different type of good, a positive income effect might not be reasonable at all. For instance, there is not reason to think that, if there is an exogenous increase in the income of employees in a firm, they will be more willing to form a union to negotiate higher wages. As a final point, one could affirm that in general these theories have been good predictors for the direction of changes in cooperation, but a lot less successful when it comes to the total provision of the collective good. There have been numerous attempts to increase the equilibrium amount of collective action. However, as it is explained in the next approaches, they do so by changing the model considerably.[24]

3.3 Changing the game that best describes collective action

Olson's collective action model has been so successful because it is simple, and at the same time, it elegantly describes an important characteristic present in collective goods namely free riding. Unlike the previous approach, the papers described here tackle precisely the free riding problem. Furthermore, they do so by changing the basic model in such a way that high degrees of cooperation emerge in equilibrium. However, despite this common approach, very different ways of modifying the model have been used. By and large, authors have used three

[23] See evidence in footnote 8.

[24] There are a few exceptions were the model is left almost intact but then strong mainly unexplained assumptions have to be introduced. See for example Morton (1991) and Uhlaner (1989) for two interesting cases in which voter turnout is achieved by dividing the larger groups into smaller subgroups.

different methods to explain cooperation, namely repetition, institutions and production technology.

One serious limitation of Olson's model is that it is a static model. In other words, it does not allow the group to interact repeatedly over time. Once we relax this assumption, it is reasonable to ask oneself if repeated interaction would be able to solve the free riding problem. The answer to this question is indeed yes, but only for some very specific cases. The intuition behind the results can be seen if we analyze the collective action problem as a repeated 2-person prisoner's dilemma. Recall that a prisoner's dilemma game requires the following relationship between player *i*'s payoffs: $\pi_i(d_i, c_j) > \pi_i(c_i, c_j) > \pi_i(d_i, d_j) > \pi_i(c_i, d_j)$, for $i, j \in \{1, 2\}$ and $i \neq j$, where $\{c, d\} \in A$ are the two possible actions, cooperate and defect, for both players *1* and *2*. Normally, such a game has only one Nash equilibrium, that is, both players defecting. However, once we repeat the game indefinitely many more equilibria arise. To see this we can examine one such case, in particular, a fully cooperative equilibrium with the use of a grim-trigger strategy. First, lets note that payoffs of player *1* in the repeated game are: $\Pi_1 = \sum_{t=0}^{\infty} \delta^t \pi_1(A_1^t, A_2^t)$, where $t \in \{0, ..., \infty\}$ indicates the time period and δ the discount rate of future payoffs. Furthermore, lets assume player *2* uses a grim-trigger strategy. More specifically, player *2* will cooperate as long as player *1* cooperates. However, if player *1* defects then player *2* will defect for the rest of the game.[25] Now, lets consider player *1*'s options in the current period. Player *1* knows that if she decides to defect in the current period then player *2* will cooperate that period but defect thereafter. Since the best player *1* can do once player *2* is defecting is to defect herself, defecting this period gives her a payoff of $\pi_1(d_1^0, c_2^0) + \sum_{t=1}^{\infty} \delta^t \pi_1(d_1^t, d_2^t) = \pi_1(d_1, c_2) + \frac{\delta}{1-\delta}\pi_1(d_1, d_2)$. On the other hand, if player *1* chooses to cooperate in the current period and every period thereafter,[26] her payoff would equal $\sum_{t=0}^{\infty} \delta^t \pi_1(c_1^t, c_2^t) = \frac{1}{1-\delta}\pi_1(c_1, c_2)$. Comparing the two payoffs one can conclude that player *1* will decide to cooperate in the current period if,

$$\frac{1}{1-\delta}\pi_1(c_1, c_2) > \pi_1(d_1, c_2) + \frac{\delta}{1-\delta}\pi_1(d_1, d_2) \Rightarrow \delta > \frac{\pi_1(d_1, c_2) - \pi_1(c_1, c_2)}{\pi_1(d_1, c_2) - \pi_1(d_1, d_2)}$$

[25] The grim-trigger strategy is quite drastic in the sense that defection is punished forever. Therefore, it might appear as very unrealistic. However, this strategy is used only as an illustration, other much more realistic strategies, such as tit-for-tat, yield the same general results.

[26] Since this is an infinitely repeated game, these two strategies are the only ones we need to consider when player *2* uses a grim-trigger strategy. Of course, player *1* could cooperate in the current period and then defect in some future period, but this gives us an identical situation as when player *1* considers a defection in the current period. See Fudenberg and Tirole (1991) for details.

Furthermore, since the same analysis can be done for any value of *t* and for any of the two players, it is safe to conclude that cooperation can always be sustained if future payoffs are valued enough (δ satisfies the above inequality). Thus, by simply adding repetition and a strategy that punishes defection we can construct a model where there are feasible cooperating outcomes.

Unfortunately, the use of repetition to solve the free-riding problem has three severe limitations. First, it relies on the very unrealistic assumption that games must be repeated over an infinite number of periods.[27] If it were not for either of these assumptions, then, no matter how long the game is, backward induction predicts that defection in every period is the only solution.[28] Second, for games in which there are more than two players, the strategies require a great deal of coordination. For example, if the group is using a tit-for-tat strategy[29] and a player defects then all the other players have to switch to defection and then back to cooperation at the same time.[30] Any player that fails to make the switch on time will be unfairly punished. Lastly and more devastatingly, this kind of game is subject to the famous conclusions of the Folk Theorem. Concisely, it has been proven that, for any set of payoffs for which $(\Pi_1, \Pi_2) >> \left(\frac{1}{1-\delta}\pi_1(d_1,d_2), \frac{1}{1-\delta}\pi_2(d_1,d_2)\right)$ holds, there exists a $\delta < 1$ such that (Π_1, Π_2) are the average payoffs arising in a sub-game perfect Nash equilibrium. In other words, any sequence of actions, which results in payoffs greater than full defection, can be a feasible outcome of the game (for details see Abreu, 1988). Thus, this model not only predicts that full cooperation is a possible outcome, but it also predicts that pretty much any kind of behavior can arise. Not a very encouraging result, if one is looking for a theory that has some kind of predictive power.

Along with repetition, a very popular method, which allows the free riding problem to be overcome, is the use of different kinds of institutions. The institutional approach focuses its attention on the different elements that encourage free riding and then uses an institution to

[27] One can also make the very similar assumption that the game has no certain end. Instead, in each period the game has only a positive probability of ending (Fudenberg and Maskin, 1986).

[28] Kreps et al. (1982) have modeled cooperation in finitely repeated prisoner's dilemma games. However, they rely on the existence of irrational players who stick to playing only one strategy.

[29] A tit-for-tat strategy is one in which the player initially cooperates and then imitates the move of the other player. Thus, she rewards cooperation by cooperating in the next period and punishes defection by defecting in the next period.

[30] In fact, for large groups such a strategy cannot be used to sustain cooperation since an agent knows that everybody defecting is a too strong punishment. The other agents would get a lower payoff if all defect than if they continue to cooperate and allow one defection. Hence, all defecting ceases to be a credible threat. See Bendor and Mookherjee (1987) for details.

correct the deficiencies. A good example of such an exercise can be found in Dawes et al. (1986). In their paper, the authors first point out how in the prisoner's dilemma a player has two incentives to defect. As was characterized by Coombs (1973), a player defects because: on one hand, she has "fear" of getting the lowest payoff by cooperating when the other player defects, and on the other, the player also has "greed" and wants to get the highest payoff by defecting when the other player cooperates. If seen this way, then an institution that controls either a player's "fear" or "greed" might be enough to prevent free riding. Dawes et al. (1986) call our attention to a couple of institutions that satisfy this requirement and which are already used by some groups in our society. In their opinion, a kind of "money-back" policy eliminates the "fear" of receiving the lowest payoff since it returns your contribution in case the other does not contribute.[31] Similarly, what they refer to as "enforced contribution" eliminates a player's "greed" for the highest payoff because it imposes a fine such that defecting while others cooperate does not give her a higher payoff than cooperating. Other authors have done similar analysis. In some cases, to investigate the effects of existing institutions such as third party monitoring (Agrawal and Goyal, 2001) or matching schedules (Guttman, 1987). In other cases, to suggest a new institutional setting that might successfully promote collective action (Levati and Neugebauer, 2001).

However, even though understating how institutions can promote collective action is very important, one must be careful when interpreting the results, especially when making policy recommendations. To elaborate, we should remember that in this approach we are assuming that free riding is indeed present and therefore an institution is warranted. However, we must acknowledge that in many cases we still do not have a clear idea why people cooperate. Therefore, if we impose an institution in a situation where, for yet unexplained reasons, there is some degree of cooperation, then, there might be unintended consequences and cooperation could even decrease (Hirsch, 1976). Furthermore, the institutional approach regrettably does not provide an explanation of how a group managed to introduce an institution. Just as selective incentives, this approach is subject to the second-order collective good argument. Without third party intervention, forming an institution can be just as difficult as providing the original collective good. To sum up, many of the institutions we observe evolved over time to solve specific and sometimes unique collective action problems. If we do not have a complete understanding of how the institution was created and what specific deficiency it corrects, we might end up using the institution in the wrong situations.

[31] See also Bagnoli and Lipman (1989) for more on the effect of a "money-back" institution on public good provision.

Lastly, the third method by which the free riding problem in Olson's model has been dealt with is what can be referred as the production technology critique. This line of research questions Olson's assumption that the technology used to produce the collective good presents increasing marginal costs with respect to contributions (both one's own as well as other's).[32] Once this assumption is relaxed, the model changes dramatically to the point that free riding might even disappear while other problems, such as coordination, become prominent. Numerous papers have been written on this subject using many different cost functions, from decreasing (Elster, 1989) to S-shaped and even discontinuous functions (Oliver, Marwell and Teixeria, 1985; Oliver and Marwell, 1988a).[33] It is this author's opinion that a big part of the disagreement comes from the types of collective goods the different authors have in mind. If one thinks of mass demonstrations it would be natural to think that the cost of protesting does not increase and might even decrease with the number of protestors. On the other hand, if one thinks of campaign contributions to "buy" votes for a candidate, more contributions by others would definitely make your own contributions less effective as voters become increasingly more difficult to convince. On the whole, this research makes a very valid point: production functions cannot be assumed to have only one shape. Many different shapes might exist and analysis of each should be undertaken. A very illustrating example of such an exercise is present in Heckathorn (1996). In the following paragraphs, we will describe in more detail a slightly modified version of it.

To illustrate what types of games might arise in a collective action setup Heckathorn (1996) uses a 2-player game and a few assumptions. In the game, each player $i \in \{1, 2\}$ can decide between two actions $\{c, d\} \in A$; each can either provide effort c or not provide effort d to the production of one unit of a collective good. For each unit produced, both players receive a benefit equal to $v > 0$, and if zero units are produced they both get zero benefits. The cost to each player of providing the collective good is given by the cost function $C(e)$, where e equals the amount of effort provided. If only one player provides the good then she has to provide one unit of effort to produce the one unit of the good, that is $e = 1$, and thus pays $C(1)$. On the other hand, if both players provide the collective good then they affect each others productivity so that each player has to provide either less or more effort per produced unit of the collective good,[34] and therefore, each pays $C(\hat{e})$ where $\hat{e} \in (0, \infty)$. Lastly, if a

[32] This assumption is described in footnote 4.

[33] For good overviews of the critiques of Olson's model, see Oliver (1993) or Udéhn (1993).

[34] Such externalities are very common and can arise from situations in which individuals help (hinder) each other when working together in a group. This effect can be seen as an additional input in the production of the good so that additional workers make other workers more (less) efficient. Of course, it can be assumed that the

player does not provide any effort to the production of the good, $e = 0$, then it faces zero costs, $C(0) = 0$. In addition, two more assumptions are made on the nature of $C(e)$: first it is assumed that costs are always positive for all $e > 0$, and second, costs increase with the level of effort,[35] $C'(e) > 0$. At this point, the game can be represented using a payoff matrix (for simplicity only the payoffs of the row player are shown),

Table 1: Payoff matrix for the row player

	c	D
C	2v – C(ê)	v – C(1)
d	v	0

Using the assumptions about *v* and *C(e)*, it is easily shown that only eight distinct games can arise from the above payoff matrix.[36] However, the eight games can be grouped into five different classes depending on the shape of the cost function. Each class has its own special characteristics representing different types of collective goods. The five classes are:

Prohibitive cost: Two of the eight distinct games fit into this class. They arise when the cost of producing the collective good is so high that playing *c* always gives a negative payoff, $v < ½C(ê)$ and $v < C(1)$. In this case, there is no dilemma at all. Both players have a dominant strategy to play *d*, which in turn gives the socially desirable outcome.

Prisoner's dilemma: This game, which was previously discussed, arises when the cost of producing one unit of the good is still higher than the utility derived from consuming one unit of the good, $v < C(ê) < C(1)$. However, the cost is no longer prohibitive if both players produce the collective good, and thus are able to consume twice the number of units, $v > ½C(ê)$. Thus, the prisoner's dilemma game is more likely to present itself when costs are relatively high and the effect of the players on each other's productivity is small.

Assurance game: This game presents itself when the cost of producing the good alone is still high, $v < C(1)$. However, when working together each player produces a significant positive effect on the other's productivity, $ê < 1$. Thus, producing the good together provides them

---

marginal benefits of help decrease with the group size, however, no such assumption is necessary in the present analysis.

[35] Note however that no assumption is made of the second derivative of *C(e)* and thus it is possible to get numerous types of cost functions.

[36] There are 4! = 24 possible distinct symmetric 2 × 2 games (distinct meaning games in which all payoffs differ). However, the initial assumptions guarantee that $\pi_i(d_i, c_j) > \pi_i(c_i, d_j)$ and $\pi_i(d_i, c_j) > \pi_i(d_i, d_j)$, and thus, only 8 distinct games remain. In his paper, Heckathorn (1996) uses an explicit form for the production function that is equivalent to imposing the condition $v > c(ê) – c(1)$. Thus, since his setup is slightly more restrictive, he gets only five distinct games.

considerable gains, $v > C(\hat{e})$. What distinguishes the assurance game is that it has two Nash equilibria in pure strategies: one in which both players choose *c*, giving them a high payoff, and one in which both players choose *d*, giving them a low payoff. Thus, the focus of this game is coordination. Both players would prefer to choose *c* but without a signaling device, they cannot be certain what the other player will decide on. Furthermore, it is important to note that as the effect on each other's productivity grows, that is $C(1) - C(\hat{e})$ becomes larger, one's risk of choosing *c* while the other chooses *d* increases. Therefore, even though coordination on *c* yields a higher payoff, if there is a lot of uncertainty on the other player's action, an individual could find playing *d* more attractive.[37]

Chicken-type games: There are three games within this class. They arise when the cost of producing the collective good is low, $v > C(1)$. However, when working together each player produces a significant negative effect on the other's productivity, $\hat{e} > 1$. Consequently, working together generates considerable losses, $v < C(\hat{e})$. What characterizes the games in this class is that players wish to choose opposite actions: if player *1* chooses *c* then player *2* wants to choose *d* and vice versa. However, even though both agree that someone defecting and someone cooperating is the best outcome, both want to be the one that defects. The focus of the game is actually bargaining; there is agreement on which is the desired outcome but disagreement on the way to achieve it (not a big surprise since $\hat{e} > 1$ already indicates that the players hinder each other when working together).

Privileged game: This game, as the games in the "prohibitive costs" class, presents no conflict. It arises when costs are so low that each of the players would find it profitable to produce the collective good no matter what is the other's choice, $v > \frac{1}{2}C(\hat{e})$ and $v > C(1)$, a situation that could be considered analogous to Olson's description of the privileged group. Again, there is no dilemma. Both players have a dominant strategy to play *c*, resulting in the socially desirable outcome.

In this manner, it is straightforward to see that although the prisoner's dilemma and its free riding characteristic are definitely present in some collective goods, it is only one of the situations that can arise. Of the five different classes of games described above, only three are actually of interest. The first and last classes, in which there is no dilemma, are trivial cases that do no require much attention. However, the other three classes can be used to understand differences between collective goods.[38] We should note that the basic difference between

[37] For more details on the coordination problem in the assurance game see §3.5.

[38] Note that only these games arise because this is a 2-player game. If we expand the analysis to n-player games then many more cases might come up. Schelling (1973) provides a few examples of the situations that present

them is the magnitude of the effect caused by adding people to the production process. If the effect is large and positive then an assurance game is more likely, if the effect is small, a prisoner's dilemma arises, and if the effect is large and negative, a chicken type of game is the result. The literature has generally concentrated on two of these games, namely the assurance and the prisoner's dilemma. Authors that refer to collective goods, such as mass demonstrations, often use an assurance game kind of setup, and argue that free riding is neither a serious limitation for collective good provision nor the most important feature of collective action. On the other hand, authors that refer to collective goods such as lobbying or solving neighborhood problems generally emphasize free riding and think of coordination as a minor problem. Unfortunately, to the best of my knowledge there is no systematic empirical study that attempts to measure the shape of the production functions and thus establish which kind of good is more common and deserves more attention.

Overall, all the three methods described here, repetition, institutions and production technology, have provided us insights on the workings of cooperation. From this work, it is clear that, although Olson's model is insightful, it might not be complete enough to capture the essence of cooperation. Furthermore, work along this lines has taught us when is cooperation more likely to occur. Namely, when interaction is not a one-time affair, when viable institutions are put in place and lastly when working together produces large positive externalities.

However, a satisfactory explanation for an underlying ability to cooperate that seems to arise in many settings is still not provided. Perhaps, the clearest indications that these theories are incomplete are the numerous experimental results of public games. From these experiments, it is clear that people do cooperate and in some cases considerably so. Hence, in situations where free riding is definitely present, where no formal institutions exist and where repetition is either clearly finite or non-existent, collective goods are still provided (Roth, 1995). In other words, even in Olson's basic model individuals cooperate noticeably above predictions. Thus, it is this author's opinion that we still should not dismiss Olson's model as being too simple. A deeper understanding of what motivates people to cooperate in such settings would provide us a strong base from which to build more complicated scenarios.

---

themselves when larger n-player games are considered. The author draws attention to cases where the optimal provision of the good is attained with the participation of most but not all individuals. Such situations can be very difficult to solve since solutions such as strict norms that sanction non-cooperators would lead to overprovision of the collective good. Vaccination is a good example of such a case.

This brings us to the next approach, which attempts to understand cooperation as a result of different types of preferences.

3.4 Changing the utility function of individuals, allowing for social preferences

In a similar manner to the last approach, the focus of these studies has been tackling the free rider problem. However, in this case, it is acknowledged that sometimes people seem able to overcome a clear free rider problem without the use of formal institutions or the effects of sustained interaction. Many of these authors reject the common assumption that people care only for their own welfare. They argue that cooperation can be explained with the use of more complex individuals whose preferences include social components such as status, fairness, morality, or even the well being of others (Fehr and Fischbacher, 2002). With the use of this kind of agents, cooperation becomes much more likely.

To illustrate the effect these kind of social preferences have on the provision of collective goods, we can use a model proposed by Holländer (1990). In his model, Holländer proposes that besides valuing the benefits of the collective good and the costs of producing it, individuals also value social status or recognition. Thus, contributing to the collective good serves two purposes. On one hand, it increases the amount of the collective good, and on the other it gives the contributor extra benefits by raising her social status. The model is formalized as the following maximization problem,

$$\max_{\gamma_i} v_i(\Gamma) - C_i(\gamma_i) + S_i(\gamma_i, \tfrac{1}{n}\Gamma) = v_i\left(\sum_j \gamma_j\right) - C(\gamma_i) + S_i\left(\gamma_i - \tfrac{\sigma}{n}\sum_j \gamma_j\right)$$
$$s.t.\ v_i(\Gamma) + S_i(\gamma_i, \tfrac{1}{n}\Gamma) > C_i(\gamma_i)$$

Where $S_i(.)$ is the function that represents the satisfaction i receives due to social status. It is assumed that $S'(.) > 0$ and $S''(.) < 0$. Note that social status depends not only on i's contribution γ_i, but also on the average contribution of the group $\tfrac{1}{n}\Gamma$. The variable σ measures how important are relative contributions for the acquirement of social status.[39] The resulting best response of individual i is given by the following set of equations,

$$\begin{cases} \dfrac{dv_i(\Gamma)}{d\gamma_i} + \left(1 - \dfrac{\sigma}{n}\right) \cdot \dfrac{dS_i(\gamma_i^*, \tfrac{1}{n}\Gamma)}{d\gamma_i} = \dfrac{dC_i(\gamma_i^*)}{d\gamma_i} & if \quad v_i(\Gamma) + S_i(\gamma_i^*, \tfrac{1}{n}\Gamma) > C_i(\gamma_i^*) \\ \gamma_i^* = 0 & if \quad v_i(\Gamma) + S_i(\gamma_i^*, \tfrac{1}{n}\Gamma) \le C_i(\gamma_i^*) \end{cases}$$

39 Note that just as other models in which utility is relative to the average, this creates a "catching up with the Jones" effect, wherein, an individual tries to increase her utility by contributing more than others, but this in turn forces the others to increase their contribution in order not to loose utility, forcing the individual to further increase her contribution and so on.

If the group is very large then we can make the simplifying assumption that your own contribution has no noticeable effect on the total amount of the collective good. In other words, we can take Γ as a constant. In this case, we arrive to the following best response,

$$\begin{cases} S_i'\left(\gamma_i^* - \sigma/_n \Gamma\right) = C_i'\left(\gamma_i^*\right) & if \quad v_i(\Gamma) + S_i\left(\gamma_i^*, 1/_n \Gamma\right) > C_i\left(\gamma_i^*\right) \\ \gamma_i^* = 0 & if \quad v_i(\Gamma) + S_i\left(\gamma_i^*, 1/_n \Gamma\right) \leq C_i\left(\gamma_i^*\right) \end{cases}$$

Thus, thanks to social recognition, contributions to the collective good might indeed exist when in its absence there would been none. Furthermore, if we use again the interpretation of the cost function as the opportunity cost of consuming the collective good, then, we can note that individual contributions to the collective good will increase whenever: the income of the individual increases, *C'(.)* decreases for all $\gamma_i > 0$; the average contribution increases, $1/_n \Gamma$ goes up; the importance of relative contributions increases, σ goes up; or when the satisfaction received from social status becomes more important, that is *S'(.)* increases for all $\gamma_i > 0$. This model also implies that, since contributions are now only made to gain status, governmental intervention will not have a "crowding out" effect on the private provision of the collective good (that is of course, as long as the fact that the government also provides the good does not affect the social status function) (Andreoni 1990).

Using, a similar argument other authors have used this kind of setup to explain cooperation as a result of different kinds of intrinsic motivations.[40] Elster (1985) provides an interesting breakdown of the various reasons different people might have for participating in collective action. He includes irrational beliefs, selfish "warm glow" feelings, altruistic as well as moral motivations, and compliance with social norms. More recently, models that take into account fairness seem to provide not only a theoretically plausible explanation to cooperation but also do particularly well explaining experimental results (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999; Falk and Fischbacher,1999).

This brings us to another important characteristic of this approach, the modeling of agent heterogeneity. The recognition that some people will value the collective good more than others is definitely not ignored in previous approaches. However, the analysis of the effects of different mixes of social preferences on the provision of collective goods had not been properly studied. The need to take into account heterogeneous types of individuals came out of the experimental results accumulated during the last twenty years. As mentioned earlier,

---

[40] Some examples include pure and "selfish" altruism (Andreoni, 1989 and 1990), duty and solidarity (Barry, 1978), sympathy and commitment (Sen, 1979), morality (Elster, 1989), social control (Heckathorn, 1990), and when it comes to voting, a sense of civic duty (Knack, 1992; Aldrich, 1997).

this evidence clearly demonstrates that in many cases cooperation does not go to extremes, full cooperation as well as total defection is rare, while intermediate ranges are more common (Ledyard, 1995). Furthermore, a closer look at the data demonstrates that some people consistently contribute more than others (Schram, 2000; Fehr, Fischbacher and Gächter, 2001). The study of different types of individuals became necessary in order to explain these results.

On the surface, it might seem farfetched to think that modeling different types of individuals would dramatically change the collective action model. After all, in equilibrium in for example Holländer's (1990) model, if there is a pair of individuals *i* and *j* for whom it holds that $S_i'(\gamma_i^*, \frac{1}{n}\Gamma^*) > S_j'(\gamma_j^*, \frac{1}{n}\Gamma^*)$, then, for the same cost function $C_i(.) = C_j(.)$, *i* will contribute more than *j*. However, individuals might have more complex preferences that shatter these simple results. A good example is individuals that are now commonly know as conditional cooperators. This type of individuals wishes to contribute only if others also contribute. In other words, they do not like to free ride but they do not like when others free ride at their expense. If this type of individuals are present, then the common assumption made so far $d\gamma_j / d\gamma_i = 0$, cannot be used anymore. If we drop this assumption then previous predictions change and situations emerge where even individuals who do not mind free riding might contribute considerably to the collective good (Kreps et al., 1982).

To conclude, different intrinsic motivations are important variables for the explanation of collective action. However, much more research must be done in order to assess their significance, and in addition, to identify which motivations play the more important roles. In fact, common criticisms to this approach are precisely these two points. To begin with, since these models rely on intrinsic motivations, things which cannot not be easily observed, it is no surprise that empirical data to support these theories is hard to obtain.[41] Secondly, without reliable data it is quite easy to come up with many possible forms of utility functions that support some degree of collective action. Which is why, if there is little empirical proof, it is important to explain not only how preferences for a specific type of characteristic can give rise to collective action, but also why people posses such preferences. In other words, why did social preferences evolve? For example, if altruism is proposed as an important variable for collective good provision, then we need to know how and specially under what conditions can altruists survive the competition of selfish individuals who free ride on their kind

[41] Some progress is being made in this aspect with the careful use of self-reports (Bosman and van Winden, 2002) and in some cases physiologic readings (Lo and Repin, 2002).

behavior. Without this additional understanding, intrinsic motivations will provide only an incomplete answer. In the next approach, precisely these questions are addressed.

3.5 Agent types and evolutionary models

One of the main limitations of the models presented so far is that they fail to explain how cooperation evolved. As we have seen, many models explain collective action by assuming the existence of specific types of institutions or social preferences.[42] Unfortunately, although these models can clarify why we observe some instances of collective action, they do not explain how cooperation emerged. Without this understanding, the recommendations these models can provide are rather limited, especially if our aim is to promote collective good provision. After all, if there is a case in which there is no collective action, these theories would require a third party to impose an institution or to promote the adoption of social preferences. Naturally, in many cases this is very difficult or even impossible. Understanding what are the necessary conditions for a group to engage in collective action on its own would be of tremendous value. With the use of evolutionary and learning models, we have begun to understand precisely what these conditions are.

On the surface, free riding dooms cooperative individuals to eventual extinction. If we have a group of only cooperators, then an individual who decides to defect would receive a higher payoff. Since in evolutionary terms a higher payoff means more chances of survival, eventually the whole group would consist of free riders.[43] Cooperation will survive only when the group is able to overcome this problem. In the following paragraphs, two different methods of achieving this are presented. In the first, cooperation is sustained by repeated interaction and in the second by the emergence of social preferences.

As it has been previously mentioned, repeated interaction of a 2-player prisoner's dilemma game can present equilibria with cooperative outcomes. Regrettably, game theory provides us a huge set of possible outcomes and strategies. Furthermore, whenever players know when the game is going to end, backward induction predicts all players defecting. Axelrod (1984) proposes a way of overcoming these two problems. First, he argues for the use of finite automata, that is, players who react only on the history of the game (this solves the backward induction prediction), and second, he proposes to have different finite automata

---

[42] Of course, these assumptions are not based on thin air. The authors base their assumptions on institutions or preferences that we know exist in our society. However, they do assume that these institutions or preferences are indeed key factors for collective action.

[43] This is confirmed by the fact that defection is the only evolutionary stable equilibrium in the prisoner's dilemma game.

compete with each other to see which is more successful (this gives us a clear prediction). As it turns out, his initial studies seem to indicate first, that cooperative automata survive competition from free riders, and second, the most successful finite automaton was in fact a very simple one, namely tit-for-tat.[44] Unfortunately, further research has shown that this outcome is not stable. Tit-for-tat does very well against strategies that defect; however, it has no special advantage against strategies that cooperate. This fact makes tit-for-tat and full cooperation an unstable outcome. To see this, imagine a situation in which everybody in a group is using a tit-for-tat strategy. If a player decides to switch to the strategy always-cooperate then she will receive the same payoff as the other group members, and therefore she has the same chance of survival as the rest. If by chance, the number of individuals playing always-cooperate increases over a certain threshold, then individuals that start playing always-defect can do even better that individuals playing tit-for-tat, and hence, cooperation can break down. For this reason, we have an unstable outcome. In fact, the instability just illustrated is part of a more general result demonstrated by Binmore and Samuelson (1992). The authors actually show that no single finite automata can be an evolutionary stable strategy. Nevertheless, even though we arrive to such a pessimistic result, it might still be useful to analyze the common characteristics of the strategies that generally lead to successful collective action. Axelrod (1997) does exactly this and points out two important characteristics: on one hand, a strategy must punish defectors by defecting, but on the other, it must also forgive past defections to return to cooperation.

Another application of evolutionary theory has been the explanation of social preferences. As we have seen, some types of social preferences have been used to explain collective action. However, such explanations generally do not describe how these specific kinds of preferences evolve. To fill in this gap, research has been conducted to see if indeed social preferences can account for collective good provision.[45]

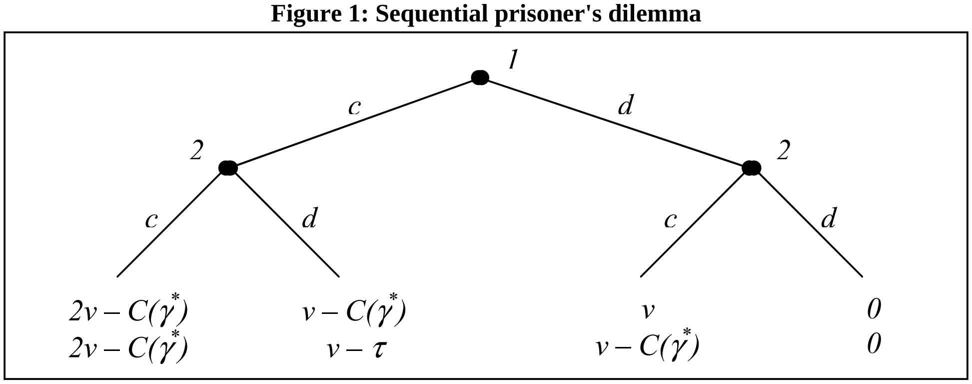
To illustrate how social preferences can evolve we can use a model proposed by Güth and Kliemt (2000). In their paper, the authors provide a good indication of how and under what circumstances trustworthiness can facilitate cooperation. To accomplish this Güth and Kliemt

[44] See footnote 29 for a description of the characteristics of tit-for-tat.

[45] It might be argued that, it should not be necessary to justify the existence of preferences that are clearly present in our societies. After all, nobody can seriously question the fact that people care for things such as morality or fairness. However, just because we can observe that people are concerned about fairness does not automatically mean that fairness evolved as a way of overcoming free riding in collective action. Policy recommendations could be terribly wrong if they are based on a characteristic that was not designed for that purpose.

use what is referred as the "indirect evolutionary approach". As in standard evolutionary models, there are a large number of individuals who are randomly matched to play a stage game. The payoffs obtained in the stage game determine the evolutionary success of the player so that the population of more successful players grows as more games are played. However, unlike the normal evolutionary setup, Güth and Kliemt make the distinction between objective and subjective payoffs. Objective payoffs are the ones that determine the evolutionary success, while subjective payoffs are the ones used to determine behavior. In this way, we can model social preferences, such as a preference for trustworthiness, which affect behavior but do not provide per se a higher chance of survival. Furthermore, it is assumed that players are fully rational, that is, in each game they will use the strategy that gives them the highest subjective payoff. This assumption narrows down the number of strategies we must consider and this makes the game easier to analyze.[46] To study the evolution of trustworthiness we can consider a stage game that consists of a sequential prisoner's dilemma, see .

Figure 1: Sequential prisoner's dilemma



This game is practically the same as the normal prisoner's dilemma, the only difference being that player *2* can observe the action of player *1* before making her choice. The structure of the game can be easily discerned in . However, a few things need to be clarified. First is the variable $\tau \in \{\tau_L, \tau_H\}$ where $\tau_L < \tau_H$.[47] This variable represents the utility loss experienced by player *2* whenever she breaks the trust of player *1*, that is, when player *2* observes player *1* cooperates and nevertheless she decides to defect. Individuals for whom $\tau = \tau_H$ can be considered to have a preference for trustworthiness (hereafter referred to as the trustworthy type), while the converse being true for individuals for whom $\tau = \tau_L$ (untrustworthy type). However, as was mentioned above, since social preferences cannot be said to affect survival

---

[46] For more details of the "indirect evolutionary approach" see Güth and Kliemt (1998), or Königstein and Müller (2000).

[47] See Güth and Kliemt (2000) for the proof of why it is sufficient to consider only two values of τ.

chances, τ affects only subjective payoffs. All other variables affect both subjective as well as objective payoffs. Lastly, note that for this game to be a prisoner's dilemma we must assume that $C(\gamma^*) > v$.

In order to see how trustworthiness evolves we can examine the case in which trustworthy and untrustworthy individuals make different decisions whenever they are in the position of player *2*. Specifically, trustworthy individuals will cooperate if player *1* cooperates, that is, $v + \tau_H > C(\gamma^*)$, while for untrustworthy individuals the converse holds, $v + \tau_L < C(\gamma^*)$. The intuition behind the main results of the model can be observed examining two different cases:

The type of individuals is observable: In this case it is clear that whenever player *1* observes a trustworthy type she will decide to cooperate, while, if she observes an untrustworthy type she will choose to defect. Thus, trustworthy types get a higher objective payoff than untrustworthy types and eventually take over the whole population. The outcome in this case is full cooperation.

The type of the individual is not observable: In this case the only information available is the population share of each type. This means player *1* will decide to cooperate only if $\mu(2v - C(\gamma^*)) + (1-\mu)(v - C(\gamma^*)) > 0 \Rightarrow \mu > C(\gamma^*)/v - 1$ where μ equals the share of the population who are trustworthy types. If μ indeed satisfies this inequality then individuals in player *1*'s position will choose to cooperate, and since free riding pays more than cooperation, untrustworthy types will get a higher payoff than trustworthy types. Thus, μ would decrease until it no longer satisfies the inequality. If μ does not satisfy the inequality, player *1* will always choose to defect. Logically, this is followed by a defection of player *2* irrespective of which type she is. Hence, although there might be some surviving trustworthy types the outcome is full defection.[48]

Thus, we can identify the two extremes of the model and the underlying intuition. Trustworthy players benefit from successfully signaling their type while untrustworthy players benefit from being mistaken as trustworthy. In a less extreme environment than the two cases mentioned here, Güth and Kliemt show that when individuals can buy a noisy signal of the other player's type then three outcomes can occur. First, if the signal is either too costly or too inaccurate, only untrustworthy types survive. Second, if the signal is not too costly and accurate enough and in addition, the initial population of trustworthy types is large, both types of individuals will coexist with stable shares of the population, and third,

---

[48] If we allow for mistakes so that player *1* sometimes chooses to cooperate even though she intended to defect, then untrustworthy types can benefit from time to time from the free rider payoff and slowly drive trustworthy types to extinction.

only when the signal is free and perfectly accurate trustworthy types will come to dominate the whole population. In conclusion, this model helps us understand how a preference for trustworthiness can lead to cooperation by making individuals feel bad if they choose to free ride. However, this is true only if individuals have a way of credibly signaling what type of preferences they posses.[49] Other authors have performed similar analyses that evaluate the evolutionary stability of different kinds of institutions and social preferences.[50]

Although, most of the research in this approach has concentrated on overcoming the free rider problem, there is some literature that has focused on the second difficulty faced in collective action, that is to say, coordination. As we have mentioned, coordination can be analyzed using the assurance game, where the only problem faced by the players is to coordinate in either one of the two Nash equilibria. Since the cooperative equilibrium is Pareto superior, it is reasonable to assume that both players would prefer that outcome. However, cooperating is still risky since if for some reason the other player defects, one is left with the lowest possible payoff. Hence, if there is uncertainty of the other player's action it might be better to defect. For example, if we consider the game already described in Table 1, we can see that the row player's expected payoff of choosing *c* is $\mu(2v - C(\hat{e})) + (1 - \mu)(v - C(1))$ where μ is the row player's belief that the column player will choose *c*. Similarly, the row player's expected payoff of choosing *d* is equal to μv. Noting that the row player will choose the action that gives her the highest expected payoff, she will play *c* only when $\mu > C(1) - v / C(1) - C(\hat{e})$. Hence, as long as this is an assurance game, that is if $C(1) > v > C(\hat{e})$ holds, there will always be a feasible value for μ for which the inequality holds. In other words, we can never rule out defection as the equilibrium outcome. Even in an evolutionary setting, defection is a stable outcome if the population of individuals defecting is large enough. In fact, we get a more pessimistic result if in addition $2v < C(1) + C(\hat{e})$ is true. In this case, defecting is the risk-dominant equilibrium as defined by Harsanyi and Selten (1988). It basically means that $\mu > ½$ and thus, the row player will cooperate only if she believes that more than half the population is also cooperating. Hence, if we allow players to make mistakes, it takes a smaller number of mistakes for a population to switch from full

---

[49] Frank (1987) proposes that emotions such as guilt can cause observable involuntary reactions (like blushing) that can serve as a cheap and reliable signal of ones preferences.

[50] To name a few examples, Cárdenas and Ostrom (2001) examine the evolution of cooperative norms (see also Witt, 1986), Heckathorn (1993) studies the effects of income heterogeneity, Bester and Güth (1998) explore the stability of altruistic behavior, Macy (1993) tests the performance of various sanctioning systems, and Flache and Macy (2002) use learning models to explain, among other things, how the type of aspiration levels affect peoples ability to learn to cooperate (for more learning models see Macy, 1990 and 1991b).

cooperation to full defection than vice versa, and therefore, full defection is more likely to occur in the long run.

Thankfully, the coordination problem can be easily solved if we simply allow players to communicate before the game (Blume, Kim and Sobel, 1993).[51] This can be easily illustrated with an example. First, we change the game slightly so that the players get the opportunity to send each other one message $\phi \in \Phi$, where Φ is the finite set of possible messages, and thus, players can condition their beliefs μ on the message they receive, that is, $\mu(\phi)\colon \Phi \rightarrow [0,1]$. Lets assume that there is an initial population of individuals who all believe their opponent will always play d with certainty, that is, $\mu(\phi) = 0$ for all $\phi \in \Phi$. Now, we introduce a group of individuals who decide to send always one previously unused message $\phi_u \in \Phi$[52] and who condition their beliefs in the following way, $\mu(\phi) = 0$ for all $\phi \neq \phi_u$ and $\mu(\phi) = 1$ for $\phi = \phi_u$. These new individuals will defect whenever they encounter members from the old group and cooperate whenever they play against members of their own group. Therefore, they will on average obtain a higher payoff and eventually drive the defecting individuals into extinction. Hence, by using a simple message, players can remove any uncertainty of the other's action and thus arrive to the Pareto superior equilibrium.[53]

Overall, using evolutionary as well as learning models have considerably expanded our understating of collective action. Nonetheless, evolutionary models do have certain drawbacks. Perhaps the most important limitation is that analyzing complex situations in an evolutionary setting is sometimes extremely hard and thus research tends to be limited to very simple models. A consequence of this is that, more often than not, the resulting equilibrium presents either full cooperation of full defection. A more complex game is needed in order to arrive to equilibria that are more realistic.[54] Another limitation of this approach is that players are matched randomly, while in reality we usually have a say with whom we decide to interact. Once we allow individuals the opportunity to choose whom to associate with, networks of ties will emerge, and as will be seen in the next approach, they might considerably affect collective action.

---

[51] Unfortunately, the same does not apply to the prisoner's dilemma game. See Farrell and Rabin (1996) for a discussion on the role of communication in this setting.

[52] By unused it is meant that individuals who do not belong to the new group will never use $\hat{s}$. If we think of messages as spoken words then this condition is easy to satisfy since one can simply come up with a "secret" word or password that nobody else outside group understands.

[53] For more details and replicator dynamics, see Weibull (1995).

[54] For further discussion see van Veelen (2002) where the author argues that repeated stochastic games are a much better model for human interaction.

3.6 Networks, information transmission, and bandwagon effects

A common limitation of the previous models is that they neglect to take into consideration the fact that in many situations people can choose with whom to interact. It is rather obvious that this choice is not random. Generally, people prefer to interact with people who are similar to them, and collective action is no exception. Empirical work has demonstrated that individuals who participate in collective action have more links to other participants than individuals who do not participate (Opp, 1989). If links were created randomly, this would not be the case. Hence, network structure becomes an important factor to take into consideration. As was demonstrated by Gould (1993), the type of network and the position of individuals in it determine, to a great extent, if collective action is successful or not.[55] When it comes to theoretical work, a substantial part of the research in this approach has concentrated on the coordination problem of collective action. A couple of examples, which illustrate two significant segments of this literature, follow in the next few paragraphs.

One of the most important results originating from the analysis of networks is the importance of costly links. The basic idea behind it is that individual choice of action might be conditioned by the cost of linking. In some cases, linking costs help collective action while in others they complicate matters. If for example, the cost of making new links is bigger than the cost of keeping the current ones, then an individual might find herself trapped in a bad outcome and unable to move to a better one. In order to elaborate we first introduce some notation. We are modeling a situation in which each individual can potentially make a link with any other individual in the group. Formally, let $G_i = \{g_{i1}, \ldots, g_{in}\}$ be the set of links formed by individual i. Furthermore, let $g_{ij} = 1$ if individual i made a link to individual j and $g_{ij} = 0$ otherwise.[56] Finally, the network $G = \{G_1, \ldots, G_n\}$ is simply the profile of link decisions of every individual. The game is constructed as follows, each individual i can attempt to make a link to any individual $j \neq i$, who can either accept or reject the link. Each link that i tries to establish costs her an amount $k > 0$. If a link is established between i and j, in other words if $max\{g_{ij}, g_{ji}\} = 1$, then they both play with each other a 2-player assurance game equal to the one that has been previously described.[57] If no link is established between i and j, then each individual gets a payoff $q \in \Re$ for not playing the game. Furthermore, we

[55] For a survey on the earlier work on the importance of networks and collective action, see Macy and Flache (1995).

[56] Since an individual does not need to form a link to herself, g_{ii} will always equal zero.

[57] Note that in this case links are one-sided. That is, for the game to be played only one individual has to make the link. Of course, this also means that only one of the individuals pays the cost. For similar models but with two-sided links see Droste, Gilles and Johnson (2000), and Jackson and Watts (2002).

assume that once i contacts j to form a link, j can observe i's choice of action $A_i \in \{c, d\}$. Consequently, j will accept to link to i only if she can get a higher payoff from playing the game than from not playing it; formally, for the given A_i, $\pi_j(A_i, A_j) > q$ for at least one $A_j \in \{c, d\}$. Finally, every individual i must choose the same action for all the 2-player games she participates in.

Goyal and Vega-Redondo (2001) analyze the characteristics of this model's equilibria. Unsurprisingly, they find that the equilibrium network and action depend on the cost of linking k and the value of the outside option q. They identify three distinct cases:

Prohibitive linking costs: This case arises if the outside option is higher than the highest payoff of the game minus the cost of making a link, that is, if $q > 2v - C(\hat{e}) - k$. If this is the case, then it is obvious that nobody will propose any links and the equilibrium outcome will be an empty network. That is, $g_{ij} = 0 \ \forall i, j \in N$ and each player receives a payoff equal to nq.

Separating linking costs: In this case, the value of the outside option is lower, so that, if individuals link to play the cooperative equilibrium, they receive a higher payoff, $q < 2v - c(\hat{e}) - k$. However, the out side option is high enough so that it is better not to play than to link and play the defecting equilibrium, $q > -k$. Here, the linking costs have a possibly positive effect. Since, it is only worth linking to play the cooperative equilibrium, then if individuals link there is no longer any uncertainty of which action they will take. Thus, we either get a complete network where everybody cooperates, that is $max\{g_{ij}, g_{ji}\} = 1$ $\forall i, j \in N$ and the payoff of each individual is $n(2v - C(\hat{e}) - ½k)$, or we get an empty network and everybody defects.[58] Note that, in the case of cooperation, in equilibrium the links are roughly evenly distributed, so that, on average an individual has to pay only $½k$ for each game she plays.

Pooling linking costs: This last case occurs when the outside option is lower than the payoff of the defecting equilibrium, $q < -k$. Here, the linking costs are not enough to differentiate between individuals choosing c and individuals choosing d. Thus, we can get any two of the equilibria. However, depending on the exact value of q, we can arrive to different situations. If $q < v - C(1)$ then the equilibrium network is complete and all individuals coordinate on the same action, but if $q > v - C(1)$ then, in addition, there exist equilibria in which the network is partitioned into two complete subcomponents where one component is cooperating while the other is defecting. This difference is due to the role of passive links. When $q < v - C(1)$, the outside option is so low that all individuals will accept any link irrespective of the action

[58] An unlikely outcome given it is so easy to signal one wants to cooperate, but still it cannot be discarded without further refinements.

of their opponent. Thus, at the end, everybody is interconnected and individuals will choose the action that is more popular. Specifically, if n^A is the number of other individuals choosing action A, then everybody will switch to cooperate if $n^c > n^d(C(1) - v)(v - C(\hat{e}))^{-1}$. In contrast, when $q < v - C(1)$, an individual who is playing c will neither make nor accept links from individuals that are playing d. Thus, separate clusters each playing a different action will form, and if they are roughly the same size, no single individual will find it profitable to switch to another cluster. This is a striking finding since it means that, even when individuals are identical, there is a possible outcome where a group of them earns a substantially greater amount than the rest. We can use a specific case to see why such an outcome can occur. Suppose there are two clusters of individuals, each cluster playing a different action. Furthermore, imagine that all the individuals in one cluster are connected with each other. In this situation, if individual i is a member of the cluster playing d it means she currently gets a payoff equal to $n^d(-\frac{1}{2}k) + n^c q$.[59] If i wants to move to the cluster playing c, then the best she can do is destroy all the links she has now and create a new link with everyone in the other cluster. This would give her a payoff of $n^c(2v - C(\hat{e}) - k) + n^d q$. Comparing the two payoffs, we can conclude that individual i will stay in her current situation if n^d is significantly bigger than n^c, that is when $n^d - n^c > 2v - C(\hat{e}) - \frac{1}{2}k$. Doing the same exercise for a member j of the cluster playing c, who whishes to switch to the cluster playing d, gives us that j will stay in her current situation if the difference between n^d and n^c is sufficiently small, that is, if $2v - C(\hat{e}) + \frac{1}{2}k > n^d - n^c$. Combining these two results, we can deduce that, for large n, there is always a range of stable cluster sizes. Furthermore, the bigger the cost of linking the larger this range will be. In other words, the fact an individual is not paying for all the links in her cluster makes switching clusters much more difficult.

A second important result in the network literature on collective action is the study of which types of network structures facilitate or hinder the provision of collective goods. Since producing a collective good involves more than one individual, the information each possesses (number of people known, their willingness to cooperate, etc.) will considerably affect their decision. This information is especially relevant in the case of coordination problems where individuals do not have a single action that constitutes a dominant strategy. Hence, studying how this information flows within a group can help us understand why groups with similar individuals can end up taking very dissimilar actions.

In order to understand how network structure determines collective action, it is important to know what basic functions networks fulfill. In an assurance game setting, Chwe (2000)

[59] Again, note that on average links are distributed equally so that the average cost of linking is only $\frac{1}{2}k$.

gives an excellent analysis of one such function, namely, the distribution of common knowledge. To demonstrate his point the author presents the following model. As in the previous setups, each individual $i \in N$ chooses an action $A_i \in \{c, d\}$. However, in this case individuals are further differentiated into two types depending on their cost function; formally, each i has a cost function $C_i(e) \in \{C_L(e), C_H(e)\}$ where $C_L(e) < C_H(e)$ for all e. Moreover, since the effort an individual has to spend to produce one unit of the collective good depends on the number of other individuals choosing c, the cost function can be written as $C_i(A_i, n^c)$ where $C_i(d, n^c) = 0$ and $C_i(c, n^c) < C_i(c, n^{c}')$ implies $n^c > n^{c}'$. To continue, if i chooses to cooperate, her payoff is equal to $(n^c + 1)v - C_i(c, n^c)$, and similarly, if she chooses to defect, her payoff is equal to $n^c v$. Hence, i chooses c only when she believes that enough people will cooperate so that $v > C_i(c, n^c)$. Finally, assume that high cost types choose d under any circumstances, that is, $v < C_H(c, n^c)$ for all $n^c \geq 0$. The communication network G is defined as above, however, in this case links are used only to acquire information. We assume that if $g_{ij} = 1$ then i knows what is j's type, and in addition, i knows with whom j has formed links.[60] Now, we will call G a sufficient network for collective action if its structure is such that all low cost individuals choose c even when they believe they are in the worst possible scenario (given their information). By worst possible scenario it is meant that every i will assume that every j for whom $g_{ij} = 0$ is of the high cost type.[61] In other words, i will choose c only if she knows that $v > C_L(c, n^c)$ is true. Finally, we will call G a minimal sufficient network for collective action if G and G' are sufficient networks, and $G \supset G'$ implies $G = G'$.

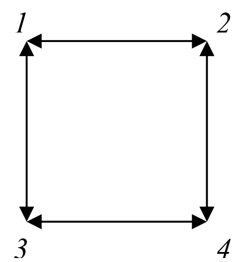
As it turns out, minimal sufficient networks have a common structure that highlights the importance of the distribution of common knowledge. That is, if G is a minimal sufficient network then there exist cliques $H_1, \dots, H_z$ which cover N such that the following three conditions hold: first, all individuals in a clique have links with each other, that is $g_{ij} = g_{ji} = 1$ for all $i, j \in H_v$ and $i \neq j$; second, i has a link to j only if everyone in i's clique has a link to everyone in j's clique, formally, $g_{ij} = 1$ only if there exists some H_v which contains i and H_w which contains j such that $g_{ij} = 1$ for all $i \in H_v$ and all $j \in H_w$; third, each relation among cliques is part of a chain that starts with a leading clique, that is, if $g_{ij} = 1$ for all $i \in H_v$, and

[60] Note that links are one-sided so that if $g_{ij} = 1$ but $g_{ji} = 0$ then i has information on j but j does not have information on i.

[61] This situation might seem extreme since individuals ignore what is the prior distribution of types. However, it reflects the cases where unsuccessful collective action can bring disastrous consequences, for example when failing to throw down an authoritarian government.

all $j \in H_{v-1}$, then $g_{im} = 1$ for all $m \in H_{v-2}, \ldots, H_1$.[62] The leading clique H_1 is characterized by the fact that its members do not have any links to other cliques, and at the same time, all low cost individuals in the clique choose to play *c*. Each of these conditions highlights important properties of the distribution of knowledge necessary for collective action. First, we should note that common knowledge is necessary within each clique and it is especially important in the leading clique. Let's illustrate with an example, suppose that $v > C_L(c, n^c)$ for $n^c \geq 3$. Furthermore imagine four low cost individuals linked in the following manner: all possible links are equal to one with the exception of $g_{13} = g_{31} = g_{24} = g_{42} = 0$. This makes a square network in which every individual is mutually linked with other two people (see Figure 2).

Figure 2: The square network



In this case, each individual knows that, counting herself, there are at least three low cost individuals. Thus, they know there are enough people willing to engage in collective action. However, since they do not have information on the individual across from them, none knows for certain if the others know that there are enough low cost people, and thus, none of them chooses *c*. In other words, knowing other people who want to act collectively is not enough; it must also be known that they know each other. Second, collective action needs a leading clique that possesses a relatively high amount of low cost individuals. This of course is easy to see; since the leading clique has no links to other cliques, it must posses enough low cost individuals to make them choose *c* on their own. This explains in part why people that engage in collective action tend to overestimate the number of people that are sympathetic to their cause (Finkel, Muller and Opp, 1989; Finkel and Muller, 1998). They simply have a frame of reference, namely their clique, in which there is a high concentration of people willing to engage in collective action. The third characteristic is that cliques further down the chain need to be aware of all the other cliques before them. This simply means that cliques with a relatively small amount of low cost individuals have to possess large amounts of information before they decide to play *c*. This clearly makes their participation much more unlikely.

[62] See Chwe (2000) for the proof.

As a final point, it is almost impossible to avoid noting the similarity between the structure of minimal sufficient networks and the various threshold models that produce bandwagon effects. The threshold model naturally arises when the costs of collective action are decreasing in the number of participants (as in the assurance game). In this situation, individuals with low costs start cooperating; this lowers the costs of other individuals who then start to cooperate, further lowering costs of yet other individuals, and so on. Threshold models were first proposed by Granovetter (1978) and Schelling (1978) and since then, they have received considerable attention.[63] Chwe's network model illustrates how the bandwagon effect comes up naturally in this setup and furthermore, it gives us insights on what kind of network structure is necessary for the occurrence of this phenomenon. More specifically, this model helps us understand under what conditions strong links are better than weak links for the production of collective goods. If we have three individuals, *i, j* and *m*, then $g_{ij} = g_{im} = 1$ can be considered weak links if it is highly likely that $g_{jm} = 0$. Thus, links tend to be spread out, making weak links particularly efficient for the quick transmission of information through the network. On the contrary, $g_{ij} = g_{im} = 1$ can be considered strong links if it is highly likely that $g_{jm} = 1$. Therefore, since links tend to be very concentrated, strong links are highly efficient in creating common knowledge among small groups. This leads us to the conclusion that if costs are relatively low, strong links will be enough to create leading cliques and quickly start the bandwagon effect. On the other hand, weak links are necessary for the effect to spread throughout the population. Furthermore, if costs are relatively high, weak links are necessary to create cliques that are large enough for individuals to start cooperating. Hence, we get a mixed result that indicates both types of links serve their purpose. Which type of link turns out to be more important would then depend on the costs of producing the collective good.[64]

On the whole, investigating the effects of networks on collective action is an important endeavor. As we have seen, networks are essential if we wish to understand the role of information in collective action and in addition, the possibility that successful and unsuccessful groups coexist. Moreover, other research has also hinted at other mechanisms by which networks can affect collective action. For example, link formation can be an effective way of punishing defectors by excluding them from playing (Riedl and Ule,

---

[63] A few examples are Macy (1991a), Oliver, Marwell and Teixeira (1985), and Oliver and Marwell (1988b).

[64] For more details on the weak versus strong link analysis see Chwe (1999) and Oliver and Marwell (1988b), for more on information and common knowledge and bandwagon effects see Gavious and Mizrahi (2000 and 2001).

2002),[65] and even the position of an individual in a network might make cooperation more likely if it creates a strong sense of responsibility (Yamagishi and Cook, 1993).

However, in this author's opinion, we still require a more basic understanding of collective action if we are to decipher what goes on in complicated settings such as networks. Networks help us explain many phenomena, however, they fail to shed light on why people cooperate in simpler settings. Moreover, learning more about the basic mechanism by which collective action emerges could even help considerably in the study of networks. For instance, understanding when people decide to punish, reward or forgive others, can help us figure out how links in social networks change over time; something which would be very useful since it would help us narrow down the usually large amount of Nash networks that commonly arise in most situations.[66] In the following section, this idea is elaborated more in detail.

4. Future Research: The need for Bounded Rationality and Reasoning

As we have seen throughout the last section, research in collective action has been extensive and varied. Multiple approaches have revealed many of the mechanisms used by groups to facilitate coordination and overcome the free rider problem. However, even though we have learned considerably about the nature of these problems, we still do not have a full-fledged theory that satisfactorily explains the emergence of collective action in the most basic settings. Good examples are the experimental results from both assurance-type and prisoner's dilemma-type games. These results clearly indicate that cooperation is significantly more than zero but also significantly lower than the full cooperating outcome (Ledyard, 1995). Of course, experimental evidence can always be criticized with arguments such as: subjects do not understand the game, or the experimental setting is so different that people do not behave normally. However, such arguments seriously underestimate the reasoning capacity of individuals.[67] After all, the games used are rather simple and the subjects are definitely not slow witted.[68] Moreover, even though the experimental setting is in some cases unfamiliar, so

[65] Similar analysis, but without the network framework, have been done by modifying prisoner's dilemma games so that players have the option to avoid playing the game at all, see Orbell and Dawes (1993).

[66] It might also help us come up with more accurate predictions than the ones provided by Nash and strict Nash networks. Apparently, preliminary studies with networks and social dilemmas seem to indicate that people find it extremely hard to arrive to the predicted network structures (Riedl and Ule, 2002; Falk and Kosfeld, 2003).

[67] A strange argument indeed if it is put forward by supporters of the rational expected-utility-maximizer model, in which, agents are assumed to possess unlimited reasoning capacity (Rabin, 2002).

[68] That is unless someone can seriously affirm that undergraduate and graduate university students should not be considered at least somewhat smart.

are many situations in which collective action emerges. In fact, if we manage to figure out how collective action can emerge even in these new and unfamiliar situations, we would gain a basic understanding of collective behavior that can be used later to explain cooperation in numerous settings.

It is quite clear that, if we are to consider basic collective action settings, we have to accept that the simple neoclassical agent that maximizes expected utility, uses rational expectations, and possesses selfish preferences, fails considerably as a model of human behavior. Therefore, it is not surprising that various authors strongly argue for the use of agents that are more psychologically realistic, in order to improve our models' predictions (Rabin, 1998; Camerer, 1998; Thaler, 2000).[69] More precisely, we should start to incorporate agents that present both bounded rationality (cognitive limitations) and bounded reasoning (emotions).[70]

Bounded rationality, a broadly used term, attempts to incorporate two rather obvious facts: first, humans are cognitively limited, and second, individuals attempt to choose what is best for them. In other words, individuals want to optimize their happiness or satisfaction, however, they cannot do so perfectly because of cognitive limitations. This point was made very clear by Simon (1957) when he proposed satisficing as a better theory of how individuals make choices.[71] It is rather unfortunate that the term bounded rationality is used rather loosely to describe almost any kind of non-optimizing behavior, including backward looking agents in learning and habituation models, as well as models that simply add a cognitive "cost" to complex strategies. However, as Selten (1998b) emphasizes, these are not models of bounded rationality. In the examples above, in one case agents do not make choices, they simply follow a heuristic, and in the other case, agents are still optimizing in the traditional sense, albeit an even more complicated problem since now they have an additional variable.

---

[69] The idea of incorporating psychological theory as a basis for our models of individual choice is by no means a new one. There are indications that this question preoccupied the minds of many economists throughout history. A couple of examples would be Jevons' (1911) pleasure/pain principle, and Menger's (1871) lexicographic interpretation of the satisfaction of human needs. Regrettably, this preoccupation more or less ended with the birth of the von Neumann-Morgenstern utility function, which easily relegated to the background all concerns regarding how preferences and beliefs are formed, reevaluated, and changed (Thaler, 2000).

[70] The term bounded rationality is first used by Simon (1957) to describe the limited capacity of humans to find optimal solutions. The term bounded reasoning is used by van Winden (2002) to describe how emotions limit choice by causing urges to execute a particular action.

[71] Unfortunately, perhaps due to a lack of a formal mathematical framework, other methods besides utility optimization have barely received attention.

Bounded reasoning adds one more element to the decision-making process, namely emotions. The importance of emotions for cognition has been largely ignored in economics as well as in other social sciences. Generally, emotions are considered simply as part of an individual's preferences, and thus, they are already represented by the shape of our utility function. However, recent research has indicated that this is far from the truth; emotions play a central role in decision-making.[72] More specifically, emotions play an important role by highlighting what actions might be more suitable and what outcomes the most desired.[73] Thus, people will make different choices when faced with the same situation if they are experiencing different emotions.

Regrettably, a complete theory that incorporates bounded rationality and reasoning still does not exist.[74] Nevertheless, we can still apply some of the main findings of this literature to the decision-making process individuals face in a collective action setting. In the remaining part of this section, I suggest promising paths for future research along these lines. In order to do so, I use a variation of the framework presented by Cárdenas and Ostrom (2001)[75] to divide the decision-making process into three separate parts; each part being characterized by a different set of considerations.

4.1 Goals – What do I want?

One of the steps an individual has to take when confronted with a new situation is to identify her most preferred outcome or goal. This might be a very simple process if an individual possesses simple selfish preferences. However, for preferences that are more complex this might involve considerable effort. To give two examples, consider an individual that has preferences not only over payoffs but also over which actions lead to such payoffs (Weibull, 2001), or similarly, an individual that cares about fair outcomes (Bolton and Ockenfels, 2000). In both cases, the individual has to observe the form of the game and decide what sequence of actions she prefers or even more problematically, what constitutes a fair

---

[72] From more discussion on the role of emotions is economics see Elster (1998) and Loewenstein (2000).

[73] Many references exist that document this finding, Damasio (1994), Bechara et al. (1997), and Evans (2001) are just a few examples.

[74] A few authors have begun the difficult work of putting together what could someday be considered the bases for a theory of individual behavior; see Güth (2000). Furthermore, for good examples of more in-depth analysis of games using a bounded rationality/reasoning framework see Güth, Huck and Müller (2001) and van Winden (2001).

[75] The authors propose that individuals use four distinct layers of information to transform the objective game into the subjective game that is actually played. This framework is also used in Cárdenas (2000) and some aspects of it can be identified in Bowles (1998).

outcome.[76] Furthermore, unless the individual is able to foresee all possible actions in a game, an unexpected action might make the individual reevaluate what she prefers. In fact, as argued by Sen (2002), reevaluation should be considered an essential element of rationality. Traditionally, the process by which an individual decides what she prefers has not received much attention. However, many factors might alter what is perceived as the preferred outcome. Recent research has highlighted two factors that play a vital role in this part of the decision-making process, specifically, intentions and emotions.

Intentions are very important because individuals seem to exhibit strong a preference for reciprocal behavior. That is, individuals like to reward good behavior and punish bad behavior.[77] In fact, in the collective action setting, cooperation seems to emerge because of the presence of two types of players: "conditional-cooperators" and "willing-punishers" Ostrom (2000). It is not hard to notice that all these players are doing is reciprocating the actions or expected actions of others. Logically, in order to reciprocate, an individual needs to judge if an action was good or bad. When doing so, intentions or more accurately perceived intentions play an essential role. Experimental evidence (Falk, Fehr and Fischbacher, 2001) as well as causal observation tells us that the same action can receive dissimilar responses if the perceived intentions behind it are different (Rabin, 1998). Thus, a better modeling of how individuals perceive the intentions of others would help explain under what circumstances people are more likely to reward cooperation or punish defection. There are a few attempts to model intentions, most notably Rabin (1993). However, to the best of my knowledge, there is not a complete model that does so in a dynamic N-player collective action setting.[78]

Emotions are the other factor that is crucial when an individual decides what are her goals in a game. As it has been shown in various studies, emotions help us decide what is more important for us (Damasio, 1994). Thus, depending on our emotional state different goals might come up as desirable. For example, an individual in a good mood will tend to be more optimistic and thus concentrate on positive outcomes; hence, she might be more willing to

---

[76] This kind of complex preferences might be represented by some kind of meta-preferences that take into account past experiences and even the experiences of others (Becker, 1996). However, this does not alter the argument that some effort is to be spent figuring out what the individual prefers.

[77] For the interested readers, there is a growing literature on reciprocity. See Dufwenberg and Kirchsteiger (2003), Falk and Fischbacher (1999), Bowles and Gintis (2002), and Fehr and Gächter (2000 and 2001), for both theoretical and experimental work. In addition, see also Carpenter and Matthews (2003) for work on social reciprocity (that is reciprocity towards strangers).

[78] Preliminary attempts in this direction include Dufwenberg and Kirchsteiger (2003), and Falk and Fischbacher (1999).

cooperate than an individual in a bad mood (Evans, 2001). Furthermore, emotions are also essential to draw our attention when, confronted with unexpected outcomes, there might be a need to reevaluate our goals (Damasio, 1994). For instance, an individual that considers cooperation as an important goal might be surprised by someone else's very low contributions, and if this causes anger, then she might reevaluate her goal to include hurting the other person as part of her objective[79]. Understanding which emotions play the most relevant roles in collective action is necessary if we wish to have a complete framework.[80] More empirical as well as theoretical work is necessary in order to accomplish this. For example, very little research has been conducted on "collective" emotions; that is, on the difference between emotions felt when people are in an individual versus a collective setting.[81]

Finally, we can briefly mention other variables which are considerably important and that, with a few exceptions, have been relatively ignored. One such variable is identity or more specifically, how much an individual identifies herself with the group. Experiments done in the laboratory (Orbell, van de Kragt and Dawes, 1988; Brown-Kruse and Hummels, 1993) as well as in the field (Cárdenas and Carpenter, 2002) hint that this might indeed be an important factor.[82] A second variable is the factors that determine the intrinsic motivation of individuals. Motivation is important in many settings, such as the workplace and of course in collective action. However, little is known on the interaction of motivation and behavior (see Benabou and Tirole, 2002).

4.2 Context – What is the game?

Besides looking within herself, an individual must also analyze her surroundings. To do so, she has to take into account many different things among which two are very important. The first is the structure of the game, meaning, among other things, an analysis of the actions that she and others can take, and of any formal and informal institutions, such as norms, that might affect behavior. The second is the characteristics of the other players; in other words, the individual has to form beliefs of what are her opponents' goals and possible actions. Of course, if the individual is playing a repeated game, these beliefs ought to be updated

---

[79] See Bowles and Gintis (2001) for more on the role of emotions and punishment on collective action.

[80] See Fessler and Haley (2002) for a considerable effort in this direction.

[81] A couple of exceptions are Lawler and Thye (1999), and Lawler (1992). In this case, the authors suggest that successful social exchange can lead people to from emotional ties with each other. Subsequently, if this leads to more altruistic behavior then collective action can be less of a problem.

[82] For a more theoretical discussion on identity, see Akerlof and Kranton (2000), or Calvert (2002).

whenever new information is obtained. In this part of the decision-making process, the bounded rationality/reasoning literature indicates two relevant factors. Specifically, when it comes to analyzing the structure of the game, framing becomes important, and when it comes to forming the beliefs of the other's actions, the existence of norms becomes crucial.

As possibly any good politician knows, framing is a key determinant of behavior. Depending on how the game is framed the attention of the players will focus on more cooperative or non-cooperative outcomes. It is beyond the scope of this paper to go through the evidence in this respect.[83] However, we should indeed point out that economists have done little effort to understand why framing has a substantial effect, and furthermore, if this effect can be consistently measured and predicted.

Even though it is by no means surprising to affirm that social norms have a substantial effect behavior, economists have tended to ignore their importance for collective action. However, recent findings demonstrate that social norms are actually very important. The reason for this is that, in many cases, people make decisions thinking in relative terms. If they do so, the reference point they use to judge each situation can largely determine the outcome.[84] When interacting in groups, individuals will often use established norms as a common reference point.[85] Hence, if in a group there is a norm to contribute a certain amount to the collective good, then an individual who intentionally contributes below (above) the norm could cause other players to feel anger (gratitude) and urge them to punish (reward) the individual.

Therefore, in order to figure out how collective action arises we first have to understand how norms that encourage cooperation emerge. If norms are established in the initial stages of the game then this can turn out to be a heavily path-dependent process that would explain why some groups are locked into bad outcomes while other groups manage to successfully cooperate. Furthermore, this highlights the significance of factors, such as leadership (Calvert, 1992; Hermalin, 1998) and communication (Dawes, van de Kragt and Orbell, 1990), which can be especially important in the first periods of a game.

4.3 Strategies – How do I get what I want in this game?

The final part of the decision-making process combines the elements of the last two parts in order to arrive to the best strategy. It is here that the player considers the strategic aspects of

---

[83] For example, see Pruitt (1967 and 1970) for results from a decomposed prisoner's dilemma.

[84] In fact, both factors mentioned above (intentions and emotions) depend on a reference point.

[85] See Keser and van Winden (2000) for evidence that contributions in public good experiments indeed move towards a common point, in this case the average contribution.

the game, that is, what is the best way of accomplishing ones goals given the structure of the game and the characteristics of the other players. Usually, at this point, we assume that individuals simply calculate what is the best action to take for each possible combination of the others' actions, that is, they compute a best reply function. However, doing so is extremely difficult in a collective action setting, especially if an individual has no information on the subjective payoffs of the game, and in addition, she has to take into account the intentions her actions will convey and anticipate what will be the emotional reactions of herself and the other players. Cognitively and emotionally bounded individuals would probably rely on other methods to arrive to the best course of action. Unfortunately, although great progress has been done identify some of these methods (Gigerenzer et al., 1999; Gigerenzer and Selten, 2001), we still do not have a complete model that describes how they fit in a coherent framework. Nevertheless, it might be useful to explore what occurs if individuals use some of these methods when faced with a collective action problem.

We can briefly discuss two basic types of methods, backward-looking and forward-looking. Backward-looking methods would include heuristics (such as imitate the most successful player in each round) or decisions that are based solely on what the individual has learned in the past. It is evident that there is a learning effect in many experiments (Isaac, Walker, and Williams, 1994) as well as the fact that in many cases simple heuristics seem to be good predictors of behavior (Gigerenzer et al., 1999). Therefore, it is important to determine what are the effects of backward-looking methods on collective action. There are a number of learning models in the literature; however, their results are highly dependent on the type of aspiration levels used (see Macy and Flache, 2002). Additional research on how aspirations are set would enrich the models considerably. Forward-looking methods include some kind of optimization in which agents attempt to choose the best alternative based on their limited understanding of the situation. There is still no consensus on how individuals do this. Some of the proposed methods include induction (Holland et al., 1986) and satisficing (Selten, 1998a; Simon, 1957). What we do know is that, in some cases individuals reach optimal decisions while in other cases they fail considerably (Conlisk, 1996). Regrettably, collective action research is usually based on either a backward or a forward-looking method. If we want to come closer to human decision-making, what is actually needed is a combination of both.

A potentially fruitful way of discovering how people combine both types of methods is through the research of emotions. Although it is not entirely clear how, emotions seem to play an important role in the "choice" of which method to use (Damasio, 1994; Bechara et al., 1997). For example, emotions such as anger will promote the use of more cognitively simple methods (Kaufman, 1999). Thus, if we are able to determine what emotions are experienced at different points in the collective action game, we might get a better idea of which method to apply at each moment. Furthermore, it is worth mentioning that emotions also directly affect both backward and forward-looking methods. In the first case, by highlighting which experiences are worth learning from (Blaney, 1986; Kahneman 2000) and in addition by affecting which experiences are recalled from our memory (Oatley and Jenkins, 1996). In the second case, by urging individuals to focus on particular goals (Bechara et al., 1997).[86] In conclusion, exploring the role of emotions in these settings is one of the lines of research that most promises to shed light on our way to a complete theory of collective action.

As a final point, it is important to realize that all three parts of the decision-making process presented here are interdependent. It is rather obvious that the preferences and beliefs formed in the first two parts are considered when picking the best course of action. However, the deliberation process undergone in picking the right action might also affect the previous parts. As for example when someone realizes that a goal might not be attainable and thus must reevaluate what goals she should strive to achieve.

5. Final Thoughts

Throughout the paper, we have seen how research on the provision of collective goods has consisted of numerous approaches, each providing insights into different aspects of the collective action phenomenon. However, what on the surface might look like a trivial everyday activity has turned out to be one of the most complex problems faced by the social sciences. After decades of research, we still do not have a satisfactory explanation of how collective action emerges, something that the development of experimental economics has been made painfully clear. Experiment after experiment has shown that our current theories cannot explain the obtained results in even the most basic of settings.[87] On the other hand, years of running experiments have also provided us with uncountable opportunities to start identifying the most important determinants of collective action. Thus, experimental evidence

[86] In economics, see van Winden's analysis of the emotions in the decision to destroy income in the power-to-take game (van Winden, 2001).

[87] Before sounding too pessimistic, it is important to note that we have learned considerably of numerous ways of increasing cooperation. Thus, we are able to predict changes in cooperation. The problem is that we do not have a clear model that predicts the level of cooperation.

coupled with the development of behavioral economics has granted us much better tools to understand such a complex situation.

In order to progress in our search for a complete theory of collective action, we have to start modeling more psychologically realistic agents. If indeed we want a comprehensive theory, we have to be able to explain much more than has been previously achieved. For example, if we are to model a collective action situation in which costly sanctions are possible, we cannot make the simple assumption the agents will simply sanction others who contribute below a given threshold. In such a situation, we should try to model how a cooperative norm might emerge and how intentional deviations from it might trigger negative emotions, which would then lead to sanctioning. Failing to consider any of these elements may result in a model that oversees important variables. For instance, in this example if we fail to model how sanctioning is caused by anger, and we simply assume that an intentional deviation from the norm will cause retaliation; then, we would miss the fact that the emotions are experienced with different intensities (depending on other variables such as how the income was earned), and depending on the intensity of the emotion sanctioning behavior might be different.[88] Hence, a complete theory of collective action must necessarily use agents which present both bounded rationality and bounded reasoning.

[88] As Bosman and van Winden (2002) find in the respondent behavior in a power-to-take game.

References

Abreu, D. (1988). On the theory of infinitely repeated games with discounting. *Econometrica* 56: 383-96.

Agrawal, A. and S. Goyal (2001). Group size and collective action: Third-party monitoring in common pool resources. *Comparative Political Studies* 34: 63-93.

Akerlof, G. E. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics* 115: 715-753.

Aldrich, J.H. (1997). When is it rational to vote? In D.C. Mueller (ed.), *Perspective on Public Choice*. Cambridge: Cambridge University Press.

Andreoni J. (1989). Giving with impure altruism: Applications to charity and ricardian equivalence. *Journal of Political Economy* 97: 1447-1458.

Andreoni J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal* 100, 464-477.

Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

Axelrod, R. (1997). *The Complexity of Cooperation*. Princeton: Princeton University Press.

Bagnoli, M. and Lipman, B. L. (1989). Provision of public goods: Fully implementing the core through private contributions. *The Review of Economic Studies* 56: 583-601.

Barry, B. (1978). *Sociologists, Economists and Democracy*. Chicago: Chicago University Press.

Baumol, W. J. and W. E. Oates (1988). *The theory of environmental policy*. 2nd ed. New York: Cambridge University Press.

Bechara, A., H. Damasio, D. Tranel, and A. R. Damasio (1997). Deciding advantageously before knowing the advantageous strategy. *Science* 275: 1293-1294.

Becker, G. (1996). *Accounting for tastes*. Cambridge: Harvard University Press.

Benabou, R. and Tirole, J. (2002) Self-confidence and personal motivation. *The Quarterly Journal of Economics* 117: 871-915.

Bendor, J. and D. Mookherjee (1987). Institutional structure and the logic of ongoing collective action. *American Political Science Review* 81: 129-154.

Bergstom, T., L. Blume, and H. Varian. (1986). On the private provision of public goods. *Journal of Public Economics* 29: 25-49.

Bester, H. and Güth W. (1998). Is altruism evolutionarily stable? *Journal of Economic Behavior and Organization* 34: 193-209.

Binmore, K. and L. Samuelson (1992). Evolutionary stability in repeated games played by finite automata. *Journal of Economic Theory* 57: 278-305.

Blaney, P. H. (1986). Affect and memory: A review. *Psychological Bulletin* 99: 229-246.

Blume, A., Y. G. Kim, and J. Sobel (1993). Evolutionary stability in games of communication. *Games and Economic Behavior* 5: 547-575.

Bolton, G. and A. Ockenfels (2000). A theory of equity, reciprocity, and competition. *American Economic Review* 90: 166-193.

Bosman, R., and F. van Winden (2002). Emotional hazard in a power-to-take game experiment. *The Economic Journal* 112: 147-169.

Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature* 36: 75-111.

Bowles, S. and H. Gintis (2002). The origins of human cooperation. Working paper #02-08-035. Santa Fe Institute.

Bowles, S. and H. Gintis (2001). The economics of shame and punishment. Working paper #02-07-028. Santa Fe Institute.

Brown-Kruse, J. and D. Hummels (1993). Gender effects in laboratory public goods contribution: Do individuals put their money where their mouth is? *Journal of Economic Behavior and Organization* 22: 255-267.

Calvert, R.L. (1992). Leadership and its basis in problems of social coordination. *International Political Science Review* 13, 7-24.

Calvert, R.L. (2002). Rationality, identity, and expression. In Katznelson I. and H. Milner (eds.), *Political Science: The State of the Discipline*, 3rd edition. New York: W.W. Norton and American Political Science Association.

Camerer, C. (1998). Bounded rationality in individual decision making. Experimental Economics 1: 163-183.

Cárdenas, J. C. (2000). Rural institutions, poverty and cooperation: Learning from experiments and conjoint analysis in the field. Ph.D. dissertation. University of Massachusetts.

Cárdenas, J. C. and J. Carpenter (2002). Using cross-cultural experiments to understand the dynamics of a global commons. Working paper #2. Middlebury College Center for International Affairs.

Cárdenas, J. C. and E. Ostrom (2001). What do people bring into the game? How norms help overcome the tragedy of the commons. Presented at the “Norms and Law” seminar organized by the Center for Interdisciplinary Studies, School of Law, Washington University in St. Louis.

Carpenter, J. and P. Matthews (2003). Social reciprocity. Working paper. Middlebury College.

Chan, K. S., S. Mestelman, R. Moir, and R. A. Muller (1996). The voluntary provision of public goods under varying income distributions. *Canadian Journal of Economics* 29: 54-69.

Chwe, S. Y. (1999). Structure and strategy in collective action. *American Journal of Sociology* 105: 128-156.

Chwe, S. Y. (2000). Communication and coordination in social networks. *Review of Economic Studies* 67: 1-16.

Coleman, J. S. (1988). Free riders and zealots: The role of social networks. *Sociological Theory* 6: 52-57.

Conlisk, J. (1996). Why bounded rationality? *Journal of Economic Literature* 34: 669–700.

Coombs, C. (1973). A reparameterization of the prisoner’s dilemma game. *Behavioral Science* 18: 424-428.

Damasio, A. R. (1994). *Descartes' error. Emotion, reason and the human brain*. New York: G.P. Putnam's Sons.

Davis, D. D. and C. A. Holt (1993). *Experimental Economics*. New Jersey: Princeton University Press,

Dawes, R. M., J. M. Orbell, R. T. Simmons, and A. J. C. van de Kragt (1986). Organizing groups for collective action. *American Political Science Review* 80: 1171-1185.

Dawes, R. M., A. J. C. van de Kragt, and J. M. Orbell (1990). Cooperation for the benefit of us: Not me, or my conscience. In J. J. Mansbridge (ed.), *Beyond Self-Interest*. Chicago: University of Chicago Press.

Droste, E., R. P. Gilles, and C. A. Johnson (2000). Evolution of conventions in endogenous social networks. Working Paper. Virginia Polytechnic Institute.

Dufwenberg, M. and Kirchsteiger, G. (2003). A theory of sequential reciprocity. Working paper. Stockholm University.

Economist (2001a). Not quite so keen. *Economist* 359(8224): 52-54.

Economist (2001b). Mike's mighty challenge. *Economist* 361(8247): 27-28.

Elster, J. (1985). Rationality, morality, and collective action. *Ethics* 96: 136-155.

Elster, J. (1989). *The Cement of Society*. Cambridge: Cambridge University Press.

Elster, J. (1998). Emotions and economic theory. *Journal of Economic Literature* 36: 47-74.

Esteban, J. and D. Ray (2001). Collective action and the group size paradox. *American Political Science Review* 95: 663-672.

Evans, D. (2001). *Emotion: The science of sentiment*. New York: Oxford University Press.

Falk, A., E. Fehr, and U. Fischbacher (2001). Testing theories of fairness: Intentions matter. Working paper #63. University of Zurich.

Falk, A. and U. Fischbacher (1999). A theory of reciprocity. Working paper No. 6. Institute for Empirical Research in Economics, University of Zürich.

Falk, A. and M. Kosfeld (2003) It's all about connections: Evidence on network formation. Working paper #146. University of Zurich.

Farrell, J. and M. Rabin (1996). Cheap talk. *Journal of Economic Perspectives* 10: 103-118.

Fehr, E. and U. Fischbacher (2002). Why do social preferences matter: The impact of non-selfish motives on competition, cooperation and incentives. *The Economic Journal* 112: 1-33.

Fehr, E., U. Fischbacher, and S. Gächter (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters* 71: 397-404.

Fehr, E. and S. Gächter (2000). Fairness and retaliation. *Journal of Economic Perspectives* 14: 159-181.

Fehr, E. and S. Gächter (2002). Altruistic punishment in humans. *Nature* 415: 137-140.

Fehr, E. and K. Schmidt (1999). A Theory of fairness, competition and cooperation. *Quarterly Journal of Economics* 114: 817-868.

Fessler, D. M. T. and K. J. Haley (2002). The strategy of affect: Emotions in human cooperation. Forthcoming in P. Hammerstein (ed.) Dahlem Workshop Report, *Genetic and Cultural Evolution of Cooperation*. Cambridge: MIT Press.

Finkel, S. E., E. N. Muller, and K. D. Opp (1989). Personal influence, collective rationality and mass political action. *American Political Science Review* 83: 885-903.

Finkel, S. E. and E. N. Muller (1998). Rational choice and the dynamics of collective action: Evaluating alternative models with panel data. *American Political Science Review* 92: 37-49.

Frank, R. H. (1987). If Homo Economicus could choose his own utility function, would he want one with a conscience. *American Economic Review* 77: 593-604.

Fudenberg, D. and E. Maskin (1986). The Folk Theorem in repeated games with discounting or with incomplete information. *Econometrica* 54: 533-554.

Fudenberg, D. and J. Tirole (1991). *Game Theory*. Cambridge: MIT Press.

Gavious, A. and S. Mizrahi (2000). Information and common knowledge in collective action. *Economics and Politics* 12: 297-319.

Gavious, A. and S. Mizrahi (2001). A continuous time model of the bandwagon effect in collective action. *Social Choice and Welfare* 18: 91-105.

Gigerenzer, G., P. M. Todd and the ABC Research Group (1999). *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Gigerenzer, G. and R. Selten (eds.) (2001). *Bounded Rationality: The Adaptive Toolbox*. Cambridge: MIT Press.

Gould, R. V. (1993). Collective action and network structure. *American Sociological Review* 58: 182-196.

Goyal, S. and F. Vega-Redondo (2001). Network formation and social coordination. Working paper. University of London.

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology* 83: 1420-1443.

Großer, J., and T. Kugler (2002). The impact of group size uncertainty on voter participation: an experimental study. Working paper, University of Amsterdam.

Güth, W. (2000). Boundedly rational decision emergence: A general perspective and some selective illustrations. *Journal of Economic Psychology* 21, 433-458.

Güth, W. and H. Kliemt (1998). The indirect evolutionary approach: Bridging the gap between rationality and adaptation. *Rationality and Society* 10: 377-399.

Güth, W. and H. Kliemt (2000). Evolutionarily stable co-operative commitments. *Theory and Decision* 49: 197-221.

Güth, W., S. Huck and W. Müller (2001). The relevance of equal splits in ultimatum games. *Games and Economic Behavior* 37, 161-169.

Guttman, J. M. (1987). A non-Cournot model of voluntary collective action. *Economica* 54: 1-19.

Haan, M. and P. Kooreman (2000). How majorities can lose the election: Another voting paradox. *Social Choice and Welfare* forthcoming.

Hardin, R. (1971). Collective action as an agreeable n-prisoner's dilemma. *Behavioral Science* 16: 472-481.

Hardin, R. (1982). *Collective Action*. Baltimore: The Johns Hopkins University Press.

Harsanyi, J. and R. Selten (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge: MIT Press.

Hechter, M. (1983). A theory of group solidarity. In M. Hechter (ed.) *The Microfoundations of Macrosociology*. Philadelphia: Temple University Press.

Hechter, M. (1984). When actors comply: Monitoring costs and the production of social order. *Acta Sociologica* 27: 161-183.

Heckathorn, D.D. (1988). Collective sanctions and the creation of prisoner's dilemma norms. *American Journal of Sociology* 94: 535-562.

Heckathorn, D.D. (1989). Collective action and second-order free-rider problem. *Rationality and Society* 1: 78-100.

Heckathorn, D.D. (1990). Collective sanctions and compliance norms: A formal theory of group mediated social control. *American Sociological Review* 55: 366-384.

Heckathorn, D.D. (1993). Collective action and group heterogeneity: Voluntary provision versus selective incentives. *American Sociological Review* 93: 329-350.

Heckathorn, D.D. (1996). The dynamics and dilemmas of collective action. *American Sociological Review* 61: 250-277.

Hermalin, B.E. (1998). Toward an economic theory of leadership: leading-by-example. *American Economic Review* 88: 1188-1206.

Hirsch, F. (1976). *Social Limits to Growth*. Cambridge: Harvard University Press.

Hirschman, A.O. (1970) *Exit, Voice, and Loyalty. Responses to Decline in Firms, Organizations, and States*. Cambridge: Harvard University Press.

Holland, J. H., K. J. Holyvak, R. E. Nisbett, P. R. Thagard (1986). *Induction Processes of inference, learning, and discovery*. Cambridge: MIT Press.

Höllander, H. (1990). A social exchange approach to voluntary cooperation. *American Economic Review* 80: 1157-1167.

Issac, R. M., J. Walker, and S. Thomas (1984). Divergent evidence on free riding: An experimental examination of possible explanations. *Public Choice* 43: 113-49.

Issac, R. M., J. Walker, and A. W. Williams (1994). Group size and voluntary provision of public goods: Experimental evidence utilizing large groups. *Journal of Public Economics* 54: 1-36.

Jackson, M. O. and A. Watts (2002) On the formation of interaction networks in social coordination games. *Games and Economic Behavior* 41: 265-291.

Jevons, W.S. (1911). Brief account of a general mathematical theory of political economy. *Theory of Political Economy*. London: Macmillian, forth edition, Appendix III, 303-314.

Kahneman, D. (2000). Experienced utility and objective happiness: A moment-based approach. In Kahneman, D. and Tversky, A. (eds.), Choices, Values and Frames. New York: Cambridge University Press and the Russell Sage Foundation.

Kaufman, B. E. (1999). Emotional arousal as a source of bounded rationality. *Journal of Economic Behavior and Organization* 38: 135-144.

Knack, S. (1992). Civic norms, social sanctions, and voter turnout. *Rationality and Society* 4: 133-156.

Königstein, M. and W. Müller (2000). Combining rational choice and evolutionary dynamics: The indirect evolutionary approach. *Metroeconomica* 51: 235-256.

Kreps, D., P. Milgrom, J. Roberts, and R. Wilson (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory* 27: 245-52.

Laffont, J. J. (1988). *Fundamentals of Public Economics*. Cambridge: MIT Press.

Lawler, E. J. (1992). Affective attachments to nested groups: A choice-process theory. *American Sociological Review* 57: 327-339.

Lawler, E. J. and S. R. Thye (1999). Bringing emotions into social exchange theory. *Annual Review of Sociology* 25: 217-244.

Ledyard, J. O. (1984). The pure theory of large two-candidate elections. *Public Choice* 44: 7-41.

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In A. Roth and J. Kagel (eds.), *A Handbook of Experimental Economics*. Princeton: Princeton University Press.

Levati, M. V. and T. Neugebauer (2001). An application of the English clock market mechanism to public goods games. Working paper, Max Planck Institute for Research into Economic Systems.

Lo A.W. and D.V. Repin (2002). The psychophysiology of real-time financial risk processing. *Journal of Cognitive Neuroscience* 14: 323-339.

Loewenstein, G. F. (2000). Emotions in economic theory and economic behavior. *American Economic Review, Papers and Proceedings* 90: 426-432.

Lohman, S. (1993). A signaling model of informative and manipulative political action. *American Political Science Review* 87: 319-333.

Lohman, S. (2000). Collective action cascades: An information rationale for the power in numbers. *Journal of Economic Surveys* 14: 655-679.

Macy, M. W. (1990). Learning theory and the logic of critical mass. *American Sociological Review* 55, 809-826.

Macy, M. W. (1991a). Chains of cooperation: Threshold effects in collective action. *American Sociological Review* 56, 730-747.

Macy, M. W. (1991b). Learning to cooperate: Stochastic and tacit collusion in social exchange. *American Journal of Sociology* 97: 808-843.

Macy, M. W. (1993). Backward-looking social control. *American Sociological Review* 58, 819-836.

Macy, M.W. and A. Flache (1995). Beyond rationality in models of choice. *Annual Review of Sociology* 21: 73-91.

Macy, M.W. and A. Flache (2002). Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences U.S.A. May 14* 99: 7229-36.

Menger, C. (1871). *Principles of Economics*, ed. by Dingwall, J. and B.F. Hoselitz. New York: New York University Press.

Morton, R.B. (1991). Groups in rational turnout models. *American Journal of Political Science* 35: 758-776.

Oatley, K. and Jenkins, J. M. (1996). *Understanding emotions*. Cambridge and Oxford: Blackwell.

Offerman, T. (1997). *Beliefs and Decision Rules in Public Good Games*. Dordrecht: Kluwer.

Oliver, P. (1980). Rewards and punishments as selective incentives for collective action. *American Journal of Sociology* 86: 1356-1375.

Oliver, P. (1993). Formal models of collective action. *Annual Review of Sociology* 19: 271-300.

Oliver, P. and G. Marwell (1988a). The paradox of group size and collective action: A theory of the critical mass II. *American Sociological Review* 53: 1-8.

Oliver, P. and G. Marwell (1988b). Social networks and collective action: A theory of the critical mass III. *American Journal of Sociology* 94: 502-534.

Oliver, P., G. Marwell, and R. Teixeira (1985). A theory of the critical mass I: Interdependence, group heterogeneity, and the production of collective action. *American Journal of Sociology* 91: 522-556.

Olson, M. (1965). *The logic of collective action*. Cambridge: Harvard University Press.

Opp, K. D. (1989). *The Rationality of Political Protest: A Comparative Analysis of Rational Choice Theory*. Boulder: Westview Press.

Orbell, J. and R. Dawes (1993). Social welfare, cooperators' advantage, and the option of not playing the game. *American Sociological Review* 58:787-800.

Orbell, J., A. van de Kragt, and R. Dawes (1988). Explaining discussion-induced cooperation. *Journal of Personality and Social Psychology* 54: 811-819.

Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action: Presidential address, American Political Science Association, 1997. *American Political Science Review* 92: 1-22.

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives* 14: 137-158.

Palfrey, T. R. and H. Rosenthal (1985). Voter participation and strategic uncertainty. *American Political Science Review* 79: 62-78.

Pruitt, D. G. (1967). Reward structure and cooperation: The decomposed prisoner's dilemma game. *Journal of Personality and Social Psychology* 7: 21-27.

Pruitt, D. G. (1970). Motivational processes in the decomposed prisoner's dilemma game, *Journal of Personality and Social Psychology* 14: 227-238.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review* 83: 1281-1302.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature* 36: 11-46.

Rabin, M. (2002). A perspective on psychology and economics. *European Economic Review* 46, 657-685.

Riedl, A. and Ule, A. (2002). Exclusion and cooperation in social network experiments. Working paper. University of Amsterdam.

Roth, A. E. (1995). Introduction to Experimental Economics. In A. Roth and J. Kagel (eds.), *A Handbook of Experimental Economics*. Princeton: Princeton University Press.

Sandler, T. and T. Tschirhart (1980). The economic theory of clubs: An evaluative survey. *Journal of Economic Literature* 18: 1481-1521.

Schelling, T.C. (1973). Hockey Helmets, Concealed Weapons, and Daylight Saving: A Study of Binary Choices with Externalities. *The Journal of Conflict Resolution* 17: 381-428.

Schelling, T.C. (1978). *Micromotives and Macrobehavior*. New York: W.W. Norton.

Schofield, N. (1975). A game theoretic analysis of Olson's game of collective action. *Journal of Conflict Resolution* 19: 441-461.

Schram, A. (2000). Sorting out the seeking: the economics of individual motivations. *Public Choice* 103: 231-258.

Selten, R. (1998a). Aspiration adaptation theory. *Mathematical Psychology* 42, 191-214.

Selten, R. (1998b). What is bounded rationality? in: G.Gigerenzer and R. Selten, eds., *Bounded Rationality: The Adaptive Toolbox*. Cambridge: MIT Press, 281-296.

Sen, A. (1979). Rational fools. In F. Hahn and M. Hollis (eds.), *Philosophy and Economic Theory*. Oxford: Oxford University Press.

Sen, A. (2002). *Rationality and Freedom*. Cambridge: Harvard University Press.

Simon, H. (1957). *Models of Man: Social and Rational; Mathematical Essays on Rational Human Behavior in Society Setting*. New York: John Wiley.

Stigler, G. L. (1974). Free riders and collective action: An appendix to theories of economic regulation. *The Bell Journal of Economics and Management Science* 5: 359-365.

Thaler, R. H. (2000). From Homo Economicus to Homo Sapiens. *Journal of Economic Perspectives* 14: 133–141.

Udéhn, L. (1993). Twenty-five years with the logic of collective action. *Acta Sociologica 3*6: 239-261.

Uhlaner, C. (1989). Rational turnout: The neglected role of groups. *American Journal of Political Science* 33: 390-422.

van Veelen, M. (2002) Altruism, fairness and evolution: the case for repeated stochastic games. Discussion Paper #02-111/1. Tinbergen Institute.

Weibull, J. W. (1995). *Evolutionary Game Theory*. Massachusetts: MIT Press.

Weibull, J. W. (2001). Testing game theory. Working Paper. Stockholm School of Economics.

Wilson, J. Q. (1973). *Political Organizations*. New York: Basic Books.

van Winden, F. (2001). Emotional hazard exemplified by taxation-induced anger. *KYKLOS*, 54, 491-506.

van Winden, F. (2002). Experimental investigation of collective action. Forthcoming in: Stanley Winer and Hirofumi Shibata (eds.), *Political Economy and Public Finance: The Role of Political Economy in the Theory and Practice of Public Economics*, Edward Elgar, Cheltenham.

Witt, U. (1986). Evolution and stability of cooperation without enforceable contracts. *KYKLOS* 39: 245-266.

Yamagishi, T. and K. S. Cook (1993). Generalized exchange and social dilemmas. *Social Psychology Quarterly* 56: 235-248.