The Formation of Social Preferences: Some Lessons from Psychology and Biology

Louis Lévy-Garboua, Claude Meidinger, Benoît Rapoport

Cahiers de la MSE

The formation of social preferences: some lessons from psychology and biology

Louis LEVY-GARBOUA
Claude MEIDINGER
Benoît RAPOPORT

2004.10

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
ISSN : 1624-0340

THE FORMATION OF SOCIAL PREFERENCES:

SOME LESSONS FROM PSYCHOLOGY AND BIOLOGY*

By

Louis Lévy-Garboua, Claude Meidinger, et Benoît Rapoport

TEAM (CNRS), Université Paris I (Panthéon Sorbonne)

* Forthcoming in *A Handbook on the Economics of Giving, Reciprocity and Altruism*, L.A. Gerard-Varet, S.C. Kolm, and J. Mercier-Ythier (Eds.), Amsterdam: Elsevier.

Abstract

The goal of this paper is to draw some lessons for economic theory from research in psychology, social psychology and, more briefly, in biology, which purports to explain the "formation" of social preferences. We elicit the basic mechanisms whereby a variety of social preferences are determined in a variety of social contexts. Biological mechanisms, cultural transmission, learning, and the formation of cognitive and emotional capacities shape social preferences in the long or very long run. In the short run, the built-in capacities are utilized by individuals to construct their own context-dependent social preferences. The full development of social preferences requires consciousness of the individual's similarities and differences with others, and therefore knowledge of self and others. A wide variety of context-dependent social preferences can be generated by just three cognitive processes: identification of self with known others, projection of known self onto partially unknown others, and categorization of others by similarity with self. The self can project onto similar others but is unable to do so onto dissimilar others. The more can the self identify with, or project onto, an other the more generous she will be. Thus the self will find it easier to internalize and predict the behavior of an in-group than an out-group and will generally like to interact more with the former than with the latter. The main social motivations can be simply organized by reference to social norms of justice or fairness that lead to reciprocal behavior, some kind of self-anchored altruism that provokes in-group favoritism, and social drives which determine an immediate emotional response to an experienced event like hurting a norm's violator or helping an other in need.

Keywords : Formation of social preferences, psychology, social psychology sociale, biology.

Résumé

Nous dégageons les leçons qui peuvent être tirées des recherches consacrées, en psychologie, en psychologie sociale et en biologie, à la « formation » des préférences sociales. Nous identifions les mécanismes fondamentaux par lesquels se déterminent les préférences sociales en fonction du contexte des interactions sociales. A long terme, ce sont les mécanismes biologiques, la transmission culturelle, l'apprentissage, et la formation des capacités cognitives et émotionnelles. A court terme, les capacités qui se sont formées sont mobilisées par l'individu pour construire ses propres préférences sociales en fonction du contexte. Un développement complet des préférences sociales exige une conscience des ressemblances et des différences entre soi et les autres, qui passe par une connaissance de soi et des autres. Trois processus cognitifs suffisent a générer une grande variété de préférences sociales : l'identification de Soi a d'autres connus, la projection d'un Soi connu sur d'autres en partie inconnus, et la catégorisation des autres d'après leur ressemblance a Soi. Ego ne peut se projeter que sur d'autres qui lui ressemblent. Plus il parvient a s'identifier à un autre ou à se projeter sur lui, plus il sera généreux envers lui. Par conséquent, Ego trouvera plus facile d'internaliser et de prévoir le comportement d'un en-groupe que d'un hors-groupe et préfèrera en général interagir avec l'un qu'avec l'autre. Les principales motivations sociales sont de trois types : l'adhésion à des normes sociales de justice ou d'équité qui engendre une réciprocité de comportements, une forme d'altruisme auto-centré qui conduit à favoriser son en-groupe, et des pulsions sociales qui déclenchent une réaction émotionnelle immédiate à un évènement ressenti, comme faire du mal à celui qui a enfreint la norme ou aider celui qui est dans le besoin.

Mots-clés : formation des preferences sociales, psychologie, psychologie sociale, biologie.

Classification JEL : B40, D63, D64, D70, D80, Z13

I. Introduction

Adam Smith wrote *The Theory of Moral Sentiments* (1759) almost twenty years before *The Wealth of Nations* (1776). The former is a book about other-regarding behavior, while the latter is justly famous for describing individuals as driven by their self-interest in the marketplace. Adam Smith cannot be suspect for ignoring "social preferences" which come into play in interpersonal relations but he likely felt that the concern for others would eventually be superseded by the forces of competition imposed by the efficient functioning of large markets. Adam Smith's intuition has proved to be right. When several experimental players compete for the best offer to a single responder (who may reject all offers, in which case no one gets anything, or accept the best offer without alteration), competition dictates that the responder take the lion's share after only a few repetitions of the game (Roth et al. 1991). By contrast, when a single offer is made to the single responder under the same conditions, as in the ultimatum bargaining game (Güth et al. 1982), the player who first receives a sum of money to be shared does not exploit her bargaining power and usually gives an equal or almost equal share to the second player. This robust observation, like many others, is plainly inconsistent with the "economic" assumption of selfishness which has become standard- by way of parsimony- since *The Wealth of Nations*. The addition of stable altruistic or envious preferences (Becker 1974) is not sufficient either to predict behavior observed in many games. For instance, Camerer and Thaler (1995) remark that, in the ultimatum game, "randomly drawn proposers often make generous offers as an altruist would, but randomly drawn responders often reject low offers as an envious person would". Thus, subjects' behavior is role-dependent and cannot be permanently described as either altruistic or envious. In recent years, there have been a few important attempts from economists for reconciling the contrasted behavior appearing in the ultimatum game, the market game and other games as well (Rabin 1993, Fehr and Schmidt 1999, Bolton and Ockenfels 2000, among others). These papers have generally substituted social utility functions for selfish, money maximizing behavior.

* We thank Ernst Fehr and the editors for very constructive remarks which helped us to improve the paper.

The goal of this chapter is not to review the fast growing economic contributions to social preferences[1], though, but to draw some lessons for economists from research in psychology, social psychology and, more briefly, in biology, which purports to explain the "formation" of social preferences. In contrast with the standard practice in economics, the biological approach does not assume a given distribution of preferences at the societal level and the psychological approach does not even assume given preferences at the individual level, since the various processes of preference formation constitute their common object of study. Wider access of economists to the important literature in psychology and biology is needed in our view to elicit basic mechanisms whereby a variety of social preferences are determined in a variety of social contexts. For instance, selfish behavior may arise out of selfish preferences (a special case of social preferences holding when Self systematically disregards Others in social contexts) but it may also arise out of non-selfish preferences as a result of repeated competition. In a similar fashion, the motives underlying pro-social behavior like helping, sharing or giving, may be altruistic but may also arise from a sense of justice. Given the special emphasis of the present handbook on giving and other pro-social attitudes, rewards are likely to have a greater weight than punishments in our review. This context-dependent bias has been contained but what remains of it should not be taken as neglect of the role of negative feelings and behavior in interpersonal relations.

Although biology and psychology have a definite empirical and experimental orientation, we will be mainly interested in lessons which can be drawn from these disciplines for economic theory. A special effort will be sometimes required of us for putting psychological theories in a choice-theoretic framework while making the least prejudice to the original theoretical ideas. The methodology and content of the work under review is well-suited for raising major questions, like the following: Can the laws of evolution predict the appearance of stable genetically-determined social types? How does the development of children's cognitive abilities and experiences permanently affect pro-social behavior? How does the specific context of social interactions determine social cognition and the "constructed" social preferences (Payne, Bettman, and Johnson 1992), and do the latter follow systematic patterns? Can social preferences arise from emotions as well?

We start this review by examining in section 2 the evolutionary emergence of stable social types in the very long run. After considering non-cultural species, we move to human societies. In the next two sections, we shift to the intergenerational transmission of social preferences which takes place through learning, cognitive development, and personal experiences of children. Section 3 deals with

[1] References can be found in Rabin 1998, and Charness and Rabin 2002.

social learning and section 4 with the cognitive theories of moral and pro-social development with special emphasis on Piaget. Finally, we study the short run construction of social preferences in the context of interpersonal relations. We suggest that a great many instances of social preference formation reviewed in the social-psychological literature can be articulated with three basic mechanisms of social cognition: identification of self with others, projection of self onto others, and categorization of others by similarity with self. They all have in common to make use of the human ability to take others' perspective. These mechanisms are presented in a simple choice-theoretic framework and serve to synthesize the wide variety of results which can be found in this literature. The first two appear in section 5, and the third in section 7. The interplay of these simple mechanisms can generate a number of context-dependent social motivations in the short run, and be either reinforced or inhibited by learning from experience in the long run. The main social motivations, though, can be simply organized by reference to social norms of justice or fairness that lead to reciprocal behavior (section 6), some kind of self-anchored altruism that provokes both in-group favoritism and out-group discrimination (section 7), and social drives which determine an immediate emotional response to an experienced event like hurting a norm's violator or helping an other in need (section 8). Finally, we summarize the main lessons to be drawn from our reading of psychology and biology in section 9.

II The evolutionary emergence of social types

The natural evolution of populations in non-cultural species is usually explained by the Darwinian hypothesis of "descent with modification". If the organisms in a population differ in their abilities to survive and reproduce and if the characteristics that affect these abilities are transmitted from parents to offspring, the population will evolve. Within this framework, when one speaks of the ability of an organism to survive and reproduce, one usually refers to the *phenotype* of that organism, i.e. "the observable properties of an organism as they have developed under the combined influences of the genetic constitution of the individual and the effect of environmental factors" (Wilson 1975:591). And when one speaks of transmission of these observable properties, one usually considers that these properties are under the control of the *genotype*, i.e. the genetic constitution of an individual organism. Thus, evolution by natural selection works in a remarkable way. Genotypes are mapped onto phenotypes that have different abilities to survive and reproduce. Then, natural selection acting

differentially on the phenotypes modifies the composition of the population as it matures to the adult stage. Fitness is a measure of the survival success of the genotypes. The genotype with the highest fitness value will increase its frequency in the population.

Within this approach, one immediately encounters the problem of selection for social behavior. A species is defined as *social* (Boorman and Levitt, 1980: 2 and 12) if "its members engage, at any point in the life cycle, in sustained cooperation that goes beyond parental care and the continued association of mated pairs". More specifically, *altruistic* behavior is defined as "any behavior involving the sacrifice of a certain amount of fitness on the part of one organism (the donor) in exchange for augmented fitness on the part of a second con-specific (the recipient)". Social and altruistic behavior offers a challenge to the theory of natural selection since the latter only predicts adaptations that maximize fitness of individuals taken separately.

In what follows, without entering into the complexity of the formal models drawn from mathematical population genetics, we first want to review the different ways which evolutionary models for non-cultural species have tried to solve the problem of selection for social behavior and second to point to what can differentiate human pathways to sociality from animal pathways to sociality.

1 The problem of selection of social behavior in non-cultural species

Alternative genetic models of altruism and cooperation are usually divided among group selection, kin selection and reciprocity selection models. For the moment, we neglect the group selection approach. Since the first formulation of this hypothesis by Wynne-Edwards (1959) and its revival as a problem in mathematical genetics by Levins (1970), this approach has received numerous competing formulations distantly related to one another and often without reduction to a common basis. For this reason, escaping from the complexity of these formalisms, the evolutionary approach to animal behavior mostly retains kin selection and reciprocity selection as principal pathways to sociality[2] (Kreps and Davies 1981, McFarland 1985, Smuts et al. 1986). Kin selection is selection for altruism toward kin. Reciprocity selection is selection for cooperation between genetically unrelated individuals.

a) Kin selection

[2] But see also the Boorman and Levitt (1980) group selection model.

The importance of kin selection for the biological evolution of altruism was first anticipated by Fisher (1930) and Haldane (1953), and fully perceived by Hamilton (1964). The term kin selection (Maynard Smith 1964) was used to describe a process by which a behavioral trait is favored owing to its beneficial effects on relatives such as siblings or cousins. Between two full sibs for instance, there is a probability (coefficient of relatedness) of 0.5 that they share a copy of the same gene. Therefore, as an extreme example of altruism, a gene that programs an individual to die in order to save the life of relatives will increase in frequency in the gene pool if on average this altruistic act saves the lives of more than two brothers or sisters. More generally, in a large population of donor-recipient pairs, with each donor giving up a fraction δ units of fitness in exchange of a fitness increment of π to the recipient, Hamilton's theory predicts that an altruist gene will be selected if $-\delta + r\pi > 0$, r being the mean coefficient of relatedness across the population of pairs.

Although the concept of altruism assumes a central position in any discussion of kin selection, behavior that appears to be altruistic at the phenotypic level turns out to be genetically selfish in Hamilton's theory (Dawkins 1976). This theory can explain acts of altruism as extreme as suicide or sterility in animal species because the sacrifice of some amount of genetic material in one organism leads on average to the preservation of a greater amount of the same material in another organism. Worker bees who attack predators approaching their nests die as a result of the act of stinging. The evolution of such behavior is explained by the fact that the beneficiaries of the altruistic act are close relatives of the worker (Michener 1974). Also sterile castes and helping have evolved in the social insects because sterile workers usually help their mother (the queen) to produce offspring (Brockmann 1984, Wilson 1971). Since the rapid development of Hamilton's ideas and the demonstration that a number of cases of sociality outside social insects also involve altruism among close kin, kin selection has become one of the most favored explanations in evolutionary socio-biology.

Kin selection requires an individual to behave differently towards individuals of different degrees of relatedness. In communal animals, this can be simply the result of living near one's relatives. There is also a growing body of evidence showing that individuals can indeed recognize kin and even distinguish close kin from distant kin. For instance, members of a social insect colony identify fellow members by colony-specific pheromones. But it is also clear that, in contrast to Hymenoptera, within social vertebrates' groups for instance most cases of altruism must account for transfers of fitness between non-sibs as well as between sibs. Since the strength of kin selection pressure rapidly weakens at a relational distance greater than that of half-sib, there is substantial reason to consider

that most social behavior, particularly in vertebrates, has been shaped by the combined effects of more than one selection principle. Therefore, in order to explain how cooperation and altruism evolve among *unrelated* con-specifics, one has to introduce reciprocity selection.

b) Reciprocity selection

When social evolution is considered, one cannot directly assign fitness to organisms viewed in isolation. As we already argued in the context of kin selection, what needs to be determined is inclusive fitness which goes beyond the physical environment and encompasses the social environment consisting of other con-specifics. Boorman and Levitt (1980) speak of fitness interlocking when the behavior of one individual directly affects the fitness of other individuals. They also note that sociality characteristically imposes fitness tradeoffs between different individuals. Some forms of cooperation, called mutualism by Krebs and Davies (1981), do not involve any altruism because each cooperating individual gains a net benefit from doing so. Pied wagtails joining together to defend a feeding territory enjoy a greater feeding rate than they would by being alone. In other cases, kin selection and mutuality work together. Lionesses in a pride are related so that not only does hunting in packs improve the chances of capturing a zebra but it also confers kin benefit to individuals. More generally, this is also the case with Trivers' (1971) altruism where both participants will gain as long as the help is reciprocated at some later date. However, only in some simple cases of social aggregation is it possible to ascribe a positive value to social participation for *all* members. Advanced cases of sociality impose tradeoffs between benefits to some participants and costs to others. To understand why such cases could be a problem, let us consider here a very simple model of evolution borrowed from Boorman and Levitt's (1980) "minimal model", recast into the framework of the Replicator Dynamics[3].

[3] This "minimal model" can also be recast into the framework of the Evolutionary Game Theory initiated by Maynard Smith (1982). The concept of an evolutionary stable strategy ESS usually leads to individuals programmed to play mixed strategies and to monomorphic population equilibria. There are known examples of mixed behavior that could be interpreted as an ESS. For instance, Brockmann and Dawkins (1979) studied the female great golden digger wasps (Sphex ichneumoneus) which lay their eggs in underground burrows that they have provisioned with grasshoppers as food for the larvae. The female wasp has two strategies open to her. She can either dig her own burrow, running a small risk of being invaded by another wasp, or she can enter an already dug burrow, saving herself the cost of digging but with the risk that the burrow is being used by the owner. The best strategy depends upon that adopted by other female wasps in the vicinity and clearly digger wasps employ a mixed strategy. This is corroborated by the fact that the success (in terms of the number of eggs laid) of the two pure strategies "Dig" and "Enter" is the same whether the wasp decides to enter an existing burrow or to dig her own. Nevertheless, to restrict ESS considerations to situations in which only a single type (possibly mixed) may exist at equilibrium may be unsatisfactory. In general, one would also like to explain polymorphic population equilibria that could arise in an evolutionary stable way.

Dawkins (1976, 1982) defines a *replicator* as anything in the universe of which copies are made, and considers that evolution is the external and visible manifestation of the differential survival of alternative replicators. A replicator is active if its nature has some influence over its probability of being copied. According to this approach, let us suppose that individuals are vehicles in which two active replicators travel about. The first one is a social replicator that programs individuals to play a pure social strategy *S*. The second one is an asocial replicator that programs individuals to play a pure asocial strategy *A*. Social and asocial fitness must now be defined. With each individual randomly paired with another member of the population, the individual fitnesses resulting from such pairwise contests are defined in Table 1 below (values in the table are the fitness to an individual hosting the replicator on the left while his opponent hosts the replicator above) In this table, it is assumed that:

- When a social individual meets another social individual, each has fitness $1+\sigma$ so that σ is the per capita benefit from membership in a partnership between socials
- When a social individual meets an asocial individual, the asocial receives $1+\sigma_1$ while the social individual has reduced fitness $1-\tau$
- When asocial individuals meet, both have fitness equals to 1.

Table 1 Payoffs in terms of individual fitness

Individual fitness	*social*	*asocial*
social	$1+\sigma$	$1-\tau$
asocial	$1+\sigma_1$	1

Therefore, in a polymorphic p-population (in which there is a fraction *p* of individuals hosting a social replicator and a fraction $1-p$ of individuals hosting an asocial replicator), the expected fitness conferred on an individual by the social replicator is $f_S(p)=(1+\sigma)p+(1-\tau)(1-p)$ and the expected fitness conferred on an individual by the asocial replicator is: $f_A(p)=(1+\sigma_1)p+(1-p)$. The Replicator Dynamics is simply determined by the comparison of these two frequency-dependent expected fitnesses. Because the fitness of a host is a measure of how frequently it gets to reproduce its replicator, replicators that confer high fitness to their hosts are going to control a larger share of hosts than those that confer low fitness. In our population, with $\sigma>\sigma_1$ exemplifying the synergistic case in which there is a more than additive advantage of cooperation (as opposed to the opposite

case $\sigma_1 > \sigma$ of a Prisoner's Dilemma), the Replicator Dynamics shows that there is a rest point $p^* = \tau/(\tau + \sigma - \sigma_1)$ solution of the equation $f_S(p) = f_A(p)$. But such a rest point which corresponds to a polymorphic equilibrium is unstable, as shown by the fact that for *p* greater (lower) than *p**, f_S is greater (lower) than f_A. Above *p**, the selection process will lead to social fixation. But below *p*,* it will lead to asocial fixation. Therefore, in a population of asocials, a small fraction of social replicators arising by mutation will be counter-selected. Then how is it possible that a social replicator appearing by mutation at a very low initial frequency can ever cross the critical *p** value?

Interestingly, Trivers' (1971) reciprocal altruism can also be reformulated in the framework of the Replicator Dynamics. Whenever the benefit of an altruistic act to the recipient is greater than the cost to the donor, both participants will gain as long as the help is reciprocated at a later date. The problem is of course the possibility of cheating, the recipient being able to refuse to repay the favor at a later date. Axelrod and Hamilton (1981) have shown in a Prisoner's Dilemma framework that a Tit-For-Tat strategy can be an evolutionary stable strategy in a population in which there is always a finite probability that two individuals will meet again, provided this probability is sufficiently large. Such a Tit-For-Tat strategy is simply "be social at the first encounter and then do whatever your opponent did on the previous encounter". This is clearly a strategy of cooperation based on reciprocity. Introduced in a Replicator Dynamics, one can thus consider a population in which there is a positive probability that two individuals will meet again and in which individuals either host an asocial replicator *A* or a Tit-For-Tat replicator *TFT*. With the payoffs previously defined in Table 1 for one encounter and with *v* the probability that two contestants meet again, one gets the following expected fitness[4] conferred on an individual by a Tit-For-Tat replicator and by an asocial replicator in a population with a proportion *p* of Tit For Tat replicators:

$$f_{TFT}(p) = \left(\frac{1+\sigma}{1-v}\right)p + \left(\frac{1-\tau(1-v)}{1-v}\right)(1-p)$$

[4] With *v* the probability that two contestants meet again :

When a *TFT* plays another *TFT*, it gets: $(1+\sigma) + v(1+\sigma) + v^2(1+\sigma) + ... = \frac{1+\sigma}{1-v}$; whereas when it plays an *A,* it gets $(1-\tau) + v + v^2 + ... = (1-\tau) + \frac{v}{1-v} = \frac{1-\tau(1-v)}{1-v}$.

When an *A* plays a *TFT,* it gets: $(1+\sigma_1) + v + v^2 + ... = (1+\sigma_1) + \frac{v}{1-v} = \frac{1+\sigma_1(1-v)}{1-v}$; whereas when it plays an *A,* it gets $1 + v + v^2 + ... = \frac{1}{1-v}$.

$$f_A(p) = \left(\frac{1+\sigma_1(1-v)}{1-v}\right)p + \left(\frac{1}{1-v}\right)(1-p)$$

Clearly, $f_{TFT}(p)$ is greater (lower) than $f_A(p)$ if and only if p is greater (lower) than p^*, with: $p^* = \frac{\tau(1-v)}{\tau(1-v)+\sigma-\sigma_1(1-v)}$. Once again, there is a rest point p^* that corresponds to a polymorphic unstable equilibrium if $\sigma - \sigma_1(1-v) > 0$ which means that: $v > \frac{\sigma_1 - \sigma}{\sigma_1}$. This is trivially verified when $\sigma > \sigma_1$, the more-than-additive advantage case of cooperation. This could also be verified in the more extreme case of the Prisoner's Dilemma, when $\sigma_1 > \sigma$, provided that the probability v that two contestants meet again is sufficiently large. But also note that in the more-than-additive advantage case of cooperation, p^* is a decreasing function of v. Therefore, the threshold p^* to be crossed before the *TFT* replicator can invade the population can be low when the probability v is sufficiently large.

2 From animal societies to human societies

According to Hamilton (1964), altruism is expected to be selectively directed toward kin and close kinship is expected to facilitate altruism. As we have seen, there is evidence that kin are responsible for much of the altruism deployed in animal species. To quote some further examples, in primate groups (Silk 1986) cercopithecine females selectively defend and support maternal kin against aggressive encounters and reserve costly aid for close kin. Also, grooming (the removal of ectoparasites and dirt from skin and hair) is primarily directed to kin. But we also know that there are several examples of altruistic behavior among unrelated individuals (Kreps and Davies 1981, Silk 1986). In many cases, such examples are instances of Trivers' reciprocal altruism such as blood sharing in vampire bats, eggs trading in black hamlet fish or alliances in primates (grooming among non-kin increases the probability of future support in aggressive encounters).

We have also seen that in a large randomly mixing population, even if it pays a species to acquire a social or cooperative trait, a frequency threshold has to be crossed that does not seem to permit to achieve the advantageous equilibrium starting from a mutation whose frequency is low. To explain animal social behavior, different evolutionary scenarios are evoked. The first one considers that cooperation could emerge at first between relatives, evolving by kin selection. Because one of the

cues favoring the recognition of relatedness could simply be the fact of reciprocation of cooperation, reciprocity cooperation could grow between relatives and thus cross the frequency threshold (Axelrod and Hamilton 1981). Independently of such an initial propagation on a kin basis, a second scenario concentrates on clusters of cooperation. This is the *cascade principle* of Boorman and Levitt (1980) according to which, in an appropriately viscous population structure, a local concentration or cluster of the social gene may favor its propagation through interactions which are more frequent than expected from random encounters. More generally, it is quite clear that, by endowing a newly introduced social trait with enough recognition capabilities to enable the socials to recognize each other with great accuracy, one can ensure the successful takeover of such a trait. If individuals hosting *TFT* replicators and individuals hosting *A* replicators are perfectly distinguishable from each other, *TFT* individuals can now interact with one another, getting a payoff: $\frac{1+\sigma}{1-v}$, and asocials are left to interact with one another, getting a payoff: $\frac{1}{1-v}$. This allows *TFT* replicators to invade the population. In many cases, it may be possible for some species to exploit substitutes for innate recognition such as barriers to dispersal, which not only make neighboring con-specifics extremely likely to be kin but also permit the persistence of a local concentration of a social gene.

Thus, as pointed out by Gintis (2000a) using a result from the group selection approach, the usual argument according to which evolution should entail the disappearance of the altruist because such an individual becomes less fit in the process of rendering the group more fit is not necessarily the end of the story. Suppose that in a population there are groups i =(1,…, n), and let q_i be the fraction of the population in group *i*. In each group, there is a social trait with frequency s_i that contributes to the mean fitness f_i of group *i*. At a given period, one therefore has $\bar{s} = \sum_i q_i s_i$ the frequency of the social trait in the population and $\bar{f} = \sum_i q_i f_i$ the mean fitness of the whole population. Now, if we consider that from one period to the next, groups grow in proportion of their relative fitness, at the next period one has $q_i^{'} = q_i \frac{f_i}{\bar{f}}$ and $s_i^{'} = s_i + \Delta s_i$ for the frequency of the social trait in the group *i* at the next period. Therefore, the variation $\Delta \bar{s}$ of the frequency of the social trait in the whole population is determined by: $\Delta \bar{s} = \sum_i q_i^{'} s_i^{'} - \sum_i q_i s_i$, which can easily be rewritten as:

$$\bar{f}\,\Delta\bar{s} = \sum_i q_i (f_i - \bar{f}) s_i + \sum_i q_i f_i \Delta s_i$$

This is Price equation, well-known in biology. It shows that, despite the fact that the social trait renders individuals bearing it less fit than other group members $(\Delta s_i < 0)$, the frequency of the trait in the whole population can nevertheless increase $(\Delta\bar{s} > 0)$. With $\Delta s_i < 0$ for all *i*, the second term in the right member of the equation (the within-group selection effect) is clearly negative. But provided that the first term is positive and sufficiently large, $\Delta\bar{s}$ can be positive. Because $\sum_i q_i (f_i - \bar{f})\bar{s} = 0$, the first term can also be written as $\sum_i q_i (f_i - \bar{f})(s_i - \bar{s})$ a between-group selection effect represented by the covariance between group fitness and group frequency of the social trait. Thus, with sufficiently high covariance (when groups with above-average frequency of the social trait also have above-average fitness), the social trait can increase in the population. Interestingly, applied to a population of both social and asocial individuals pairing off in each period with payoffs given for instance by a Prisoner's Dilemma defined by Table 1 with $\sigma_1 > \sigma$, the Price equation shows that a small number of socials can invade a population of asocials. This occurs (Gintis 2000a) if there is a sufficient degree *r* of assortative interaction (*r* being the probability that each type meets its own type and $1 - r$ being the probability to meet a random member of the population).

Nevertheless, disregarding the complex of issues and models that can evolutionarily explain social behavior in animal species, the fact remains that empirical and theoretical evidence combine to suggest that there are multiple and often delicate conditions for evolutionary emergence of social adaptation (Boorman and Levitt 1980). For perhaps one million of presently existing animal species, at most ten thousand can be considered as social in any significant way. This fact has be related to other factors that have not yet been considered - *learning*, *rational decision* and *cultural transmission* – the existence of which is generally considered important for explaining social behaviors in the human species.

a) Learning and rational decision

Evolutionary biologists and ethologists noticed that, as we move up the evolutionary ladder into the higher animals and toward humans, intra-specific variability in social adaptation becomes substantial. Social vertebrates are for instance a more heterogeneous collection of species than social insects.

This variability reflects the increasing importance of environmental, as opposed to genetic, factors. Surely, in an uncertain world, genes are more likely to be successful if phenotypes are free to adapt flexibly to the environment in which they find themselves[5]. And, like many other organisms, humans adjust their phenotypes in response to their environment through individual learning and rational calculation.

All organisms appear to possess mechanisms that allow them to modify their phenotypes adaptively in response to environmental contingencies. When an animal learns, it changes its behavioral repertoire forever. Such a change is likely to alter its fitness and therefore is subject to natural selection. From Pavlov's work on conditioning to Thorndike and Skinner's approaches to instrumental learning, it is well known that, given a criterion of reinforcement such as a sense of pain or a taste for rewards, even random errors in behavior can be conditioned to elaborately adaptive behavior. Besides such simple learning, whether animals exhibit intentional behavior that would give them greater flexibility remains a controversial issue. Certainly, as our knowledge of animal behavior has improved, the difference between humans and animals has appeared to diminish (McFarland 1985). Nevertheless, even if one can adopt an intentional stance to predict some animals' behavior, treating them as "rational agents with beliefs and desires and other mental stages exhibiting rationality" (Dennett 1987: 15), it seems that human beings differ not only quantitatively but also qualitatively from animals as far as intentionality is concerned. According to Lumsden and Wilson (1981) for instance, *reification* – a process by which the human mind produces concepts and continuously shifting reclassifications of the world – is a unique activity that fully separates mankind from the most advanced social animal species[6].

In order to understand the importance of intentionality and rational decision for cooperative behavior, let us return to our society of individuals hosting social and asocial replicators, with payoffs displayed in Table 1. If both types are perfectly distinguishable from one another, we saw that social replicators can interact selectively with another at no cost and invade the population. But suppose

[5] Investigating the biological basis of economic behavior and asking the questions: "Why might utility functions exist? Are they adaptive?" Robson (2001, p. 13) notes that "the evolutionary rationale for an hedonic internal evaluation system is to permit an appropriate response to novelty and complexity". With his distinction between homo economicus (who can only be manipulated via his preferences) on one hand and homo behavioralis (programmed directly with behavior that Nature can directly manipulate) on the other hand, Binmore (1994: 151) emphasizes the same point. He notes that in a principal-agent problem, with a rapidly changing environment, Nature has to choose homo economicus rather than homo behavioralis because the former adapts more quickly than the latter. This is why she made an expensive investment in brainpower.

[6] This was linked to the enormous increase of the cerebrum of man during a relatively short span of evolutionary time. Over a period of approximately three million years, the brain tripled in size so that no scale has really been invented that can objectively compare the intelligence of man to that of chimpanzees or other primates (Wilson 1975).

now that the two types cannot be distinguished at a glance. An individual has to pay, in terms of fitness, a cost of scrutiny *c* if he wants to know the type of the individual with whom he randomly meets. If not paid, the two types are indistinguishable. Thus, an individual hosting a social replicator and paying the cost of scrutiny will interact only with another individual of the same type and his payoff will be equal to: $1+\sigma-c$. And if he refuses to pay the cost, his expected payoff will be: $p(1+\sigma)+(1-p)(1-\tau)$, denoting by p proportion of socials in the population. Thus, it is plain that for: $p < p^* = \frac{\sigma+\tau-c}{\sigma+\tau}$, every social will pay the cost of scrutiny and get a payoff: $1+\sigma-c$ by interacting with another social. Moreover, it will not be in the interest of the asocials to bear this cost because no social will want to interact with them anyway. Therefore, if $\sigma > c$, starting below p^*, a very small proportion of intelligent socials arising by mutation and rationally deciding to pay the cost of scrutiny will expand in the population until they reach a population's share equal to p^*. Above this value, it no longer makes sense to pay the cost of scrutiny. Socials and asocials will then interact at random and the fitness payoff difference will cause the proportion of socials to shrink. The Replicator Dynamics leads here to a stable polymorphic equilibrium characterized by a proportion p^* of socials in the population.

By introducing such a rational decision element in a simple evolutionary model, Frank (1988) wanted to emphasize the fact that rational calculations play only an indirect role in solving social dilemmas. Just consider two self-interested persons who can engage in a potentially profitable venture, but with each having opportunities to cheat. The payoffs are presented in Table 2, similar to Table 1 but with: $\sigma_1 > \sigma$.

Table 2 Payoffs in a potentially profitable venture

payoff	*not cheat*	*cheat*
not cheat	$1+\sigma$	$1-\tau$
cheat	$1+\sigma_1$	1

In this Prisoner's Dilemma, both persons would profit from making a binding commitment not to cheat. But, once the venture is under way, self-interest guided by material incentives dictates

cheating. Suppose now that there are persons in the population who are unable to cheat because of strong feelings of guilt. In order for such a non-cheater to benefit in material terms, she must both be able to be recognized as a non-cheater and to recognize non-cheaters with a sufficiently low cost of scrutiny to make the venture still profitable. Thus, a genuinely trustworthy person must be observably different in a way that is partly insulated from purposeful control in order to solve the commitment problem – to defeat opportunists that could attempt to mimic the symptoms of trustworthiness. Without such a recognition mechanism, rational decision cannot solve the social dilemma except if the venture leads to repeated interactions with a sufficiently high probability. When both persons are involved in repeated interactions, indefinite cooperation is a possibility based on the threat that if someone ever deviates, the opponent will punish him with cheating thereafter. With a discount rate v representing the probability that the venture will remain constituted for at least one more period, this implies: $v > \frac{\sigma_1 - \sigma}{\sigma}$. Therefore, in cases where socials do not have sufficient capabilities to recognize each other or in cases where a venture between self-interested agents disbands with high probability, cooperation among individuals cannot necessarily be sustained. Some research in evolutionary psychology has suggested that humans may be evolutionary predisposed to engage in social exchange using mental algorithms which identify and punish cheaters (Cosmides and Tooby 1992, Hoffman, McCabe and Smith 1998)[7]. And, of course, if intelligence partly evolved from the need to interact with fellow beings, to have a "theory of mind" – a model of the beliefs and preferences of others – can permit a flexible response in strategic situations and so would be evolutionary favored (Robson 2001).

b) Cultural transmission

If humans share their proclivity to adjust phenotypes in response to their environment through learning and rational calculation with many other organisms, they are almost unique in their ability to culturally transmit the so acquired phenotypes to the next generation. In non-cultural species, even with a large range of individual learning spanning from trial-and-error to rational choice, variants so acquired and other forms of phenotypic flexibility are lost with the death of the individual because inheritable changes of phenotypical features are only accessible via the one-way road of genotypes. In such species, a given distribution of genotypes G_t in the population gives rise to a distribution of

[7] Barnett (1968) has noted that the use of punishment in the attempt to train their young in anything other than avoidance seems exclusively limited to humans.

phenotypes F_t through a process of ontogeny (including individual learning). Natural selection, acting on the phenotypic characteristics of individuals, leads to a modified distribution of genotypes G_{t+1} that gives rise to the distribution of phenotypes in the next generation F_{t+1}. However, since all the phenotypic variants acquired by individual learning are lost for the next generation and have to be learned again, one has only to know G_{t+1} to predict the distribution of phenotypes F_{t+1} in a given environment. It is not necessary to know F_t. In cultural species, things are very different (Boyd and Richerson 1985). In general, *cultural exchange* is defined as the passage of information capable of affecting individuals' phenotypes from one generation to the next by non-genetic means (McFarland 1985). In this case, it is no longer sufficient to know G_{t+1} in order to predict the distribution of phenotypes F_{t+1}. One must also know which cultural traits of F_t are transmitted to the next generation through cultural transmission mechanisms. Vertical (between generations) cultural transmission combined with individual learning thus acts to create a Lamarckian effect by which acquired variation can be inherited[8]. This Lamarckian effect introduces a force of variation in which the origination of novel traits does not result from the play of chance but from "an exercise of will on the part of individuals in actively responding to perceived needs, which they do by initiating constructive adaptations that are subsequently transmissible to offspring" (Ingold 1986).

There is no doubt that biological explanation of cooperation based on kin altruism and reciprocal altruism may apply to human and nonhuman species alike. But there is also no doubt that human cooperation is based in part on capacities that are unique to Homo sapiens. As we have seen, the Price's equation shows that in populations composed of groups characterized by a higher level of interaction among members than with outsiders, the evolutionary process may be decomposed into between-group and within group selection effects. For a social trait whose expression benefits the group, but imposes fitness loss on those who adopt it, it follows that the former effect can offset the latter effect when circumstances heighten and sustain differences between groups relative to within-group differences. Therefore, according to many contributors in this field[9], if we want to seek an explanation of cooperation that works for humans and does not work or works substantially less well for other species, we must in particular look for distinctive human characteristics than enhance the relevance of group selection for humans. And central to this relevance are psychological and cultural

[8] Besides vertical cultural transmission, there is also a horizontal cultural transmission (within generations, because adults may copy adults and children may imitate other children) that contributes to the distribution of phenotypes (Cavalli-Forza and Feldman 1981).

[9] See for example Bowles and Gintis (2003), Gintis (2000a, 2000b), Gintis (2003), Gintis, Bowles, Boyd and Fehr (2003)

human capacities to suppress within-group phenotypic differences and simultaneously sustain a high frequency of intergroup conflicts.

Cultural transmission leads people to internalize norms of behavior through vertical or horizontal socialization and these norms are followed principally because people value the transmitted behavior for its own sake, in addition to or in spite of its effect on personal fitness or well-being. Thus in cases where cooperation between self-interested people cannot be sustained, an internalized norm of cooperation (individually fitness-reducing but fitness-enhancing at the group level) may be a considerable benefit to a group. Insofar as such norms are largely widespread inside the groups, they contribute (like many other constructed institutional environments) to limit within-group competition and to reduce phenotypic variation within groups. Correlatively, the formation of groups on such non-kin characteristics limits the between-group migration allowing the possibility that group selection pressure can co-evolve with cooperative behaviors because within-group cooperation and hostility toward outsiders co-evolves.

They are some convincing examples of such internalization of norms because recent experimental research has revealed forms of human behavior difficult to explain in terms of self-interest (Fehr and Gächter 2000, 2002, Falk, Fehr and Fischbacher 2001, Fehr, Fischbacher and Gächter 2002). Strong reciprocity, for example, is a predisposition to cooperate and to punish those who violate the norms of cooperation, at personal cost and even when it is very implausible to expect that these costs will be repaid. There are many ways to evolutionarily explain strong reciprocity. Using Price equation to chart the dynamics of strong reciprocity, for example, Gintis (2000b) shows that for a sufficient amount of harm which an individual can inflict on non-cooperators at a sufficiently personal low cost of retaliation, a small fraction of strong reciprocators can always invade a population of self-interested agents when group extinction threats are relatively common. Besides cultural arguments linked to the internalization of norms that point to the decline of cost disadvantage of retaliation as defectors become rare, Gintis points out that, contrary to animal dispute for which victory often involves great cost even to the winner, Homo sapiens has the superior ability to inflict punishment at a low cost to the punisher. This is perhaps one of the most interesting properties that contributes to cooperation inside groups in human societies. Moreover, under assumptions approximating likely human environments over the 1,000,000 years prior to the domestication of animals and plants, agent-based simulations shows that the proliferation of strong reciprocators is very likely (Bowles and Gintis 2003).

Is it possible to say that, because humans acquire so much of their behavior culturally rather than genetically, the human evolutionary process is fundamentally different from that of other animals? This question is highly controversial. Many examples of cultural exchange and tradition in animal species show that culture and tradition do not necessarily require great intelligence on the part of individuals (McFarland 1985, Nishida 1986). Thus, in animal species, individual behavior is certainly under the control of two sets of instructions, genetic and cultural, with culture itself under genetic control. For people who think that individuals are the products of gene pools and cultures, and that humans do not cease to be animals with the advent of culture, this has led to a number of formal models of so-called gene-culture co-evolution (Cavalli-Sforza and Feldman 1981, Lumsden and Wilson 1981, Boyd and Richerson 1985, Gintis 2003) These models are designed to show how genetic and cultural evolution can interact through programs of individual developments. Because culture is an inheritance system that makes a pool of cultural traits to co-evolve with the gene pool, one cannot abstract from the details of cultural transmission that are likely to be essential for understanding the social evolution of human behavior. In such a framework, culture may have a variety of structures (patterns of socialization by which a given set of traits is transmitted in a given society) and one has to understand the conditions under which different structures of cultural transmissions might evolve. Cultural transmission leads to the persistence of behavioral traits through time. Because not every individual is equally invented and because experimenting directly with the environment may be dangerous, tradition or culture can be a cheaper and safer way of acquiring information. Thus, "in evolving a reliance on cultural transmission, the human species may well have traded high rates of random error caused by individual learning in variable environments for a lower rate of systematic error (with respect to genetic fitness) due to the partial autonomy of cultural evolution" (Boyd and Richerson 1985:289).

III Social learning

1. Theoretical perspectives

Pro-social behavior encompasses in social psychology any voluntary action which aims to benefit an other (see Eisenberg, 1996). Altruistic behavior is generally considered as a subtype of pro-social behavior, and motives underlying pro-social behavior may be altruistic or not. Pro-social behavior includes helping, sharing, giving, and overlaps with moral behavior. This subsection aims to describe the main theories that argue that pro-social attitudes are taught by adults and learned by children.

Such hypotheses are not absent from economic literature. For example, Bisin and Verdier (2001) study population dynamics of preference traits in a model of intergenerational cultural transmission. Although their paper is not specially designed to study how altruistic or reciprocal preferences can be sustained in a group, especially when these are not dominant cultural traits, their model can be applied to this kind of preferences. They assume that parents socialize and transmit their preferences to their offspring, who can be socialized either by their parents or by the society (cultural and social environment). The parents are altruistic toward their children, and thus might want to socialize them to a specific cultural model if they think this will increase their children's welfare. However, this altruism is only imperfect, as parents can only use their own preferences to evaluate their children's choices[10]. The authors study the long run stationary state pattern of preferences in the population, assuming that family and society are substitutes in the transmission mechanism. Parents will socialize children more intensively when the set of cultural traits they wish to transmit is common only to a cultural minority of the population. Those parents who belong to a cultural majority will be able to save on their own resources to socialize their children since they anticipate that the latter will adopt with high probability the cultural traits of the majority that they themselves wish to transmit. Bisin and Verdier show that such mechanisms are inefficient because parents invest too many resources to affect their children's preferences.

In the same vein, based on Duesenberry's work (1949), Cox and Stark (see Stark 1995) propose to explain how altruistic preferences might be passed on by the parents to their children, by demonstration and imitation. In order to make their children aware that they will have to help them in the future, by means of services or possibly monetary transfers, they conspicuously help their own old parents in front of their children's eyes, that is they demonstrate to their children how they behave with their own parents with the hope that their children will imitate this attitude in the future.[11]

Those economic models are to be related to an important trend among social psychologists: the social learning school. Bandura (e.g., 1986), is probably the most important representative of this school, rooted in behaviorism, that postulates that moral behaviors are mainly induced by modeling

[10] The terms used by Bisin and Verdier (2001) to designate this form of altruism are paternalistic altruism or imperfect empathy. Anticipating on section 5, it is a clear case of projection of parents onto their children.
[11] This model has received mixed empirical support, in particular when the probability of imitation is endogenous (see Jellal and Wolff, 2002; see also Arrondel and Masson, 2001, for other empirical evidence by economists). See also Chapter 11 of this Handbook for a critical presentation of this hypothesis.

and learned by imitation: children learn to behave pro-socially by imitating models (generally adults, but also peers), who behave pro-socially.

2. Empirical findings: the role of the family

These hypotheses have received mixed empirical support. Psychologists have investigated how socialization within and outside the family may induce pro-social behavior among children. In particular, they have studied the effects of two main parental disciplinary practices: induction (parents give explanations or reasons for requiring the child to change his behavior) and power-assertive or punitive techniques (physical punishment or deprivation of privileges). Punitive techniques appear to be generally unrelated or negatively related to children's pro-social development. In particular, immediate compliance has often been observed, but effects generally disappear over time (see, e.g., Grusec 1981), although social disapproval, compared to material punishment, may have a positive effect. Researchers have found a positive relation or no relation between parental use of inductions and pro-social responding. When various types of induction are considered separately, there is at least some evidence of a relation between pro-social behavior or sympathy and inductions focused on the state or the feelings of others (see Eisenberg and Fabes (1998) for an extended review of this literature).

Another related point of importance in enhancing pro-social behavior is the quality of the parent-child relationships. Once again, empirical findings are contrasted. Most studies, but not all, have found a positive relation between warm socializers and pro-social children, just as between parental empathy and children empathy. Moreover, support for a positive relation between parental emphasis on pro-social values and children's pro-social responding is mixed (for example, Hoffman (1975b) and Eisenberg et al. (1992)). Oliner and Oliner (1988) report nevertheless that people who rescued Jews during the Holocaust often recall learning values of caring from parents. In the same way, adults involved in civil rights activities often report that their parents were themselves involved in altruistic or social activities and discussed their altruistic involvement with their children (Rosenhan, 1970). Thus affection of parents and their altruistic values seem to be determinants of children's acquisition of altruistic behavior, as argued by Hoffman (1975b). For example, Chase-Lansdale et al. (1995), investigating the construct of caring, argue that families are instrumental in the promotion of caring through attachment, peer relationships, pro-social behavior, empathy, agency and self-control, and review empirical evidence supporting this hypothesis.

As mentioned above, another way to induce pro-social behavior by socialization is modeling. Most studies involving models are laboratories studies. Experimenters have generally implemented a kind of dictator game under two controls: before donating, the child views or does not view a model. Results indicate that those who view a generous or helpful model are more generous or helpful than children who view no model, as are subjects who view a generous rather than a selfish model. Moreover, multiple models seem more effective than single or inconsistent models, and the more generous the model is, the more effect he or she seems to have. Some researchers have also found that children imitate rewarded models. Besides, some models, in particular those who control valued resources or are perceived as competent, are more imitated than others. In real life, evidence has been found that children model parents' pro-social behaviors, but Eisenberg and Fabes (1998) indicate that the data are scarce and correlational.

Psychologists have also examined the effects of non-disciplinary verbalizations. Statements of intentions appear to have less effect than does directly viewing the model, although they may foster generosity even a few months later. Except in certain situations, preaching and exhortations seem to have little effect, although preaching emphasizing the emotional consequences of the pro-social act seems more effective. Directives are generally effective and often last, but efficiency depends on the nature of the directive, and on the age of the child. Moreover, assigning responsibilities to a child appears to have a positive effect.

Concrete and social reinforcements have often been found to increase children's pro-social behavior at least immediately, although they may have a negative effect in the long run by undermining intrinsic motivations (Lepper 1983). Ten-year-old children generalize socially reinforced pro-social actions to new situations, whereas younger children don’t (Grusec and Redler 1980).

Other techniques of fostering pro-social behavior have also been investigated. Provision of internal attribution, for example by telling the children that they are helpful, has a positive effect compared to no provision and to attributing pro-social behavior to the fact that it was expected. Observations generally support the mediation of an enhanced pro-social self-image (Grusec and Redler 1980). Moreover, according to Staub (1992, quoted in Eisenberg and Fabes (1998); see also the same review for empirical references), children's participation in pro-social activities seems to enhance pro-social behavior in the long run (learning by doing), although boys sometimes show some reactance in the short run. For Staub (1971), teaching by assigning responsibility first focuses responsibility on the child externally; then the desire to help others in need may be internalized.

3. The role of other socializers

Other socializers have received little attention. Children, even 1- to 2-year old, exhibit pro-social behavior toward their siblings, although the findings related to the effect of rank of birth on pro-social behavior are inconsistent. Moreover, mother's behaviors are related to pro-social behavior between siblings, although mother's unavailability has been found positively related to pro-social behavior of older, specially daughters, toward their young siblings. This influence of peers is complex and may be interpreted in different ways. We will return later on this point.

Next, little is known about the effect of school program, in particular because of the low frequencies of pro-social behavior observed in the classroom and because pro-social behaviors are rarely reinforced or encouraged by teachers. Comparisons between children who attend school and children who don't are equivocal. However, Eisenberg and Fabes (1998) mention a few studies conducted in Israel showing that more pro-social behaviors are observed when there is age heterogeneity in the classroom and when cooperation and individualized learning are encouraged, as in certain kibbutz. Some evidence of the effects of the child-teacher relationships has also been found. Last, Eisenberg and Fabes (1998) indicate that some natural experiments, in particular programs to enhance pro-social values, behaviors and attitudes, appear to be partly effective. Staub (1981) also emphasizes the importance of learning by doing or by participation that could be enacted in school.

Last, television can be counted among other "socializers", or at least as displaying models (see, for example, Rushton, 1981). Most studies have investigated effects of violence on aggressive behaviors, but some have also examined the effect of pro-social models on pro-social behavior. Hoffman (1988) argues that empirical evidence is equivocal and inconsistent, in particular because it takes longitudinal data to conclude on causality.

4. Cross-cultural differences

In relation with this family of hypotheses, it is worth investigating cultural factors influencing pro-social development. As asserted by Eisenberg and Fabes (1998), psychological research on this subject is relatively sparse, while societies seem to greatly vary in the degree to which pro-social and cooperative behaviors are normative. In fact, even sub-cultural variations may be important. For example, Eisenberg and Mussen (1989) indicate that, in the United States, children from traditional rural and semi-agricultural communities and from relatively traditional subcultures (Mexican

American children) are more cooperative than children from urban and westernized cultures. Other studies by Kagan and Knight (see references in Eisenberg and Fabes, 1998), in which children are asked to share chips, confirm this pattern of results. However, no consistent evidence appears to exist on this subject. Several studies (quoted by Eisenberg and Fabes, 1998) show that Israeli kibbutz children and Israeli city children do not differ in sharing behavior at the age of 5, although fifth-grade Israeli kibbutz boys (but not the girls) share more than their city equivalents. Only few differences in moral reasoning values and beliefs about social responsibilities exist among industrial Western cultures, whereas differences between Western and non-Western cultures have been found (see Turiel (1998) and, for example, comparisons between Indians and Americans by Miller and Bersoff (1992) and Miller et al. (1990)). Differences may simply reflect differences in degree to which helpfulness and social responsibilities are emphasized (Eisenberg and Fabes; 1998).

5. Conclusion

Although learning and enforcement theories have received some empirical support, many psychologists argue that these factors cannot fully explain the acquisition and development of pro-social behavior. As asserted by Krebs and Van Hesteren (1994), "[t]here is relatively little disagreement among developmental psychologists that children construct their social worlds in terms of cognitive structures" (p. 107), and not only by copying "whatever the environment presents to them" (Flavell, 1992, p. 998).

IV. Cognitive theories of moral and pro-social development

Jean Piaget was probably one of the first to acknowledge the importance of cognitive factors in the development of moral judgment and moral behavior. However, the most interesting feature of his work appears to be the assumption that children learn to behave pro-socially by interacting with their peers, and not because they have been taught to behave in this way. In his 1932 book, *Le jugement moral chez l'enfant* (*The moral judgment of the child*, 1997), Piaget, before studying how children form moral judgments, showed how the rules of a game are used by the children, and how conscious of the rules they are. In this study, Piaget mainly observed boys playing a game of marbles, and asked them to explain him the rules of the game. Although not directly related to pro-social development, this preliminary step helps to understand how children develop and perceive rules and how they comply with them.

1. How children use the rules of a game and how conscious of these rules they are

Piaget distinguished four stages concerning how the rules are put into practice. The first one is purely individual and motor: the child manipulates the marbles according to his desires and motor habits. The egocentric stage appears between the age of 2 and the age of 5, when the child receives from the outside the example of codified rules. He imitates the models, but either plays alone or plays with other children without seeking to win or to standardize the different ways of playing. Around the age of 7 or 8, appears a form of cooperation. Every child tries to win, so that appears the need for mutual control and unification of the rules. Understanding of the rules remains vague and information about the rules given by the children is still different, if not conflicting. By the age of 11 or 12, the rules are fully codified, the games are fixed up in minute detail, and everybody knows the rules.

In the same fashion, Piaget distinguishes three stages in the understanding and respect of the rules. First, the rule is motor, thus not coercive. Then, it is considered as sacred and intangible. It is of adult origin, and unchangeable (from the point of view of the practice of the rules, this stage corresponds to the second half of the egocentric stage and to the first half of the cooperative stage). Lastly, the rules are seen as emanating from mutual agreement. Thus they are compulsory to the children themselves, but can be transformed provided that the modification wins general agreement. In the second stage, the coercive rule is obeyed out of the hierarchical respect that the children have for their parents. In the third stage, the rule is obeyed because of the mutual respect the children have for each other.

2. How children form moral judgments

Piaget next studied how children form moral judgments. Although not directly concerned by giving or redistribution, Piaget investigated moral dilemmas like theft, clumsiness, and untruthfulness. Unlike the study of the rules of the marbles game, in which Piaget was able to observe the children playing the game and to question them on the rules, the analysis on the formation of moral judgments hinges only on discussions with the children, mainly on their reactions and judgments about told stories and situations presented by Piaget and his collaborators. This is an important limitation of his work, but judgments in moral dilemma are of interest even though one cannot be sure that children would act as they prescribe. In fact, Piaget noticed many times that verbal thought and conceptualization are generally behind action, that is children relate that they previously have

acted as would prescribe children older than they are. As Piaget studied only children between 6 and 12 (sometimes 13), because younger children may encounter difficulties understanding the stories, no motor stage is observed. Piaget showed quite convincingly that one can observe two stages in moral development, corresponding to two worlds of constraint, and to two types of respect.[12]

According to Piaget, in the first stage of "moral realism", the child is under the adult constraint. The morality is essentially heteronomous. The moral rules are external to the child: the just is what conforms to the rules enacted by the adults. In particular, disobedience is always unfair. Among younger children, intentions are generally not taken into account. Rightness or wrongness of an act is judged only on the basis of the magnitude of its consequences, because any deviation from the rule results in punishment. At this stage, the children obey the rules because they respect their parents (more generally adults), so that this stage is characterized by unilateral respect[13] (see Turiel (1998) for empirical references contrasting the Piagetian view of children's understanding of authority relations, showing that children take into account the nature of act commanded, and the attributes (like social position) of persons giving orders).

In the second stage of "morality of reciprocity", the morality is fully autonomous. As explained by Piaget (p. 157[14]): "The conclusion we will reach is that the feelings of justice, although they can of course be strengthened by the precepts and the example of adults, are, for a large part, independent of these influences and only need mutual respect and solidarity between children to develop." Piaget

---

[12] Eisenberg (1986, see also Eisenberg and Mussen, 1989) distinguishes five levels in the development of thinking about pro-social moral questions. Some studies have found a positive relation between the level or the stage of moral reasoning and the tendency to behave pro-socially, among adults (see, for example, Underwood and Moore, 1982) and sometimes among children. Note that empirical evidence may be limited for children, perhaps because the range of moral stages is more limited among them, so that the relation is probably moderated by other factors, like sympathy (Miller et al., 1996). Moreover, the correlation appears to be greater when the moral dilemma concerns sharing or helping (see Eisenberg, 1986, for a review) and when dilemma and pro-social behavior are similar in content (Levin and Bekerman-Greenberg, 1980). Among preschoolers, the positive relation was clearer for spontaneous sharing behaviors than for helping or responding to a peer's request behavior. Among elementary or high school students, pro-social behavior involving high costs (donating money or time) has been found more frequently associated with moral reasoning than low-cost behavior (helping). According to Eisenberg and Shell (1986), the reason is that the latter is performed automatically (as discussed in section 8.3) whereas the former might entail cognitive conflict (as discussed in section 6.2). In addition, "types of reasoning that clearly reflect a self- versus other-orientation and are developmentally mature for the age-group are likely to predict pro-social responding." (Eisenberg and Fabes 1998:732).

[13] Piaget drew from Bovet [see Piaget], who proposed two necessary and sufficient conditions for the appearance of the consciousness of the duty: that an individual receives orders from another, and that the former respects the latter. This is in opposition with the Kantian position according to which respect for others follows from the fact that rules are regarded as compulsory.

[14] Our translation. The page numbers refer to the 1973 French edition.

distinguished in fact two types of justice: retributive justice, a notion inseparable from the notion of sanction, and distributive justice, which only implies the idea of equality. In the first stage, sanctions are considered by children as just and necessary and the more severe the more just. Among older children, although the first kind of opinion subsists among them and even among adults, Piaget noticed that sanctions are not a moral necessity. Only those which require a "restoration", making the guilty party conscious of the consequences of its act, or which are a reciprocal treatment, are fair.

Piaget argues that those distinctions have important consequences from an educational point of view. Blame and explanations are viewed as more efficient than sanctions. This allows discriminating between expiatory sanctions, which are related to the constraint which presses on the child and arbitrary, and reciprocal sanctions which emphasize that the social ties have been broken. Piaget showed that younger children referred more frequently to expiatory sanctions, and older children to reciprocal sanctions. These two types of attitudes can be related to the two types of morality. As noticed by Piaget, the first one probably originates from instinctive reactions of the child (the compassion and vindictive tendencies observed among children) but it is first and foremost shaped by the adults. Afterwards, the transition to the second type is a particular case of the general evolution from the unilateral respect to the mutual respect. Hence, even if, at the beginning, the idea of reciprocity appears as a sort of tit-for-tat, the material element of punishment tends to disappear. Thus, as the children begin to interact, mutual respect develops between peers.

When obedience and equality conflict, Piaget showed that the youngest always say the adults are right, whereas the oldest defend equality, even if it is in opposition to obedience. In fact, the unilateral respect seems to raise obstacles to the free development of the feeling of equality even though the parents attempt to instill this feeling into their children, first because no equality is possible between parents and children, and second because equality among children cannot be dictated. Piaget gave an interesting answer to the question of why the democratic practice is so developed in the game of marbles, played by boys of 11 to 13, whereas it is so unfamiliar to adults. For him, an explanation is that these boys have no seniors to impose rules, as the game is generally dropped by the age of 14, whereas adults in many spheres of life are subject to the weight of previous generations. It is also interesting to note with Youniss (1980) that 6 to 14-year-old children define pro-social behaviors as giving, sharing, playing, when directed towards peers, but as being polite, obey, etc. when directed towards adults (parents).

In fact, Piaget distinguished three stages in the development of distributive justice, by differentiating between equality and equity. In the first stage, justice is not differentiated from authority. Then, egalitarianism develops (by the age of 7 or 8), conflicting with obedience. Last, from 11 or 12, children qualify equality and give precedence to equity[15] by taking each particular situation into account. For instance, oldest children agree to favor the youngest when they play together, in order to compensate for the differences in abilities.[16] These three stages in moral development are not clear-cut, though, and the two moral worlds of constraint and cooperation coexist in childhood and persist in adulthood due to the weight of previous generations. In particular, Piaget observes that the different types of morality coexist at a given age, but that the proportion of children who refer to a given type differs across ages. Nevertheless, there is a gradual shift from one to the other, and attitudes of parents and their relationships with their children may favor or delay the development of cooperation and of reciprocal morality.

To summarize, even though adult influence is obviously huge, authority cannot be the source of justice because the development of justice assumes autonomy. Justice can only develop as both cooperation and mutual respect increase, first among children then between children and adults. For Piaget, distributive justice is equivalent to the notions of equality and equity. The notion of distributive justice certainly has individual or biological roots, but according to Piaget, from an epistemological point of view, such concepts can only be *a priori*, in the sense that they are norms "towards which the reason cannot not tend, as it refines" (p. 253). Equity and reciprocity norms are thus an ideal equilibrium. There must exist a collective rule, *sui generis* product of the common life: "The consciousness of a necessary equilibrium that compels and limits both the *alter* and the *ego* must

[15] Some parallels with the psychoanalytic approach can be drawn. According to Freud's theory (see, for example, Freud, 1968 [1923]), a newborn is driven by the demands of the id, which require immediate gratification. The dictatorship of the id partly corresponds to the Piagetian egoistic stage. Next forms the ego, which appears with the understanding that immediate gratification is generally impossible, and which acts as a form of repression or a control of the urge. The fear of punishment, in particular by the parents, is a mean used by the ego to repress the urge. Ego can be related to moral realism. By the age of 4 to 6, appears the superego, when the child begins to internalize these external sources of punishment. The superego uses guilt to enforce these internalized rules that may become values. Pro-social behavior may be the consequence of the superego's action. It is worth noting that, for Piaget, the rules are not internalized, but are continuously constructed by the group.

[16] When analyzing distributive justice and equity, Piaget always refers to equity according to needs. According to twelve or thirteen-year old children, equality should be tempered, but only to correct initial inequalities. In particular, they never mention that differences in abilities should be rewarded. It should be noticed that equity is not absent in the world of retributive justice, but it differs from the notion of equity associated to distributive justice. In the domain of retributive justice, equity consists in taking extenuating circumstances into account.

arise out of the actions and reactions of the individuals on others." (p. 254). As stated by Piaget and emphasized by Carpendale (2000), the central element for the development of morality is not the weight of society, as in a Durkheimian perspective, but the existence of social interactions. Hence morality is a social concept, but only insofar justice and reciprocity always concern at least two interacting subjects. For Piaget, Durkheim's error is that "there is no more society as a being than there are isolated individuals. There are only relationships […]" (p. 290).

For an economist, the Piagetian theory of pro-social development has interesting consequences. First, pro-social attitudes are not innate, and thus do not seem to be part of individual preferences, at least if preferences are viewed as given. Second, they are not shaped by parents. Third, they ultimately develop while children interact with peers, and follow from mutual respect. Parental authority can only enhance the conditions of the development of pro-social behaviors. Last, it requires that people are able to understand each other, so that cognitive development is a prerequisite to it.

3. Other theories of stages

After Piaget, numerous developmental psychologists have proposed theories of stages.[17] The most famous follower of Piaget is Kohlberg (1984; Colby and Kohlberg, 1987; see Campbell and Christopher, 1996 for a presentation), who, as noticed by Carpendale (2000), "followed Piaget in rejecting an explanation of moral development as a simple transmission of moral rules from parents to children as incomplete because this view cannot account for how such moral norms arise in first place, and it simply equates morality with conformity to moral rules." Kohlberg distinguished six stages of moral development extending and refining Piagetian stages. At the preconventional level (Stage 1 and Stage 2), subjects do not explicitly understand moral rules and social conventions. At Stage 1, actors base moral judgments on the material consequences of actions for them, and they consider that they behave good when they submit to authority and when they have avoided punishment; at Stage 2, moral judgments are based on what instrumentally satisfies the subject's needs. The conventional level (Stages 3 and 4) is the level of conformity, in which people only strive to conform to the rules of the group, first in order to please others (Stage 3), then in order to maintain the social order (Stage 4). At the post-conventional level, morality is distinguished from

[17] See, for example Krebs and Van Hesteren (1994) who propose a table comparing the stages elaborated by several developmental theorists (Table 1, pp. 114-115). They show that in spite of differences, there is a important degree of correspondence between the stages proposed by these theorists.

social convention. The Stage 5 has clearly a social-contract orientation. Moral rules are necessary to the good functioning of the society, and result from general agreement. At this stage, morality is subjective, and relative. The Stage 6 is characterized by universal ethical principles (justice, equal rights and respect for individual dignity), that concern everyone and thus are not revisable. These principles are those which any perfectly rational agent would choose.

In Kohlberg's views, individuals are supposed to progress orderly through these stages (although Stage 6 is very rare[18]) and this sequence cannot be changed by cultural factors (only speeded up, slowed down or stopped). The six Kohlbergian stages constitute qualitatively different modes of thinking, form structured wholes (a Piagetian concept), develop in an invariant sequence, and integrate previous stage structures in a hierarchical manner ("hard" stages in the words of Krebs and Van Hesteren, 1994). It follows from these properties that all the moral behaviors of a subject have to be consistent. However, as emphasized by Carpendale, there is much evidence of inconsistency. People do not always use the stage of moral development which they are supposed to have reached (Denton and Krebs, 1990), although the degree of consistency appears to increase with the age of children (see references in Eisenberg and Fabes, 1998). Children may develop differently[19] in each of the domains (friendship, justice and fairness, obedience and authority, social rules and conventions) and use different concepts to judge or to act (Damon, 1977). In particular, according to Damon, differential social knowledge in the different domains may enter in conflict.[20]

Developmental psychologists do not unanimously accept the existence of stages, especially that of "hard" stages. The evidence points to a softer definition of stages (see Eisenberg and Fabes, 1998). According to "softer" models (for example Damon, 1977; Eisenberg, 1986 or Krebs and Van Hesteren; 1994), development is characterized by the acquisition of increasingly complex forms of

---

18 This stage is thus generally left out in empirical studies.

19 As recognized by Damon (1977), it raises the problem of the comparison between the advancement in the different domains.

20 As Damon (1977), Turiel (1983) has argued that moral and pro-social thinking may apply differently in different domains. Turiel (1983; see Hoffman, 1988) distinguishes between moral and conventional thinking, which are viewed as distinct domains. Moral rules aim to regulate behavior which affects others' rights or well-being whereas conventional rules are used to promote behavioral uniformities that coordinate interactions within a social group. It follows that conventions are context-dependent (Turiel, 1998). Turiel (1998) reviews studies conducted by several psychologists (in particular, Turiel and Nucci) that support the hypothesis that moral issues are judged by children and adolescents as obligatory, not contingent on authority dictates, rules, consensus or accepted practices within a group. Moreover, judgments about moral issues appear structured by concepts, welfare and rights. On the contrary, certain social judgments are justified based on understandings of social organization, are linked to existing social arrangements, and are contingent on rules. This distinction clearly indicates, according to Turiel, that social judgments are not simply based on acceptance of societal values. Domain theory however lacks of developmental components (Glassman and Zan, 1995).

thought but stages are defined in terms of the content of thought, affective orientations and behavioral styles. In particular, old stage structures may be retained and invoked (Levine, 1979) after new ones are acquired, so that less consistency in pro-social behavior is expected. How children progress from one stage to another has nevertheless not received much consideration (see, however, Walker et al. 2001). Empirically, age differences in pro-social behavior appear to be complex and sometimes inconsistent. However, according to a meta-analysis involving 155 studies conducted by Eisenberg and Fabes and presented in their 1998 survey, age is positively related to the likelihood that pro-social behavior occurs. For example, Harbaugh et al. (2002) show that bargaining behavior in ultimatum and dictator games clearly changes with age among children between 7 and 18: the older the children, the more generous their proposals. They also found that very young proposers (second graders) earn more in ultimatum games than other age groups; but they also make the smallest offers in dictator games.[21]

As argued above, children do not develop autonomous morality as soon as they are exposed to other children. They must first develop cognitive and emotional skills that allow them to understand the needs and the position of others.

4. Cognitive correlates of pro-social development

Like Piaget, several theorists have hypothesized that cognitive skills like perspective-taking and moral reasoning foster pro-social behavior (Batson, 1991; Eisenberg, 1986; Hoffman, 1982; Staub, 1979). Among cognitive skills, some are personal (intelligence), whereas the others concern the relations between individuals.

Probably the most important cognitive skills for the development of pro-social behaviour are perspective-taking skills. The latter are often related with identifying, understanding, and sympathizing with others' distress or need skills, in particular with the capacity to differentiate between own and others' distress and thus to enhance empathy and sympathy. Individuals may acquire information about others' internal states by imagining themselves in another's position. They may also use other processes (see Karniol and Shomroni (1999) for a presentation), like developing a theory of others' psychology and using heuristics such as mental associations to "channel the

[21] This result is interpreted by the authors as an indication that the behavioral differences result from differences in preferences for fairness and not in abilities to play strategically.

memory search required for making predictions about other people's thought and feelings in any given context" (p. 148).

Three types of perspective-taking skills have been distinguished (see Underwood and Moore, 1982): perceptual (the ability to take another's perspective visually); affective (the ability to understand another's emotional state); and cognitive or conceptual (the ability to understand another's cognition). Most empirical studies have reported a positive relation between these three types of perspective-taking and pro-social behavior, although some have found no significant relation. Only a few studies have found a negative relation (for reviews of empirical work on this topic, see Underwood and Moore, 1982; Eisenberg and Fabes, 1998; Eisenberg et al., 2001). Besides, despite Piaget's assertion that children do not acquire the role-taking skills necessary to behave pro-socially because of insufficient cognitive abilities, some studies have shown that even very young children are able of taking roles (Hoffman, 1975a, for example).

Higgins (1981) argues that the ability to take an other's point of view into consideration when making judgments and decisions becomes more sophisticated with age, in particular as judgments become more abstract (Miller et al., 1970). Role-taking is a process by which one determines certain attributes of others, but it also involves, according to Higgins, "going beyond the information given" (p. 120), and is thus inference rather than just categorization. Last, as noticed by Eisenberg and Fabes (1998), the use of perspective-taking skills may depend on the context. Moreover, the effect of those skills may be moderated by lack of either relevant social skills or emotional motivation.

Other cognitive skills have sometimes been related to pro-social attitudes, although generally not consistently. Such is the case of intelligence (Bar-Tal et al., 1985), the level of expressed motives (see however Eisenberg, 1986), sociability or extrovert tendencies (Eisenberg et al., 1996), social competencies that are nevertheless often correlated to sympathy (Eisenberg and Fabes, 1995) and empathy (Adams, 1983, Eisenberg and Miller, 1987), popularity (Hampson, 1984), or self-esteem. However, Eisenberg and Fabes (1998:736)) remark that "[g]iven the correlational nature of associations between personality variables and pro-social behavior, causal relations are difficult to prove."

Last, there has been a great debate about gender differences. In particular, Gilligan (1982) distinguished justice (not to treat others unfairly) and care (not to turn away from persons in need), and has argued that, because most of the theory of morality has been formulated by males, the morality of care, supposed to be mainly present among females, has not received enough attention

(see Turiel (1998) for a discussion). Empirical findings are however ambiguous and gender differences in pro-social behavior appear to differ greatly with the situation (see Eagly and Crowley (1986) for a meta-analysis involving older adolescents and adults; and once again the meta-analysis by Fabes and Eisenberg summarized in Eisenberg and Fabes (1998) for children; see also Eisenberg et al. (2001) for recent results and references). In particular, differences are greater with self-reported and other-reported than with observational measures, in in-the-fields than experimental studies (this effect disappears when controlling for other characteristics of the study)[22], and when the target was an adult or unspecified than when it was another child.

V. Social cognition

1. Perspective-taking, identification with, and projection of self onto others

In this section, we propose a simple theoretical apparatus for describing how people make inferences about others and construct their own social preferences in interpersonal contexts. Social cognition is made possible by the development of the *perspective-taking ability* during childhood, reviewed in section 4.4. The formation of the perspective–taking ability is probably distinctive of human sociality because it requires sophisticated cognitive abilities and an extended period of development (childhood). Ants and bees, which have a detailed division of labor, have a social life; but they don't have a social mind.

In an attempt to simplify the processes described in a vast social psychological literature for the purpose of economic modeling, we retain essentially three mechanisms that will serve as building blocks for all subsequent analysis: *identification of self with an other, projection of self onto an other, and categorization of others as either similar or dissimilar to self*. These mechanisms rely on the development of the perspective–taking ability. Further elaboration of the third mechanism, i.e. categorization of others into similar and dissimilar others, can be found in section 7.1.

We first give a short definition of the terms being introduced. We refer to « perspective-taking » as the ability to exchange roles with one another in mental life. Perspective-taking is the basic tool that an individual possesses for making social inferences about others and constructing his own social preferences. Identification and projection are two distinct ways, which we think are the most

[22] Harbaugh et al. (2002) found significant differences between boys and girls, but only in dictator game. However, those differences disappear after controlling for height.

common, of adopting an other's perspective in specific contexts. They are opposite and extreme processes for making social inferences, since identification can be said to make the maximal use and projection the minimal use of one's perspective-taking ability. Self "identifies" with an other by mentally reincarnating in Other, while Self "projects" herself onto an other by merely imagining what she herself would have done if she played the role of Other[23]. Whenever individuals differ in more than one characteristic, like skills and preferences for example, a combination of these two processes is conceivable along different dimensions.

In what follows, we describe a rational perspective-taking person as viewing all her potential roles or identities as states of the world and ascribing subjective probabilities to each. If she knew nothing about her own future role or identity, she would ascribe equal probabilities to all states. We retain here the latter assumption, not only for convenience, but because it is actually justified by a variety of reasons like the impartiality of judgments, the similarity of group members, and the anonymity of relations in large markets as in experimental conditions. Social judgments of this kind will be reviewed by comparing the division of a given cake of size c between n members of a group both in the "identification" and in the "self-projection" treatment.

In the identification treatment, individual k must behave like an impartial judge *à la* Harsanyi (1955) who knows the initial distribution of wealth $(w_1,...,w_n)$ and all the individual preferences. If the judge is asked to share a cake between all members of the group, he determines his preferred allocation $(x_1,...,x_n)$ so as to maximize his expected utility

$$W_k = \frac{1}{n}\sum_{i=1}^{n} U_{ki}(w_i + x_i) \qquad (1)$$

$$\text{s.t. } 0 \le x_1,...,x_n \le c \text{ and } \sum_{i=1}^{n} x_i = c \qquad (2)$$

with: $U'_{ki} > 0, U''_{ki} < 0$ for all i. U_{ki} is the Von Neumann-Morgenstern utility function of k defined over the indirect utility of final wealth that k attributes to i.

In the self-projection treatment, individual k must behave as an impartial judge too, but one who, by lack of knowledge of others, evaluates the situation of others by his own standards. When asked to

[23] In our terminology, the self projects her known preferences onto others to determine her own behavior in social contexts. Another meaning of the word in psychology refers to the self projecting her known behavior to guess others' behavior. The problem with the latter concept is that it cannot be used to determine the behavior of self in social contexts.

share a cake between all members of a group, this person fails to perceive that the initial wealth and preferences of others will generally be different from his own and merely chooses the allocation that maximizes, subject to the constraints (2), his expected utility

$$V_k = \frac{1}{n}\sum_{i=1}^{n} U_k(w_k + x_i) \qquad , \qquad (3)$$

as if he were to play himself anyone of the potential roles.

We see the impartial judge's utility functions (1) and (3) as tractable ways of describing an individual's social preference[24] arising out of either his or her *identification* with, or *self-projection* onto, a group. The constrained maximization of (1) reduces to the Maximin criterion (advocated by Rawls (1971) in a non-utilitarian framework) when the judge (who runs the risk of reincarnating in the worst identity) has infinite risk aversion. Intermediate forms lying between (1) and (3) may be relevant for describing the social preferences of an impartial judge who can partially identify with others, say because he merely knows the wealth of others (for an application, see section 6.4).

These two mechanisms of social inference have two desirable properties. First, they respect Pareto-dominance insofar all participants share the same definition of goods. This is consistent with the fact that a judge's role is ideally to neutralize the inefficiencies caused by the strategic behavior of parties and reach the allocation which could have been reached by themselves had they accepted to use all the possible means of cooperation and exchange (Habermas (1979) comes close to this definition). Second, they can both generate judgments that do not depend on whether one belongs, or does not belong, to the group. For example, an impartial judge will share a cake alike whether he will eventually eat one of the shares or not.

2. Identification with a known other and self-projection onto unknown others

Identification and projection have different informational requirements. Identification requires good knowledge of Other while self-projection requires good knowledge of Self. As most people know themselves better than others, self-projection will be more common than identification in social interactions. However, identification should be more frequent with natural groups than experimental groups in which anonymous relations prevail. This is corroborated by a study of Jetten, Spears, and

[24] Although social choice theory seeks to derive society's preference from individual preferences, the derivation of society's preference is beyond the scope of the present paper. Kolm (2001) offers a recent detailed discussion of how individual preferences over distributions may converge towards a unique society-preferred distribution.

Manstead (1996) which compares the same measure of identification for experimental groups (students from the University of Amsterdam playing anonymously) and natural groups (students from the University of Amsterdam who believed that they were playing anonymously with students from its rival university in Amsterdam, the Free University). A test on the group identification data showed that the degree of identification was significantly higher in the context of natural social groups. Even though the scope for identification seems to be limited in social interactions, it is of the utmost importance in specific circumstances. A young child may identify with a parent along several observable dimensions if he has more limited knowledge of himself than of his parent. A parent may know her child well enough to identify with him even beyond the natural boundaries of her own lifetime (see Becker and Barro 1988). The ease of identification of self with an other under complete information reflects the economic definition of pure altruism.

In many situations, though, the informational requirements of identification are too stringent to be met. This point was vividly raised by Adam Smith (1982: 9 and 19) who derived one's "sympathy" for others from the faculty of projecting oneself onto the situation of another:

"As we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation. […] Every faculty in one man is the measure by which he judges of the like faculty in another. I judge of your sight by my sight, of your ear by my ear, of your reason by my reason, of your resentment by my resentment, of your love by my love. I neither have, nor can have, any other way of judging about them."

A closely related question raised by the theory of social cognition concerns the mechanisms that are used by self to predict the behavior of others when these are not well known. The same question may be raised about Rabin's (1993) model in which players react to the "kindness" of the intentions of the other player toward themselves. On the basis of what information can the self guess an other's intentions? In the experimental conditions of anonymous relations, numerous psychological studies have shown that individuals rely on self-information and are biased in viewing their own position as normative.

A preponderance of the research on the role of self-knowledge in social prediction has investigated the "false consensus effect" (for reviews, see Mullen et al. 1985, Marks and Miller 1987), first revealed by the studies of Ross, Greene and House (1977). The false consensus effect refers to the tendency of people to overestimate consensus for their own position, whether the estimated variable concerns an attitude, trait, behavior, or performance. People wrongly anticipate that other people think or behave like themselves in the same role. In a careful study, Alicke and Largo (1995) manipulated the own position variable (thus removing its endogeneity) and were able to assess the direction of the causality unambiguously. Many examples of egocentric biases have been found. For

instance, Frankenberger (2000) argues that, when adolescents and young adults start recognizing that other people think thoughts of their own, they anticipate that those thoughts will center on them, which results in adolescent egocentrism. Lind, Kray and Thompson (1998) noted that people place greater weight on their own experiences of injustice than on the injustice of others when formulating fairness judgments. Van Boven, Dunning, and Loewenstein (2000) observed that own perceptions of the endowment effect (i.e., the propensity to over-value an object that one owns and under-value an object that one doesn't own) contaminate estimates of others' perceptions, even when individuals know or suspect that others' perceptions are systematically different from their own. However, the egocentric biases which affect the prediction of similar others' behavior tend to level-off on average. Charness and Grosskopf (2001, table 1), asking subjects playing one role to indicate what they would have chosen in another's role, obtained a remarkable similarity of the average hypothetical choices with what another group of subjects actually chose in this role. The same was true when participants in one role were simply asked to estimate the other players' choices. Offerman (2002: 1433) made exactly the same observation[25]. Self-projection onto others thus appears to be a statistically unbiased mechanism for predicting the independent behavior of others when the Self lacks knowledge of others.

Alicke and Largo (1996) also demonstrated that people do not ignore case information about another person's position either but do not uniformly over-generalize from the latter information. The score of another individual is particularly valuable when a person has little or no information upon which to base a judgment. People who have similar preferences and values can serve as surrogates when judgments are required about objects or events that we have not experienced. People may also rely more on the opinion of others in estimating consensus when they know that their own opinion is idiosyncratic, or when other people clearly have more expertise in a judgment domain. On the whole, the results which have been obtained on the false consensus effect convey the impression that people use available information consistently while often treating self-information as being more precise than the same information supplied by an other. People rely on the position of others when they think that these detain relatively valuable information. Otherwise, as will often be the case in experimental conditions, people make inferences about others' positions by projecting their own.

[25] The average prediction of second players' behavior by first movers was no longer accurate when first movers predicted how the second players would react to their own intentional move (Offerman 2002: 1433-1434). Then first movers underestimated the probability of a reciprocal response, perhaps because they failed to predict the emotional component of the latter (discussed in sections 8.2 and 8.3).

The unrestricted use of self-projection would lead to large egocentric biases and costly errors. The purpose of *categorization* is to restrict the use of self-projection as far as possible to those cases for which it yields the more precise inferences. Whenever people lack knowledge of others, they first categorize others with respect to their similarity with self, then they rely on self-information to anticipate similar others' behavior (Dunning and Hayes 1996, Cadinu and Rothbart 1996, Gramzow et al. 2001). But they feel unable to project onto the group of dissimilar others. The most natural assumption in the latter case is that they will refrain from making inferences for this category and, as far as possible, take the behavior of dissimilar others as given. Categorization and the treatment of dissimilar others will be further examined in section 7. In section 6, we focus on self-projection onto similar others.

According to the social cognition story, the concern for others may grow out of three factors: the perspective-taking ability, knowledge of self and/or others, and the perceived similarity of others with self. In this view, an "egoist" is an individual who either lacks a perspective-taking ability or who systematically perceives others as being dissimilar to self. Since the ability to take others' perspective normally develops during childhood (as discussed in sections 4.4 and 5.3), a young child would be a natural egoist who would not share goods with his playing partners (see Harbaugh, Krause and Linday 2002). Adults who were led to believe, through their own experience, education or culture, that most others are not like themselves would form a very different kind of egoists, and one which fits nicely with the current conception of an egoist. The extensive evidence that people have heterogeneous social preferences, some being egoists and many others fair-minded or altruists, is thus wholly consistent with the principles of social cognition mentioned here.

3. Social cognition and the stages of pro-social development

In the end of this section, we give two illustrations of the general applicability of the foregoing analysis of social cognition for the development and construction of social preferences. As a first illustration, we use these processes to recover, quite simply, the stages of pro-social development observed by Piaget and Kohlberg (described in sections 4.1 to 4.3). For making the analysis more concrete, we examine here the four stages of empathic distress considered by Hoffman (e.g., 1981) in the face of an other in need. We show that these four stages may simply result from the combined development of perspective-taking skills and empathic concern. Empathy is an affect (see section 8.3), but it has a cognitive component since one has to use knowledge about others and thus be conscious of the self-other distinction.

Table 3 The development of perspective-taking skills, empathic concern and the four stages of empathic distress

Information of Self about Other:	*Perspective-taking skills:* *No*	*Yes*
Low Consciousness of Other	Global empathy (babies) ↓	
Consciousness, but little knowledge, of Other	Egocentric empathy (2-3 year old) →	Sympathy ↓
Deep knowledge of Other		Empathy-altruism

Table 3 summarizes the argument along two dimensions, with the three rows describing the development of consciousness of an other by the self and the two columns representing the development of perspective-taking skills. These two dimensions are not independent as pointed out by the two incompatibilities listed in table 3, which leaves us with four possibilities. The arrow underlines four successive stages of development for empathic distress. Research has shown that even very young children respond to the cry of distress of other babies (Simner 1971). During the first year of life, children cannot distinguish the distress of another person from the unpleasant feeling aroused in the self. They experience global empathic distress response ("global empathy"). Then and until 2 or 3 years of age, they are able to differentiate self from others but they still lack the perspective-taking skills to discern the internal states of others ("egocentric empathy"). As they develop perspective-taking abilities, children become increasingly aware that others' feelings may differ from their own. Inevitably, they must realize that they have better knowledge of themselves than they have of others. Thus, in the third stage, they are unable to identify with others and must use the self-projection mechanism to make sense of others' feelings and behavior. At this stage of their development, they sympathize but do not yet empathize with others. As they acquire enough knowledge of others through the experiences of childhood and adolescence, they will eventually be

able to identify with some others[26] and empathize with them. The development of empathy is likely to be gradual and extend to an increasingly wider range of emotions and people.

4. Choosing and valuing an income distribution

A second illustration of the self-projection mechanism is offered by the way people choose and value an income distribution and thus resolve the equity-efficiency issue. This problem has been studied by the psychological literature on behavioral justice. Several orientations have been proposed. One is that people acquire stable preferences for equity and efficiency, very much like a conventional economist would state the problem. Rohrbaugh, McClelland, and Quinn (1980) have defended this approach and thus sought to measure how much negotiators on a labor-management contract valued both total utility and equity. Total utility was measured by the sum of individual utilities accruing to both parties, while the inequity of a specific contract was defined by the absolute difference between the two individual utilities. In general, utility was shown to be over twice as important as equity in participants' determination of acceptable contract settlements. Yet, equity was treated as such an important value to the participants that Pareto optimality was routinely violated. Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) have shown that the mere addition to own utility of a concern for equity was sufficient to predict much (though not all) cooperative and reciprocal behavior within an otherwise conventional game-theoretic framework.

Experimental research on behavioral justice leads to the extended assumption that people compromise between these two goals in a *context-dependent* fashion. Mellers (1982) found that subjects used a context-dependent merit rule to set "fair" salaries[27], but salary levels were constrained by a

[26] Stinson and Ickes (1992) have shown experimentally that male friends were more able than male strangers to accurately read their partner's thoughts and feelings about imagined events in another place or time.

[27] In Mellers (1986), people choose "fair" allocations of salaries and taxes among hypothetical faculty members on the basis of their merit ratings. They do not follow the rule of proportionality of salaries to merits or contributions (equity theory of Adams 1965, relative ratio model of Anderson 1976). A better fit is obtained by asserting that subjects assign salaries in such a way that the relative standing of a person's salary in the distribution of salaries matches the relative position of his or her merit in the distribution of merits (relative equity theory). The relative position of a person's merit in the distribution of merits is assumed to be given by Parducci's (1965) range-frequency compromise. It is a weighted average of the person's subjective value of merit (a cardinal measure taking values of zero for the minimal value and one for the maximal value) and the person's rank in the distribution of merits. Therefore, the relative position of a person's merit is steeply increasing in merit for values where the distribution of merits is highly concentrated. The adjusted function for the subjective value of merit was a cubic function of merit. It was found that the weight of the rank, which measures the sensitivity to the form of the frequency distribution of contextual values of merit, is fairly constant across distributions of merits and budgets. For the salary allocations, the estimated value of the weight is 0.44. In the salary

floor paralleling the stated poverty line. Frolich, Oppenheimer and Eavey (1987) found that experimental groups allocated rewards according to a rule that maximizes average income after allowing that no group member fall below a certain income level. To compare these trade-off approaches, Mitchell et al. (1993) had subjects judge the relative fairness of income distribution in hypothetical societies with varying efficiency and equality. Using a hypothetical society paradigm[28], they manipulated the mean income (representing efficiency) and income variability (representing equality) of distributions of income and the correlation between income and effort within a society. Subjects made all pairwise comparisons of distributions within societies of differing meritocracy. Rawls's (1971) maximin principle of justice received considerable support whenever subjects believed effort and reward were only loosely related. People maximized minimum income within a society. However, a compromise principle best described preferences when income was tightly linked to effort. People then rejected distributions in which some citizens fell below the "poverty line" but maximized efficiency above this constraint.

Differences in the results obtained by various studies are related to the procedures used for eliciting preferences over distributions of income. When people must decide on the highest-valued allocation of income given the contextual distribution of merits (as in Mellers 1982, 1986), they focus on efficiency and merely soften their evaluation by consideration of the minimum income. By contrast, when they must compare given income distributions which unambiguously differ on efficiency and equity grounds and the correlation between income and merit is not salient, many subjects focus on equity and give more emphasis to the lower end of the distribution. As a matter of fact, preference reversals on income distributions parallel those that have been extensively observed on lotteries (e.g. Lichtenstein and Slovic 1971). People often choose the lottery yielding a small gain with high probability but set a higher selling price for another lottery of similar expected value which yields a smaller probability of a higher gain. Similarly, in making judgments of distributive justice, people often give a high value to efficiency while they prefer the less efficient but more equal and just society when asked to compare between hypothetical societies. The reference-dependence of inequality aversion has been nicely shown in a recent paper of Dolan and Robinson (2001).

Therefore preferences on income distribution revealed by opinions of distributive justice support the view that people represent themselves (or a hypothetical self) as being randomly assigned a position

---

allocation tasks, the minimum living allowance for the lowest merit person ranges from 54% to 69% of the average salary in each condition.

[28] Participants are told that "econometric studies" can accurately determine which effect various policies, emphasizing either efficiency –i.e. the overall standard of living- or equality-i.e. the difference in average income between classes-, would have on the income distribution.

in the distribution of income[29]. They take all the different perspectives and project themselves onto each position. In judgments involving lotteries, the lower ranks are often over-weighted in the comparison framing, and the higher ranks are over-weighted in the valuation framing. In judgments about societies alike, the lower ranks –and therefore equity- are over-weighted in hypothetical choices, and the higher ranks –and therefore efficiency- are over-weighted in valuations. Risk-averse people do not wish to imagine themselves at risk of being in the lower ranks of a society who treats the poor badly, but they demand a high price in order to forego the upside risk of being in the upper ranks of a society who treats the deserving nicely.

VI. Social norms and reciprocity

1. The fairness heuristic

Fairness judgments are essentially needed when one moves into a relationship with other people or with an organization. Fair treatment leads to a shift from responding to social dilemmas in terms of immediate self-interest, which might be termed the "individual mode", to responding cooperatively, which might be termed the "group mode" (Lind 2001).

Fairness heuristic theory emphasizes the cognitive function of fairness. Fairness gives people prior information as to the extent to which they can trust others not to exploit or exclude them from important relationships and groups. People pay more attention to fairness when such information gets more valuable, that is when they are uncertain about things such as the outcome of others (Van den Bos, Lind, Vermunt and Wilke 1997) or an authority's trustworthiness (Van den Bos, Wilke and Lind 1998).

Fairness heuristic theory was initially concerned, not with the fairness of outcomes but with the fairness of procedures (Folger 1977). Lind and Tyler (1988) noted that information about procedures often affects people's fairness judgments more strongly than information about outcomes. For instance, in Tyler and Lind (1992), people want to have information about whether they can trust the authority. When this information is not available, people of bounded rationality will resolve the uncertainty by relying on impressions of fairness and will react more positively toward the outcomes of the authority's decisions if the authority is using fair as opposed to unfair procedures.

[29] This is not to say that people actually maximize a normative expected utility like (3), as such behavior would be inconsistent with preference reversals between bids and choices. Our inference is based on the striking similarity of behavioral anomalies concerning judgments over lotteries and over distributions of income.

To be functional as heuristic, judgments of justice should be used more than they are revised. Once people have established fairness judgments, perceived fairness will serve as a heuristic for interpreting subsequent events. Therefore, fairness heuristic theory suggests that fairness judgments are more strongly influenced by information that is available in an earlier stage of interaction with the authority than by information that becomes available at a later moment in time. Second, in many situations, information about the procedure is available before information about the outcome. For example, the manner in which a court trial is conducted is usually known before the verdict becomes apparent. Thus people form their fairness judgments on the basis of the fairness of the procedure and the perceived procedural fairness positively affects how people later react to their outcome. This "fair process effect" (Folger et al. 1979) is one of the most replicated findings in social psychology. The fair process effect has been found consistently both in experiments (e.g. Folger et al. 1979, Lind et al. 1990, Van den Bos, Vermunt and Wilke 1997) and in survey studies. The Lind et al. (1990) experiment, for instance, manipulated whether participants were or were not allowed an opportunity to voice their opinion about the number of tasks they were assigned. A fair process effect was found. Those who were allowed to voice their opinions not only judged the procedure as more fair, but also judged their outcome (the tasks assigned to them) as more fair than participants who were not allowed to voice their opinions.

Fairness heuristic theory views fairness judgments as being formed under uncertainty in an *early* stage of the cognitive process and strongly conditioning behavior. This general argument is not restricted to procedural fairness. Van den Bos, Vermunt and Wilke (1997) tested the prediction that early information sets the stage for the interpretation of the later fairness information. By making outcome information available before or after process information, they found indeed a primacy effect: the first information, whether procedural or distributive, affected people's fairness judgments more strongly than the later one. Lind, Kray, and Thompson (2001) conducted a further experiment to show that the primacy effect holds as well with a single type of fairness information. Participants working on a series of three tasks experienced delays caused by equipment failure and always had the possibility of explaining problems to a supervisor. The supervisor refused to consider explanations in one of the three work trials but did consider explanations on the other two trials, and the timing of voice denial was manipulated. Even though all of the participants received the same number of positive and negative fairness experiences, those who encountered the unfair experience early in their relationships with their supervisor viewed the supervisor as much more unfair as did those who encountered the unfair experience later. Roch et al. (2000) further demonstrate the cognitive

function of fairness by showing, in a resource-sharing task, that *thoughts* of anchoring on equality preceded thoughts regarding adjusting from this anchor. They also manipulate high cognitive load[30] and show that the two-stage reasoning only applies to individuals with sufficient cognitive resources. Those subjects with high cognitive load stopped once they applied the equality heuristic, presumably because they were prevented to perceive the self-serving arguments that they would have normally perceived in the second stage.

Even though the fairness heuristic strongly conditions later behavior in social dilemmas, all players don't play fair and systematic deviations are observed. These are important facts which require theoretical efforts in the future. Roch et al. (2000) propose a two-stage model in which individuals first anchor on the equality heuristic and then adjust their requests in a self-serving manner from the amount prescribed by the equality heuristic. Güth (1995) suggested a two-stage process for describing the reasoning of two players involved in an ultimatum game but did not elaborate a formal model. Lévy-Garboua and Rapoport (2002) propose a model of rational behavior under dynamic uncertainty which predicts the formation of fairness norms and allows for individual self-serving deviations from the norm in the second stage.

2. Social norms of fairness in proposal-response games

Psychological work on social dilemmas (Dawes 1980, Messick and Brewer 1983, Komorita and Parks 1995) often attributes to norms the tendency of people to cooperate (Kerr 1995). Social norms can be defined as *enforceable tacit coordination rules*. Social norms of fairness are effective in many contexts of interpersonal relations (e.g., Allison and Messick 1990, Allison, McQueen, and Schaerfl 1992, Samuelson and Allison 1994, Van Dijk and Wilke 1995). As the simplest and most pervasive instance of fairness heuristic is certainly the equality heuristic, we consider here for illustration the social norm of sharing a given cake equally among all members of a party sitting around the table. The purpose of this section is to suggest that the fairness heuristic may be interpreted as a social norm and, further, to relate the emergence of this social norm with the self-projection mechanism spelled out in section 5.

First of all, the fair division rule cannot arise from players' identification with the group, that is some form of altruism. The maximization of (1) under the constraints (2) usually entails very unequal

[30] A high cognitive load was operationalized by requiring participants to remember an eight-digit number while performing the task, a manipulation successfully used in previous studies investigating the impact of cognitive load.

sharing. Thus it is unable to explain why most of us will usually divide the cake in equal shares notwithstanding existing differences in wealth and preferences. This occurs even if all identities have been given equal weight, except under very special circumstances like all individuals' having identical preferences and initial wealth.

Let us imagine the following thought experiment which, we believe, offers a close description of how thoughts of anchoring on equality first come to mind in a resource-sharing task. Before dividing a cake, the people sitting around the table stand a few seconds in a symmetrical position of not knowing who will be asked to share the cake. At this moment, they play an *n*-person proposal-response game in which a single player will be given the role of a "proposer" and others will act as "responders". The proposer first offers shares to all players; then, responders react to her offer, say by accepting or rejecting it as in the ultimatum game. The projection mechanism allows one to anticipate (perhaps wrongly) that other proposers behave like self in the same role, as if they shared one's preferences and initial wealth. Each player will choose the whole distribution of shares that she may get depending on whether she will share the cake or not. If she has an equal probability of playing any role, she maximizes her expected utility (3) subject to the constraints (2). The solution of this simple program is to cut the cake in equal shares notwithstanding the player's risk aversion, initial wealth, and number of players or cake's size.

Two remarkable results come out of the projection mechanism. First, the preference for equality applies to wealth increments, not to final wealth. People may share a cake with their friends, workers may share rents with their co-workers, taxpayers may share their marginal income; but it is certainly uncommon to see a man share his fortune. Second, this solution is independent of the player's index. Thus there exists a *prior common preference* for equality of shares, which all rational players must be aware of before the game begins. A player's prior preference defines his or her *intention*. Thus intentions of other players are common knowledge. Knowing with certainty that all potential proposers intended to share the cake equally, a responder will be entitled to object to receiving a smaller share. This creates in turn the conditions for the *reciprocity* of responses to proposals being common knowledge as well and, therefore, enforceable. In the words of Kahneman, Knetsch and Thaler (1986), "the rules of fairness define the terms of an enforceable implicit contract". If prior intentions are effectively enforced in the actual game, proposers who do not cooperate will incur a sanction by being deprived of an excessive share or, more frequently so, by facing social disapproval for their unkind manners. This unique combination of a prior common reference of all players with

an expectation of sanction imposed on deviant behavior fits the definition of a *social norm* as a tacit coordination rule. The latter fills both a cognitive function, by eliciting others' intentions and making all players feel certain about them, and an incentive function, by driving individuals to respect their prior intentions. However, the incentive provided to one player by the knowledge that other players know their own intentions is weak. In a public good game, for example, social norms may become excessively vulnerable to free riding of a minority of players if the power of social norms exclusively relies on their being common knowledge to all players. The vulnerability of social norms can only be overcome by punishing free riders. In a repeated public good game, the early defection of some players signals to loyal players that the social norm can no longer be trusted and effective punishments are needed to maintain the flow of voluntary contributions in the long run (Fehr and Gächter 2000). Even the threat of social disapproval cannot deter defection in the long run if it bears no opportunity cost to free riders (Masclet et al. 2003).

Social norms and the fairness heuristic seem to provide a promising avenue for positing the role of intentions and reciprocity in proposal-response games and develop the insights so far provided by the recent economic models of reciprocity (Rabin 1993, Levine 1998, Dufwenberg and Kirschteiger 1999, Falk and Fischbascher 1999, Segal and Sobel 1999, Kolm 2000 and Charness and Rabin 2002).

3. Some evidence on social norms

The results obtained in the ultimatum game and public good game literature give credence to the social norm interpretation. In the ultimatum game, fair sharing is the rule with most proposers giving between 40 and 50% (almost never more) to the responder, almost no offer is found below 20%, and low offers are frequently rejected. These robust findings (Fehr and Schmidt (1999), for instance, derive these quantitative conclusions from ten studies) agree with the social norm of sharing but refute the narrow self-interest (subgame-perfect Nash equilibrium) interpretation. In the public good game, the fact that people's willingness to contribute to a public good depends on their perception that other people are also willing to give provides good evidence that common knowledge of others' intentions to respect the norm of fairness matters (e.g., Gächter and Fehr 1999). Finally, based on coordinated ultimatum experiments in 15 small-scale societies, Henrich et al. (2003) found strong support for the enforcement of norms within each group since group level differences explained two-thirds of total variation in ultimatum game offers.

Strong additional evidence of the social norm of equal sharing is indirectly provided by recent experiments of Charness and Rabin (2002). In one experiment, 85% of subjects A gave up a very advantageous but unequal allocation of 900 to self versus 450 to an other B to let B, who had the option of sharing equally (400 to each), make the decisive choice. Since B could also make another choice that would have preserved his share of 400 but would have been damaging to A by only giving her 200, the great majority of A's took the risk of losing 700 for the sake of respecting the social norm. Apparently, the latter exhibited great confidence that B's would in turn respect the social norm when given the opportunity-which about two-thirds did- despite the fact that A's move surely deprived B's of 50. In another experiment, after 61% of A's had let down an allocation (375, 1000) giving 1000 to B, 97% of B's preferred an equal split (400, 400) to the strongly Pareto-dominated allocation (250, 350). Since the latter would have been the way to punish A for "depriving" B of an opportunity to get a much bigger sum, this result means that almost no B's wanted to punish the "unkind" A's. This finding may come as a surprise under an egocentric interpretation of "kindness", but it corroborates the social norm interpretation, because many A's and B's together understand that the fairest allocation among the three alternatives they had is the equal split. Thus the two players manage to coordinate through the use of this social norm. It is not normal for an A subject to punish herself by accepting too small a share when there exist a fair option, and therefore there is no reason for the second player to punish someone who behaved normally (and not unfairly).

Equality of shares is not the only social norm of fairness that people adhere to. Van Dijke and Wilke (1995) demonstrated that, when players possess different endowments and are fully aware of these differences, Resource dilemmas evoke different norms than Public Good dilemmas. Whereas one-shot Resource dilemmas appear to evoke the equality rule, participants to a one-shot Public Good dilemma appear to coordinate behavior through a proportionality rule. According to the latter, each member of the group should contribute to the (fixed amount of) public good in proportion of his or her ability to pay. The use of a proportionality rule does not lead to an equality of final outcomes. Van Dijke et al. (1999) further showed that incomplete information of players as to others' endowments or investment returns had in some cases a profound impact on the norm of fairness. In each case, a single specific rule was followed fairly closely by a great majority of players. Moreover, the players' reported own notion of fair choice was highly correlated with their actual decisions. These results suggest that players use the available information to determine a game-specific norm of

fairness and what would be fair behavior for each group member, and that they anchor their own actual behavior on what would be fair for them.

Since it is the presence of (often implicit) social sanctions which ensures the effectiveness of norms, a surprising implication of the functioning of norms is that financial inducements to perform a socially desirable task may undermine the prior willingness to perform this task by signaling to players that the norm is no longer in use. This is one facet of what is called the "crowding-out of an intrinsic motivation" by an extrinsic reward. This phenomenon (Deci 1971, and Frey 1997 for an economic exposition) will occur when the financial inducement is less effective than the norm in driving individuals to perform the socially desirable task. Unexpected interactions between material incentives and non-pecuniary motives are extensively discussed by Fehr and Falk (2002).

4. The working of a social norm: Homans' "cash posters"

We end this section on social norms with a formal illustration of their working. We chose to examine the celebrated case of the "cash posters", studied by Homans (1953, 1954), for its extreme clarity and simplicity. This case will be used to demonstrate how the projection/identification mechanism lying at the heart of social judgments may condition the revealed preference for equal hourly wages and a norm of minimal effort within this specific group of workers. Showing the interplay of social preferences and norms of sharing will further help us to reconcile the theoretical intuitions of Homans and Adams (1963, 1965) and to fit them rather naturally into a unified economic framework when the conditions of perfect competition do not prevail [31].

Cash posting consisted of recording daily the amounts customers of a utilities company paid on their bills. A group of 10 young women who worked in the same large room were interviewed and observed over a period of six months. The speed at which individual cash posters worked was recorded. Anyone who worked below the rate of 300 per hour received a mild rebuke from the supervisor. The average number of cash postings per hour was 353, well above the company's minimum standard. Only two workers had productivities slightly above the company's norm (namely, 306 and 308), and only two reached more than 400 per hour. In spite of observable differences in their human capital $(h_1,...,h_n)$, all the cash posters (e.g. $n=10$) received the same hourly wage $\overline{w}$.

[31] The reader may also refer to the discussion of Fehr and Falk (2002).

The firm acts as principal toward its employees. It proposes two enforceable rules for the division of output value, first by setting the labor's share p ($0 < p \leq 1$) relative to the firm's profits, second by choosing a pay scheme which allocates wages to workers of varying abilities after the wage bill has been set. Let the allocation of wages result from either of two pay schemes: team compensation, or piece-rate. Assuming that the definition of the equitable share of labor p does not critically depend upon the payment scheme, we normalize p to one. The essential difference between compensation of effort being based on either collective or individual output lies in the revelation of social preferences. All workers have a say in the distribution of wages when output is provided collectively, but each worker is denied voice and property rights on others' outputs when these are singled out. For simplicity, the group of workers is further assumed to determine the group's preferred pay scheme by majority voting.

The social preferences of workers who see their co-workers as forming a homogeneous group with themselves will be captured by assuming that each worker i "socially" choose the distribution of wages $(w_{i1},...,w_{in})$ and the distribution of efforts $(e_{i1},...,e_{in})$ for their own team[32]. Social preferences can be viewed here as a special kind of peer pressure (Kandel and Lazear 1992). Since each worker knows the human capital of all members of the group (who accomplish the same task in the same room) but doesn't observe the preferences of others for income and effort, it is natural to assume that worker i makes her personal judgment about all the wages and efforts by both projecting her own preferences onto her co-workers and identifying with their human capital. Worker i's behavior can thus be described by the maximization of a social utility function[33] which is a combination of (1) and (3), subject to her perception that the sum of wages must equal the labor's share of total output:

$$\max_{(w_{ij},e_{ij})} \sum_{j=1}^{n} \frac{1}{n}\left[u_i(w_{ij}) - c_i(e_{ij})\right] \qquad (4)$$

$$\text{s.t. } f(h_1 e_{i1},...,h_n e_{in}) = \sum_{j=1}^{n} w_{ij} \qquad (5)$$

[32] For a purely egoist worker, the perceived size of "team", i.e. group of similar workers, is simply: $n = 1$.
[33] Effort in the workplace is assumed to be separable from home goods and leisure.

with: $u_i' > 0, u_i'' < 0$, $c_i' > 0, c_i'' > 0$, human capital and effort taking real positive values, and *f( )* designating the production function. In the cash posters' example, the latter is simply additive: $f(h_1e_1,...,h_ne_n) = h_1e_1 + ... + h_ne_n$. The first-order conditions are:

$$u_i'(w_{ij}) = \mu_i \qquad (6)$$

$$c_i'(e_{ij}) = \mu_i h_j f_j' \qquad (7)$$

The *2n* equations (6) and (7) and the wage-effort constraint (5) determine *i*'s social preference for the distributions of efforts and wages, and her positive Lagrange multiplier μ_i. Clearly, these conditions imply the unanimous preference of workers for equal wages and determine the uniform wage and total output preferred by this individual. This conclusion does not rest on the special form of production function which applies to cash posters. It is also worth noticing that the same prediction would derive from a maximin social utility which assumes infinite risk aversion. If most workers are not egoists, the firm knows that uniform wages are the team's norm and that these workers intend to reciprocate the firm's policy to respect this norm in such way that it may be profitable to set uniform wages at the level which is sufficient to attain the desired level of output.

This description of the formation of social judgments predicts the normative preference for equal wages in a not-too-heterogeneous team and associates the norm of minimal production with the expected minimum productivity. The cash posters observed by Homans conform to this description. They formed a roughly homogeneous group and the group's norm was very close to the minimum productivity attained by two of the employees. All workers actually intended to respect this norm and even agreed to produce at a faster rate depending upon their own human capital and preferences. The more productive are willing to help their less productive co-workers if they cannot access to another better group[34] because they still benefit from the cooperation of other productive workers in their own group. The norm of minimal production that we describe is obviously easy to enforce but still needs to be recalled because, once an hourly wage has been set for workers the latter have an incentive to make less effort than they implicitly promised. In actuality, the norm is not restricted to minimal production and extends to all workers since lower-than-expected efforts of anyone are costly to all co-workers with a lower ability than herself. This is made possible by the observability of individual abilities and outputs by co-workers who can thus reward the more productive with a

[34] The cash posters were quite young (21.1 years on average) and had low tenure on the job (with a maximum of 3 years and 5 months). Turnover costs may be an important factor in determining the overall efficiency of team compensation.

higher social status and prestige among the group, and punish any worker's negative deviation from their normative expectation by downgrading her status or excluding her from informal relations. Note that each worker's efforts are monitored by lower-ability workers who stand to lose from shirking on the part of their higher-ability co-workers. Only the lowest-ability workers need to be monitored and punished by the firm's supervisor.

Interestingly, team compensation need not be less productive than a piece-rate scheme (assuming that the latter is feasible) when the workers control their own effort. One basic reason is that the piece-rate scheme constrains wages to parallel productivities that will be partly chosen by each worker, whereas team compensation doesn't. For instance, a piece-rate scheme could have been implemented for the cash posters who were all accomplishing the same task independently at an observable rate. Under piece-rate (normalized to one), wages follow individual output: $w_i = h_i e_i$, for all *i*. With the piece-rate scheme, the wage-effort constraint is automatically verified and worker *i* merely maximizes her private utility function[35] given that her wage is tied to her own productivity. The corresponding first-order condition is

$$c_i'(e_i) = h_i u_i'(w_i) \qquad (8)$$

By contrast, under the same production frontier and team compensation, individual efforts would be determined by

$$c_i'(e_i) = h_i u_i'(\overline{w}).$$

More able workers who would be paid higher than average in the piece-rate scheme will end up making more effort under team compensation for lower pay, if the average wage is held constant. To see that this is definitely possible, let us assume identical utility functions of the form: $\ln w_i + \beta \ln(1 - e_i)$, where $\beta > 0$ and $0 < e_i < 1$. Then the solution is easily calculated. Under team compensation, $\overline{w} = \frac{\overline{h}}{1+\beta}$ and $e_i = 1 - \frac{\beta}{1+\beta}\frac{\overline{h}}{h_i}$; and in the piece-rate scheme, $e_i = \frac{1}{1+\beta}$ for all *i* and $w_i = \frac{h_i}{1+\beta}$. Hence, total output coincides in both regimes $\left(= \frac{n\overline{h}}{1+\beta}\right)$ and so does the average wage. Since more able workers $(h_i \geq \overline{h})$ produce less effort for more pay in the piece-rate scheme,

[35] The private utility function is a special case of (4) when $n = 1$ and $j = i$.

they will tend to favor this regime while the less able workers prefer team compensation. If the distribution of human capital is skewed to the right, as it usually is, team compensation would be chosen over piece-rate under majority voting.

VII. In-group favoritism and self-anchored altruism

1. Categorization and the preference for conformity

The sharing problem discussed in section 6 illustrated how the projection of self onto others can generate a social norm of equality, which is often taken for granted. This is a natural assumption whenever individuals believe that others are "similar" to themselves, for instance because they know their own preferences and have no information about others'. However, subjects tend to categorize others as soon as they receive distinctive information about the latter. Let us assume that others can be classified by an individual as belonging either to her *in-group*, made of similar others, or to her *out-group*, made of dissimilar others. More precisely, it is assumed that individuals can project onto similar others, but can neither project onto nor identify with dissimilar others of whom they have no detailed information. Thus they may reason that similar others behave like self and take the choices of dissimilar others as given.

With this minimal information, an individual and similar others are willing to keep x for themselves if they receive a sum c, give d to their out-group and share the rest equally between other members of their in-group if they play the role of a proposer. Under the same context, out-group members will keep y for themselves, give e to their out-group (*i.e.* the first individual's in-group), and share the rest equally between all members of their in-group (*i.e.* the first individual's out-group) if they become proposers. The representative player of the in-group perceives her in-group to be of size I (including herself) and her out-group to be of size O, with: $I + O = n$. She reasons that other players will share her own estimates (given the fact that her in-group may be their out-group). Thus in-group members determine x and d for given values of y and e in order to maximize

$$\max_{x,d}\left\{\frac{1}{n}U(w+x)+\frac{n-1}{n}\left[\frac{I-1}{n-1}U(w+\frac{c-x-d}{I-1})+\frac{O}{n-1}U(w+\frac{e}{I})\right]\right\}$$

$$st.\quad 0\le x\le c$$

$$0\le d\le c$$

Even though in-group members do not know which shares out-group members would keep for themselves or give to them, they do not need this information in order to determine their own

behavior thanks to the additive separability of the expected utility function. The solution consists in sharing the cake equally between all members of one's in-group and giving nothing to out-group members, irrespective of the individual's risk aversion and initial wealth

$$x = \frac{c}{I} \quad , \quad d = 0$$

Since the optimal sharing rule is independent of out-group's behavior and other characteristics of the proposer except which group she belongs to, anyone will eventually be able to infer her out-group's behavior by symmetry. An individual will favor her in-group in order to just compensate for her expectation of being discarded by her out-group. Her expected share of the cake is always c/n and does not depend upon how dissimilar others are effectively. However, out of risk aversion, she would still prefer to play with a group of similar people than with a heterogeneous group because, in the first case, she would be "certain" to receive the share that she only "expects" to receive in the second case. The modal preference for conformity is a well-known fact in social psychology, pioneered by a famous experiment of Schachter (1951). It is worth noticing that such preference for conformity does not require that people *a priori* evaluate themselves positively or prefer their in-group than their out-group. It is only the consequence of the ability to project oneself onto perceived similar others and the inability to do the same on perceived dissimilar others. We manifest a universal preference for sharing with family, friends, and other people we know rather than strangers; and we are inclined to like people who, we believe, are like us.

Consistent with the present analysis, Henrich (2000) found that the Machiguenga of the Peruvian Amazon proposed only 26% of the money in ultimatum games with very few rejections, well below the 40-50% range usually observed. Machiguenga people still primarily rely on their own family for their living but now live in small communities gathering a number of extended families and households. Therefore, Machiguenga proposers probably perceived high responders' heterogeneity because they faced a high probability of being matched with a member of another family. This also provides a potential explanation for the rise of selfishness caused by the expansion of markets, as noted by Adam Smith (1776), which it relates with the necessity to trade with unknown parties of different origins and customs. The emergence and persistence of selfish behavior may have resulted from a feeling of social heterogeneity that did not arise in smaller long-established segments of the society. It may also explain the emergence and persistence of prejudice and social discrimination on the basis of race, ethnicity, nationality, religion, or even gender. In societies visibly divided in two

large groups, most members of an in-group who can benefit from economic rents (the "favored" group) will be willing to share rents with similar others and refuse to do it with dissimilar others.

2. In-group favoritism and out-group discrimination in minimal groups

Favoritism toward similar others (in-group) and discrimination against dissimilar others (out-group) is a widespread phenomenon which can take the opposite forms of liking or attraction toward in-group members and disliking or aggression against out-group members. One interesting feature of this phenomenon for economists is that discrimination between the in-group and the out-group may arise even when the two groups do not compete for scarce resources. For example, Ferguson and Kelley (1964) showed that participants who had been working independently in two groups judged their own group product more favorably than the other group product, irrespective of any objective differences in output between the two groups. Even more surprising is the finding that in-group favoritism also occurs in a "minimal group" setting, first introduced by Rabbie and Horwitz (1969) and Tajfel, Billig, Bundy, and Flament (1971).

In the typical minimal group paradigm, subjects are assigned anonymously to one of two novel groups, and there is no direct interaction between or within groups during the experiment (Brewer 1979). Group membership is determined by an arbitrary or trivial criterion like being rated as a person preferring the art of Klee versus the art of Kandinski (Tajfel et al. 1971) or flipping a coin to decide which of the two groups would receive a gift (Rabbie and Horwitz 1969). Based on an anonymous categorization into two experimental groups, these studies revealed that members of the novel in-group were better rated (e.g. Rabbie and Horwitz 1969) and were favored over members of the novel out-group in their reward allocations (e.g. Tajfel et al. 1971). Even when researchers describe the groups using objectively identical information, perceivers indicate nevertheless that the in-group possesses more favorable attributes than the out-group (Howard and Rothbart 1980).

Two leading explanations of in-group favoritism and out-group discrimination successively emerged in the psychological literature. Social identity theory was first proposed by Tajfel and Turner (1979, 1986) who tried to give a theoretical underpinning to one, and certainly the most salient, conclusion of Tajfel et al. (1971). The latter had been surprised to discover that their subjects –young boys of the same age and from the same school-did not hesitate to sacrifice income to their in-group in order to give the latter a winning position relative to their out-group. They wrote: "in a situation in which subjects' own interests were not involved in their decisions, in which alternative strategies were available that would maximize the total benefits to a group of boys who knew each other well, they

acted in a way determined by an *ad hoc* categorization". In order to explain this puzzling observation, Tajfel and Turner postulated that individuals have a need for identity which has both a personal and a social component. Social identity is defined by groups one belongs to (in-groups) as opposed to groups one doesn't belong to (out-groups) and the social status of these groups relative to one another. Group members have a need for positive social identity that can be fulfilled by favorable comparisons between in-group and out-group members. By establishing positive in-group distinctiveness, the self-concept can be enhanced. However, social identity theory was not confirmed by a number of recent experiments on minimal groups (e.g. Cadinu and Rothbart 1996, Dunning and Hayes 1996, Otten and Wentura 2001). Cadinu and Rothbart (1996) showed that, when almost no information about the groups is available, in-group perception is anchored on self-perception rather than self-perception being based on in-group perception. Cadinu and Rothbart's (1996) self-anchoring theory is a cognitive alternative to the motivational theory of Tajfel and Turner (1986). The self construes the new in-group to be similar to the self. As the self tends to be evaluated positively (e.g., Baumeister 1998), this self-anchoring process typically implies projecting a positive image onto the in-group.

Our discussion of the last section suggests that the higher ability to project oneself onto an in-group than an out-group is solely responsible for in-group boasting or favoritism. Consistent with this theory is Park and Judd's (1990) observation that perceivers refer more frequently to their own behavior when describing an in-group than when describing an out-group during a "think-aloud" procedure. In a similar vein, the false consensus effect (Ross, Greene and House 1977) indicates that perceivers use the self to estimate the prevalence of a particular attribute in the general population. Therefore, the behaviors that perceivers rate as typical in the population likely will be those that they consider to be self-descriptive. This implies that self-knowledge serves as an expectancy for the in-group in a minimal group context (Gramzow, Gaertner and Sedikides 2001). All of these findings suggest that in-group favoritism and out-group discrimination is largely a function of bolstering the in-group rather than debasing the out-group (also, in Brewer 1979).

A fatal blow was given to social identity theory by Yamagishi, Jin and Kiyonari (1999). Following a suggestion of Rabbie, Schot and Visser (1989), they showed that Tajfel et al.'s(1971) puzzling observation critically depended on their special experimental design in which each subject allocated rewards to two other subjects and, consequently, his own reward was determined by others. When subjects allocated rewards between an in-group member and an out-group member without being the target of other subjects' allocation behavior, they gave on average the same amount of money to

the two groups and in-group favoritism vanished. Yamagishi et al. (1999) conclude from a series of clever experiments that, even in minimal group conditions that preclude reciprocal exchanges between two particular subjects, a system of "generalized exchanges" is taking place. In such setting, people receive favors, but not necessarily from the ones to whom they provided favors. Generalized reciprocity of this kind is a direct implication of the prevalence of self-projection[36] onto unknown others in minimal groups.

3. Self-anchored altruism

Self-projection with categorization implies a specific form of altruism which is consistent with self-anchoring (Cadinu and Rothbart 1996) or generalized reciprocity (Yamagishi et al. 1999) and may thus be called "self-anchored altruism". Let us consider a simple reward allocation problem in which the allocator chooses one distribution of rewards between Self and Other, or between other members of an in-group and an out-group (with a fixed reward for herself, in the latter case). Several economists have recently used the first design to elicit social preferences (e.g., Andreoni and Miller 2002, Charness and Grosskopf 2001, Charness and Rabin 2002) and Yamagishi et al. (1999) have used the second design to replicate Tajfel et al.'s (1971) in a "more minimal" group condition. Participants in some experiments were explicitly made to view themselves as playing the two roles alternatively. Charness and Grosskopf (2001, study 2) asked them to make decisions as if they were in one role given that, for payment purposes, their actual role would be determined at the end of the session. Charness and Rabin (2002) told their participants that they would be playing the same game a second time in the other role with another anonymous player. In the experimental conditions of anonymous relations and minimal groups, allocators can be described as impartial judges seeking to identify with the beneficiaries of their rewards under incomplete information and choosing the feasible distribution (x_i, x_j) which maximizes their own type (1)-social utility function

$$V(x_i, x_j) = \frac{1}{2} EU_i(w + x_i) + \frac{1}{2} EU_j(w + x_j) \qquad (9)$$

Equation (9) postulates that the (non- indexed) allocator mentally reincarnates in the identities of her two beneficiaries and projects her own initial wealth onto them. The allocator assesses each beneficiary's similarity with Self on the basis of the information that she received about these identities, and attributes to each beneficiary an expected utility

[36] as defined in section 5.1. Yamagishi et al. (1999: 181) give another meaning to this word, discussed in note 10.

$$EU_k = \lambda_k U + (1-\lambda_k)\overline{U}\,, \qquad (10)$$

with: $k = (i, j)$ and $0 \le \lambda_k \le 1$. The expected utility of an other's final wealth is a weighted average of one's own concave utility U and a reservation utility level $\overline{U}$ which is out of control to the allocator. λ_k is the subjective probability that beneficiary k be similar to self, and takes value 1 when the beneficiary coincides with Self. If an allocator has no information whatsoever or exactly the same information about the beneficiaries of her rewards, she will infer that $\lambda_i = \lambda_j$ and choose a fair allocation, whatever group they belong to. But if the allocator thinks that beneficiary i is more likely to be an in-group member and j an out-group member, she will infer that $\lambda_i > \lambda_j$, which means that she can project better onto her in-group than her out-group. Consistent with this interpretation, researchers often manipulate empathy-altruism by asking subjects to imagine how the other feels or by making the latter perceive their similarity with the other (see section 8.3). It is easily derived from (9) and (10) that the choice of an allocation of rewards by subjects is described (if $\lambda_i \neq 0$) through the maximization of the linear "altruistic" utility

$$W(x_i, x_j) = U(w + x_i) + \lambda U(w + x_j)\,, \qquad (11)$$

with $\lambda = \lambda_j / \lambda_i \quad (0 \le \lambda \le 1)$.

The results obtained for the reward allocation problems in the recent experiments are consistent with the "self-anchored altruism" resulting from incomplete information about others. A person will prefer keeping money for herself than giving it to an unknown other; and she will prefer giving money to an in-group than to an out-group. However, she may sacrifice money to the benefit of an unknown other, or sacrifice her in-group to the benefit of her out-group, if this sufficiently raises the social surplus. The more she can identify with, or project onto, an other the more generous she will be.

What recent experiments show is that a majority of subjects care for others and many are willing to sacrifice money to maximize the social surplus. In sharp contrast with the results reported by Tajfel and his co-authors which gave rise to social identity theory, people do not maximize the difference of rewards between their in-group and their out-group. For instance, denoting an allocation of rewards by (Other, Self), everybody preferred (800,200) to (0, 0); 89% preferred (600, 600) to (400, 600); and

still 66% preferred (900, 600) to (600, 600)[37]. This suggests that one-third of subjects at most exhibited narrow self-interest (*i.e.*, $\lambda = 0$) or difference aversion (*i.e.*, $\lambda < 0$). Non-negligible degrees of difference aversion are practically ruled out by the unanimous preference for the strong Pareto-improvement (800, 200) over (0, 0), but narrow self-interest is consistent with the occasional rejection of a weak Pareto-improvement which only benefits Other. Subjects who cared for others formed a large majority but exhibited a variable degree of altruism. In Charness and Rabin (2002), 67% preferred (300, 600) to (700, 500), showing that one-third projected themselves onto others to the point of sacrificing 100 to increase an other's reward by 400 and another third at least accepted to sacrifice less than 100 to see an other's reward increase by 400. Another interesting result from the same study is that 54% preferred an equal allocation of 575 to (Other, Other, Self) than the unequal allocation (900, 300, 600) even though, with the second allocation, the inequality concerned the two others and Self was sure to get 25 more than with the first allocation. Equation (11) implies for the modal choice:

$$U(575) + 2\lambda U(575) \geq U(600) + \lambda[U(900) + U(300)]$$

This condition will be met with a sufficient combination of risk aversion and altruism. In general, the combination of risk aversion and self-anchored altruism implies that people care more about others who are relatively worse off ("charity") and that they especially have little taste for being themselves at a relative disadvantage. When both implications are tied together, the model's predictions will come close to inequality aversion à la Fehr and Schmidt (1999) without implying, though, that people may engage in Pareto-damaging inequality reduction.

4. Comparing behavior in social dilemmas and in social choices

Our review of psychological research clearly reveals that empathy-altruism and equity- fairness are two distinct pro-social motives, even though they may both derive from the same basic mechanisms of social cognition. Moral action is defined by its relation to some evaluative standard or social norm. Empathy-altruism provides no such standard; it provides a partial identification with an other's welfare. Thus altruism and fairness can be in conflict. Batson et al. (1995a) showed experimentally that inducing empathy for one of the individuals a person can help, but only at the expense of others, can lead the person to show partiality toward that individual, consciously violating the moral principle of fairness. In a somewhat less vivid fashion, the experiments of Blount (1995) and Offerman (2002) bring further evidence on this matter. They compare reciprocity-free choices of

[37] The first result is given by Charness and Rabin (2002) and the other two by Charness and Grosskopf (2001).

allocations with reciprocal behavior. Blount only studies negative reciprocity, while Offerman also includes positive reciprocity. The subjects observed by Offerman (2002) responded to an allocation (Other, Self) chosen by a proposer among a set of two possibilities (8, 14) and (11, 6). The first choice by the proposer, which gave 14 to the subject, was relatively "helpful" and the second choice, which gave him only 6, was relatively "hurtful". After receiving one proposal, subjects decided whether they would accept it as it is or modify it. For a small cost of 1, they were entitled to modify the sum received by the proposer by a larger amount of 4 in either direction. Increasing this sum is helpful to the proposer, and decreasing is hurtful. Starting from a helpful proposal (8, 14), subjects could thus make a helpful response $(8+4,14-1) \equiv (12, 13)$, maintain the status quo (8, 14), or make a hurtful response $(8-4,14-1) \equiv (4, 13)$. Starting from a hurtful proposal (11, 6), the helpful response was (15, 5), the status quo (11, 6), and the hurtful response (7, 5). There were two treatments. In the "Nature" treatment, proposals were randomized. A subject who receives a random proposal should not view himself as responding to an intentional proposal and should take the offer as given by Nature. Thus he should be left with a social choice between three possible allocations. With a social utility function like (11), he would never hurt an other who cannot be held responsible for her choice, since this is a dominated option, and he would even be willing to help her if he were sufficiently altruist. In the "Flesh and Blood" treatment, proposals were intentional. A subject who receives an intentional proposal should reason that it was preferred by an other to an alternative proposal. He should thus act like a responder in a modified ultimatum game offering an opportunity to reciprocate in either direction (i.e., "propose back"). On getting the helpful proposal (8, 14), he would be sure that the initial proposer respected the social norm. As a result, he would never reject a fair proposal by choosing a hurtful response. He would either opt for the more equal sharing (12, 13), on the ground that the first player did not have the opportunity to give equal shares right away, or for the status quo. After getting the hurtful proposal (11, 6), he might want to hurt back at a small cost by choosing (7, 5), as this is the way of rejecting an initial proposal which violated the norm. He might as well want to accept the offer of 6, but he would not want to help at a cost as this is a dominated option for accepting the offer. The asymmetry of fairness-driven responses after either a helpful or a hurtful proposal in the Flesh and Blood treatment contrasts with the relative independence of reactions of the subjects who were exposed to either a helpful or a hurtful proposal in the Nature treatment. This prediction is consistent with Offerman's (2002) main result: his subjects were 67% more likely to hurt after being hurt by an other than by luck; but they were only 25% more likely to help after being helped by an other than by luck. Another result worth noticing is

that, even in the Nature treatment, subjects have a weak tendency to reciprocate. No selfish or difference averse reaction can be found after a helpful unintentional proposal, whereas about one-third of participants to the Charness and Rabin's (2002) experiments seemed to show this type of preferences. On the other hand, a small fraction (17%) of hurtful responses can be found after a hurtful unintentional proposal. One part of the answer lies in the fact that the degree of self-anchored altruism being formed in the Nature treatment depends on the beliefs concerning which group the other belongs to. Both positive and negative surprises are treated as information by subjects and generate both upward and downward revision of the estimated proportion of in-group among participants. For instance, if a helpful proposal indicates a similar other with probability 1, the baseline estimate λ_0 will be revised into: $\lambda_1 = \beta\lambda_0 + (1-\beta)1 > \lambda_0$. If a hurtful proposal indicates a similar other with probability 0, the baseline estimate will be revised into: $\lambda_1 = \beta\lambda_0 + (1-\beta)0 < \lambda_0$. This interpretation is suggested by the literature on minimal groups discussed earlier according to which people use even trivial information about others when they lack more relevant information. Knowing that an other has made a helpful or a hurtful move by the toss of a dice is the kind of trivial information which Rabbie and Horwitz (1969) manipulated in the first minimal group experiment.

VIII. Social drives and emotions

1. *Social comparison*

The tendency of people to judge their own outcome by comparison with some referent is a well-established fact in social psychology (e.g., Crosby 1976, Folger 1986). Often, though not necessarily, this referent is another person. Therefore, outcome satisfaction and perceptions of fairness and equity have little to do with own outcome. Rather, they result from the value of one outcome relative to another. Veblen (1934), Duesenberry (1949) and Easterlin (1974) were the first economists to draw attention on this point. For instance, when average Americans are compared to major league baseball players, the players seem to make princely sums. However, the players often feel inequitably treated (Harder 1992). Presumably, this is because their referents are other major league baseball players and not "average" Americans. In one of the most influential studies of social sciences, Stouffer et al. (1949) described the adjustment of American soldiers during Army life. They observed that agents outside the Air force had low opportunities for promotion but were nevertheless satisfied with their job; by contrast, Air force soldiers, who had much higher opportunities for promotion, were rather dissatisfied. This result is puzzling because promotion opportunities that would seem to

raise income and utility actually brought dissatisfaction. Stouffer and his colleagues reasoned that soldiers were concerned with their relative income. Soldiers outside the Air force were satisfied with their condition because they all followed the same progression and no one was left behind. But soldiers from the Air force who faced greater prospects were often dissatisfied because they could not all be promoted and inevitably some of them would lag behind. In other words, many of those who belonged to the wealthier group were unhappy because they were "relatively deprived" (Davis 1959, Pollis 1968, Crosby 1976, Runciman 1986).

Loewenstein, Thompson and Bazerman (1989) attempted to elicit the individual "social utility functions" by taking the level of self-reported satisfaction as a utility index. They regressed the level of self-reported satisfaction of hypothetical disputants with each of the possible 42 outcomes of the dispute as a function of outcome to self and other. They were able to give within-subjects estimates and found on average a good fit for a function of outcome to self and inequality (the difference between outcome to self and outcome to other) both in quadratic form. They interpreted the satisfaction curve as a "social utility function" which exhibited a strong inequality aversion for disadvantageous inequality and a weaker inequality aversion for advantageous inequality. Subjects always disliked receiving a lower payment than the other party, but their attitude toward advantageous inequality was mixed. They disliked receiving more than friendly people whereas they were satisfied getting more than selfish people. The importance of this work lies in the claim that social utility is indeed observable.[38] More recently, Clark and Oswald (1996) have made a similar claim by showing that job satisfaction is increasing in earnings and decreasing in comparison income, the latter being estimated as the predicted variable of an earnings function. However, Kahneman, Wakker and Sarin (1997) have recently concluded from an "objective" measure of feelings that satisfaction feelings should not be confused with decision-utility, which is the preference index used in modern economics to predict choices. They described satisfaction feelings as "experienced utility", in reference to an alternative concept of utility suggested by Bentham (1789) for measuring pleasure and pain. Now, if the mere observation of satisfaction judgments and feelings does not elicit decision-utility but experienced utility, it is no longer possible to infer from satisfaction data that people make choices that maximize the kind of social utility function exhibited by Loewenstein et al. (1989). This remains an open question.

There is a simple way of interpreting the role played by social comparisons even if the outcome of other did not enter the utility function of self. Using information about others might as well be an

[38] The issues of interpersonal comparability and ordinality of satisfaction scales are not raised here.

economic way of gathering knowledge about self. This is the thrust of Festinger's (1954) theory of social comparison processes contained in these two important propositions: (i) "When an objective, non-social basis for the evaluation of one's ability or opinion is readily available persons will not evaluate their opinions or abilities by comparison with others" (corollary II B, p.120); (ii) "Given a range of possible persons for comparison, someone close to one's own ability or opinion will be chosen for comparison" (corollary III A, p.121). In Festinger's mind, as these two corollaries indicate, comparison to others is essentially a way to acquire self-knowledge when direct information on self is not available or too costly.

It is necessary to understand that social comparisons operate like a *drive* for capturing the role attributed to them by social psychologists in the formation of social preferences. Whatever deviation from their normative expectation people experience in the course of an action operates like a drive. When normative expectations are met by the experienced outcome, people are satisfied with the outcome and feel that they have been treated fairly. As a result, they are more committed and more willing to sacrifice for the social good (Lind and Tyler 1988, Tyler and Lind 1992). On the other hand, when experienced outcomes fall short of their normative expectations, individuals are dissatisfied and angry to have been treated unfairly. They are less willing to cooperate and make efforts, and may even engage in hostile demonstrations like theft and aggression (e.g., Greenberg and Scott 1996). Following Festinger (1957), people react dynamically to the cognitive dissonance that they perceive when their experience does not conform with their normative expectation (for a further discussion of cognitive dissonance and an economic model of dynamic drives, see Lévy-Garboua and Blondel 2002, Lévy-Garboua and Montmarquette 1996).

2. Reducing inequity

For Adams (1963, 1965), who explicitly refers to Festinger (1954, 1957), a social exchange between two agents is deemed equitable when the perceived value of outcomes is proportional to the perceived value of inputs. The formula has later been extended to more agents (e.g. Anderson 1976). The value of outcome to workers is the offered wage inclusive of non- pecuniary income, and it is balanced with the value of their labor inputs, designated as "effort". Both outcome and input are valued as perceived by workers. The "fair" wage (rate) is *the normative expectation* of the wage rate, i.e. the perceived ratio of wage to effort: $fair\ wage \equiv E(\frac{wage}{effort})$.

However, equity theory does not provide a full description of how this normative expectation is formed. Adams implicitly recognizes that equilibrium wage rates in a perfectly competitive economy would be fair because job inputs (like education, experience, and effort) would then perfectly correlate with outcomes (like pay). "Indeed, it is because they are imperfectly correlated that we need at all be concerned with job inequity" (Adams 1963: 424). In the experiments supporting his theory, the norm is described by the observable situation of referent others. But such definition raises a reflection problem (see Manski 2000) since the norm of one person's reference group will generally depend upon this person's situation. It suggests a dynamic adjustment process towards equilibrium, the equity drive. The worker whose wage is lower than that of a comparison worker who works in another accessible firm feels a drive to move into another firm. But the existence of an equity drive does not suffice to determine the norm of fairness. The latter must be given otherwise. We gave a specific example of how the fairness norm might be determined in the previous discussion of Homans' (1953, 1954) cash posters. Another instance is provided by Akerlof and Yellen (1990: section 4) who assume that the fair wage is a weighted average of the wage received by the reference group and the market-clearing wage. Kahneman, Knetsch and Thaler (1986) designed experiments in which the standard of fairness was simply the reference transaction. Once defined, the fair wage functions like a social norm prescribing agents how to behave in social relations and letting them anticipate others' behavior.

Workers being offered a "low" wage feel inequity because this is dissonant with their normative expectation. The subjects of Kahneman, Knetsch and Thaler (1986) confirm this prediction. They thought that workers were entitled to their current wage and the firm was not entitled to expand its profits by setting lower wages. Only when profits are threatened may firms set lower wages. However, the subjects of these experiments could only express their feelings of unfairness. A worker experiencing enough cognitive dissonance will also find effective ways of reducing effort, like shirking, absenteeism, or quitting the job, in order to maintain her wage rate at a fair level. In some circumstances, she might even force an other to increase his effort until both wage rates level-off. Akerlof and Yellen (1990) have made use of this prediction of equity theory to derive a theory of involuntary unemployment. In their model, there exist one equilibrium type in which the wages of low-paid workers are set at a fair level above the market-clearing level (since low-paid workers compare to the high-paid group), which causes unemployment for this category of workers. Employers adopt this behavior because they fear negative reciprocity from their low-paid employees if the latter felt underpaid. This kind of reciprocal behavior was neatly confirmed by the experimental

gift-exchange game of Fehr, Kirschteiger, and Riedl (1993) and by the employers' description of their own wage-setting behavior (Bewley 1999).

Another prediction of equity theory is even more surprising: workers whose wage rate exceeds the norm also feel inequity. Overpayment too is dissonant with their normative expectation so that they will reduce inequity by increasing effort until the wage rate falls back to a fair level. The reciprocal nature of inequity feelings is suggested by Adams' (1963: 427) statement that "whenever inequity exists for Person, it will also exist for Other, provided their perceptions of inputs and outcomes are isomorphic or nearly so".

These two kinds of inequity feelings induce two forms of reciprocal behavior: negative in case of an unfavorable comparison with an other, and positive in case of a favorable comparison. For instance, the worker who feels under-compensated in comparison with a co- worker stands in the position of one responder of a proposal-response game who does not receive her normal share of the surplus. She refuses a proposal that falls short of a prior social agreement and punishes the norm's violator. By contrast, the worker who feels over-compensated in comparison with a co-worker stands in the position of a proposer who committed herself to give a fair share of any surplus she might receive to the responders. She may wish to respect the implicit promise that she made.

We believe that the puzzling observations of Tajfel et al. (1971), discussed in section 7.2, are a good instance of inequity-reducing behavior. Young boys who knew each other well did not hesitate to sacrifice social surplus and income to their in-group in order to give the latter a winning position relative to their out-group when their own income was determined by the likewise allocation of an unknown other. These observations refute the "self-anchored altruism" assumption expressed by equation (11) and need to be explained. What makes Tajfel et al's (1971) experiment interesting is that it provides a rare instance in which subjects are unable to manipulate their own reward, but can manipulate the reward of others. Consequently, whenever maximizing the joint payoff was detrimental to their in-group, Tajfel's boys understood that such behavior would essentially increase their comparison income instead of increasing their own income. Widening the negative gap between own reward and its normative expectation generated feelings of inequity (Adams 1963) or dissatisfaction (Lévy-Garboua and Montmarquette 2003) among these boys. The related emotions of envy and anger eventually drove them to reduce the comparison income under their control in order to reduce feelings of inequity or dissatisfaction. Depending on the circumstances, the equity drive may push individuals to hurt an advantaged friend, punish a norm's violator, reward a generous employer or help a stranger in need.

3. *Helping others in need: is the motivation truly altruistic?*

Do the people who help others in need or distress have an altruistic personality? Answering this question requires a general agreement on the personality measures that should be used. When Staub (1974) and Rushton (1980) first claimed that there is an altruistic personality, the notion of altruism that they used, which only ruled out the individual's quest for external rewards, was criticized for still being too broad and including compliance with internalized social or personal norms. If moral obligation were the reason for helping, the act of helping would not be truly altruistic, many psychologists say, because it would not be ultimately directed toward others (Bar Tal 1976, Batson and Shaw 1991). So the question should be restated: Is there an altruistic personality which goes beyond the disposition to help others for getting peace of mind by avoiding shame and guilt? In other words, those psychologists want to make a distinction between the vicarious emotions of *empathy* and *personal distress*.

Empathy is defined as an other-oriented emotional response congruent with the perceived welfare of another person (e.g., Batson et al. 1995a, Hoffman 1988). The list of emotions associated with empathy includes adjectives like "sympathetic, moved, compassionate, warm, soft-hearted, and tender". Since the empathic emotion stems from the apprehension of another's emotional state to which it is similar (Eisenberg and Strayer 1987), it requires at least a minimal awareness of the differences between self and other. Thus empathy implies self-other merging or identification, which we also described as the economic definition of pure altruism in section 5.1. Contrasting with empathy which is assumed to be truly altruistic, personal distress is a self-oriented emotional response to another person in need or distress. It is usually associated with adjectives like "alarmed, grieved, troubled, distressed, upset, disturbed, worried, and perturbed". The personality measure for empathic concern correlates with the measured disposition for perspective-taking while the personality measure for personal distress correlates with low self-esteem or sadness (Batson et al. 1986). As for many emotions (Zajonc 1980), both empathy and personal distress function like a drive indicating one's current preferences and generating the context-dependent motivation to relieve another person's need. Empathy can be experimentally manipulated by asking subjects to imagine how the other feels (high empathy) versus to take an objective and detached perspective (low empathy) by trying not to get caught up in how the other feels (e.g., Batson et al. 1995a). Empathy can also be manipulated by making subjects perceive their similarity (high empathy) or dissimilarity (low empathy) with the other (e.g., Batson et al. 1995b).

There is extensive evidence that empathy increases helping and other pro-social behavior. However, the empirical significance of this relation seems to depend on how empathy is measured. For instance, using four personality measures of self-esteem, social responsibility, ascription of responsibility, and *dispositional* (i.e., out of context) empathy, Batson et al. (1986) found no evidence that any of these four "altruistic" personality variables was associated with truly altruistic motivation in helping. But Eisenberg and Miller (1987) show by a meta-analysis that this relation is significant for older adolescents and adults (see other references in Eisenberg and Fabes, 1998), when either picture/story or self-report measures of *situational* (i.e., in context) empathy[39] are used (the relation is less clear for children). Similar results using different physiological markers of empathy or personal distress have been obtained by numerous studies, both for children and for adults (see the references in Eisenberg and Fabes, 1998). Lastly, subjects who are induced experimentally to empathize with an other in need or distress help significantly more than those induced to have an impartial attitude. For example, 62% of Dovidio et al.'s (1990) empathy-induced subjects helped versus only 34% for those in the low empathy condition.

In order to distinguish whether the personality measures are associated with a truly altruistic motivation or with compliance to norms, Batson (e.g., 1991) compared helping behavior in two treatments. In one condition, subjects had an easy escape to helping another in need, while escape was difficult in the other condition. An altruist wants to help because she identifies with the other and derives utility from his relief. Making escape easy will not change her motivation. By contrast, a person who just feels morally obliged to help under pressure is less likely to do so when offered an easy escape because she then finds herself in the position of a dictator in a proposal-response game (see the discussion in section 5.1). Indeed, Batson (1991) found that empathy was more likely related to helping than is personal distress when it is easy to escape contact with the needy person. This is consistent with his "empathy-altruism hypothesis" that empathy is mainly other-oriented, whereas personal distress is self-oriented. For instance, 68% of Batson et al.'s (1995a) participants to the two experiments in the high empathy condition had an altruistic motivation for helping versus only 37% in the low empathy condition.

Although the empathy-altruism hypothesis seems to have gained weight in recent years among psychologists, there is still no consensus about whether the motivation for helping is truly altruistic

[39] A subjective index of empathy is obtained by showing subjects a picture or a story that represents another person in a situation of need or distress and asking them how they feel about it. A subject is supposed to empathize with the other person when his or her reported emotion, either is close to what the picture was meant to convey by the experimenter or scores high on a subjective scale. Objective sympathy indexes using physiological markers like heart rate, skin conductance, or facial reactions can also be found (see Eisenberg and Fabes, 1998).

or "egoistic". Cialdini, Kenrick and Baumann (1982) exemplify the egoistic approach with their "negative state relief" model. They suggest that negative affects, like sadness, can motivate helping because helping can be perceived as an instrumental act that will relieve these negative feelings. The helping motivation of a "sad" person is egoistic and can be removed by the anticipation of another mood-enhancing event, such as listening to a comedy tape or having the opportunity to help another person (Schaller and Cialdini 1988). However, the results of Dovidio et al. (1990) suggest that empathic concern associated with a problem mediates helping on that problem in a way that is independent of sadness.

IX. Some lessons from psychology and biology: A summary

Which lessons for economic research can be drawn from our survey of the social-psychological and biological literature dealing with the formation of social preferences? The first lesson is that biological evolution can explain the emergence of restricted forms of social preferences like altruism toward close relatives, but multiple and often delicate conditions are found for the evolutionary emergence of social adaptation at large. For perhaps one million existing animal species, ten thousand at most are social in any significant way. As we move up the evolutionary ladder, environmental- as opposed to genetic- factors increasingly come into play and account for an increasing intra-specific variability in social adaptation. Cultural transmission, for instance, is a human-specific type of learning which is transmitted to next generations by making a pool of cultural traits to co-evolve with the gene pool.

Studies on pro-social development conducted by social psychologists since Piaget's seminal work generally confirm that pro-social dispositions are not innate. Although learning and enforcement of pro-social values may seem an obvious explanation, the empirical and experimental evidence is disappointing. Children do a lot more than copying models; they construct their social worlds in terms of cognitive structures. Their pro-social dispositions seem to develop mainly during the first 10 or 12 years of existence, together with cognitive and emotional skills like perspective-taking and empathy. The formation of the perspective–taking ability is probably distinctive of human sociality. Ants and bees, which have a detailed division of labor, have a social life; but they don't have a social mind.

The full development of social preferences requires consciousness of the individual's similarities and differences with others, and therefore knowledge of self and others. The frequent asymmetry of

one's knowledge of self and others is the origin of two distinct cognitive processes for generating social preferences: identification of self with known others, and projection of known self onto partially unknown others. These two basic mechanisms of social cognition obviously require perspective-taking skills. They combine with the process of categorizing others with whom an individual interacts into similar others (in-group) and dissimilar others (out-group) to form a variety of social preferences. The self can project onto similar others but is unable to do so onto dissimilar others. Thus the self will find it easier to internalize and predict the behavior of an in-group than an out-group in social interactions and will generally like to interact more with the former than with the latter. The more can the self identify with, or project onto, an other the more generous she will be. Several context-dependent pro-social motives may derive from the same basic mechanisms of social cognition, and the social-psychological literature commonly distinguishes fairness and empathy-altruism. The fairness motive and the empathy-altruism motive are quite different and can even be in conflict.

Fairness is a social norm, present in proposal-response games, that functions like an enforceable implicit contract among a group of players. It brings to all players a precise knowledge of others' intentions and an incentive to respect their own intentions. However, it often needs to be enforced by sanctions. Self-anchored altruism originates from the partial identification of self with an other's welfare, present in a dictator game, by a purely cognitive process. Under incomplete information, Self finds it easier to identify with in-group than out-group.

However, both fairness and empathy-altruism can also arise from the emotional response to the perceived inequity or dissatisfaction of an experienced deviation from one's normative expectation, whether the latter is common to one group or specific to one individual. Disadvantaged responders feel angry and are driven to hurt or punish the norm's violator; symmetrically, many people empathize with the need or distress of another person and are driven to help. Such affective modes of response coexist with the cognitive processes of social cognition to form a rich variety of context-dependent social preferences.

References

Adams, G.R. (1983) "Social Competence During adolescence: Social Sensitivity, Locus of Control, Empathy, and Peer Popularity", *Journal of Youth and Adolescence 12*, 203-211.

Adams, S.J. (1963) "Toward an Understanding of Inequity", *Journal of Abnormal and Social Psychology* 67, 422-436.

Adams, S.J. (1965) "Inequity in Social Exchange", in *Advances in Experimental Social Psychology* (Vol. 2), L. Berkowitz (ed.), New York: Academic Press, 267-299.

Akerlof, G.A. and J.L. Yellen (1990) "The Fair Wage- Effort Hypothesis and Unemployment", *Quarterly Journal of Economics 105*, 255-284.

Alicke, M.D. and E. Largo (1995) "The Role of Self in the False Consensus Effect", *Journal of Experimental Social Psychology 31*, 28-47.

Allison, S.T. and D.M. Messick (1990) "Social Decision Heuristics in the Use of Shared Resources", *Journal of Behavioral Decision Making 3*, 195-204.

Allison, S.T., McQueen, L.R. and L.M. Schaerfl (1992) "Social Decision Making Processes and the Equal Partitionment of Shared Resources", *Journal of Experimental Social Psychology 28*, 23-42.

Anderson N. (1976) "Information Integration Theory applied to Equity Judgments", *Journal of Personality and Social Psychology 33*, 291-299.

Andreoni, J. and J. Miller (1998) "Giving according to GARP: An Experimental Test of the Consistency of Preferences for Altruism", *Econometrica 70*, 737-753.

Arrondel L. and A. Masson (2001) "Family Transfers Involving Three Generations", *Scandinavian Journal of Economics 103*, 415-443.

Axelrod R. and W.D. Hamilton (1981) "The evolution of cooperation", *Science 211*, 1390-1396.

Bandura, A. (1986) *Social Foundations of Thought and Action*, Engelwood Cliffs, NJ: Prentice-Hall.

Barnett S.A. (1968) "The instinct to teach", *Nature 220*, 747-749.

Bar-Tal, D. (1976) *Prosocial Behavior – Theory and Research*, New-York: John Wiley and Sons.

Bar-Tal, D., Korenfeld, D. and A. Raviv (1985) "Relationships Between the Development of Helping Behavior and the Development of Cognition, Social Perspective, and Moral Judgment", *Genetic, Social, and General Psychology Monographs 11*, 23-40.

Batson,C.D., Bolen, M.H., Cross,J.A. and H.E. Neuringer-Benefiel (1986) "Where is the Altruism in the Altruistic Personality?", *Journal of Personality and Social Psychology 50*, 212-220.

Batson, C.D., J.G. Batson, C.A. Griffitt, S. Barrientos, J.R. Brandt, P. Sprengelmeyer and M.J. Bayly (1989) "Negative-state Relief and the Empathy-Altruism Hypothesis", *Journal of Personality and Social Psychology 56*, 922-933.

Batson, C.D. (1991) *The Altruism Question: Toward a Social-Psychological Answer*, Hillsdale, NJ: Erlbaum.

Batson, C.D., Batson, J.G., Slingsby, J.K., Harrell, K.L., Peekna, H.M. and R.M. Todd (1991) "Empathic Joy and the Empathy-Altruism Hypothesis", *Journal of Personality and Social Psychology 61*, 413-426.

Batson, C.D. and L.L. Shaw (1991) "Evidence for Altruism: Toward a Pluralism of Prosocial Motives", *Psychological Inquiry 2*, 107-122.

Batson, C.D., Allison, S.T., Klein, T.R., Highberger, L. and L.L. Shaw (1995a) "Immorality from Empathy-Induced Altruism: When Compassion and Justice Conflict", *Journal of Personality and Social Psychology 68*, 1042-1054.

Batson, C.D., Turk, C.L., Shaw, L.L. and T.R. Klein (1995b) "Information Function of Empathic Emotion: Learning That We Value the Other's Welfare", *Journal of Personality and Social Psychology 68*, 300-313.

Batson, C.D. (1997), "Self-other Merging and the Empathy-Altruism Hypothesis: Reply to Neuberg et al. (1997)", *Journal of Personality and Social Psychology 73*, 517-522.

Batson, C.D., Sager, K., Garst, E., Kang, M., Rubchinsky, K. and K. Dawson (1997) "Is Empathy-induced Helping Due to Self-other Merging?", *Journal of Personality and Social Psychology 73*, 495-509.

Baumeister, R.F. (1998) "The Self", in *The Handbook of Social Psychology* (Vol. 1), D.T. Gilbert, S.T. Fiske, and G. Lindzey (eds.), New York: Mac Graw Hill, 680-740.

Becker, G.S. (1974) "A Theory of Social Interactions", *Journal of Political Economy 82*, 1063-1093.

Becker, G.S. and Barro, R.J. (1988) "A Reformulation of the Economic Theory of Fertility", *Quarterly Journal of Economics 103*, 1-25.

Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation*, reprinted in 1948, Oxford: Blackwell.

Bewley, T. (1999) *Why Wages Don't Fall During a Recession*, Cambridge, Mass.: Harvard University Press.

Binmore K. (1992) *Fun and Games*, D.C. Heath and Company, Lexington.

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

Birch, L.L. and J. Billman (1986), "Preschool Children's Food Sharing with Friends and Acquaintances", *Child Development 57*, 387-395.

Bisin, A. and T. Verdier (2001) "The Economics of Cultural Transmission and the Dynamics of Preferences", *Journal of Economic Theory 97*, 298-319.

Blount, S. (1995) "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences", *Organizational Behavior and Human Decision Processes 63*, 131-144.

Bolton, G.E., and A. Ockenfels (2000) "ERC: A Theory of Equity, Reciprocity, and Competition", *American Economic Review 91*, 166-93.

Boorman S.A. and P.R. Levitt (1980) *The Genetics of Altruism*, New York: Academic Press

Borke, H. (1978) "Piaget's View of Social Interaction and the Theoretical Construct of Empathy", in *Alternatives to Piaget – Critical Essays on the Theory*, L.S. Siegel and C.J. Brainerd (eds.), New-York: Academic Press, 29-42.

Bowles S., Gintis H. (2003) "The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations", forthcoming in *Theoretical Population Biology*.

Boyd R. and P.J. Richerson (1985) *Culture and the Evolutionary Process*, Chicago: The University of Chicago Press.

Brewer, M.B. (1979) "In-group Bias in the Minimal Intergroup Situation: A Cognitive-Motivational Analysis", *Psychological Bulletin 86*, 307-324.

Brockmann H.J. and R. Dawkins (1979) "Join nesting in a digger wasp as an evolutionary stable preadaptation to social life", *Behavior 71*, 203-245.

Brockmann H.J. (1984) "The evolution of social behavior in insects", in Kreps J.K. and Davies N.B. (eds), *Behavioural Ecology: An Evolutionary Approach*, 2nd ed., 340-61, Blackwell, Oxford.

Cadinu, M.R. and M. Rothbart (1996) "Self-anchoring and Differentiation Processes in the Minimal Group Setting", *Journal of Personality and Social Psychology 70*, 661-677.

Camerer, C. and R.H. Thaler (1995) "Ultimatum, Dictators and Manners", *Journal of Economic Perspectives 9*, 209-219.

Campbell, R.L., and J.C. Christopher (1996) "Moral Development Theory: A Critique of its Kantian Presuppositions", *Developmental Review 16*, 48-68.

Carpendale, J.I.M. (2000), "Kohlberg and Piaget on Stages and Moral Reasoning", *Developmental Review 20*, 181-205.

Cavalli-Sforza L.L. and M.W. Feldman (1981) *Cultural Transmission and Evolution: A Quantitative Approach*, Princeton: Princeton University Press.

Charness, G. and B. Grosskopf (2001) "Relative Payoffs and Happiness: An Experimental Study", *Journal of Economic Behavior and Organization 45*, 301-328.

Charness, G. and M. Rabin (2002) "Understanding Social Preferences with Simple Tests, *Quarterly Journal of Economics 117*, 817-869.

Chase-Lansdale, P.L., Wakschlag, L.S. and J. Brooks-Gunn (1995) "A Psychological Perspective on the Development of Caring in Children and Youth: the Role of the Family", *Journal of Adolescence 18*, 515-556.

Cialdini, R.B., Schaller, M., Houlihan, D., Arps, K., Fultz, J. and A.L. Beaman (1987) "Empathy-based Helping: Is It Selflessly or Selfishly Motivated?" *Journal of Personality and Social Psychology 52*, 749-758.

Cialdini, R.B. (1991) "Altruism or Egoism? That is (still) the Question", *Psychological Inquiry 2*, 124-126.

Cialdini, R.B., Brown, S.L., Luce, C., Lewis, B.P. and S.L. Neuberg (1997) "Reinterpreting the Empathy-Altruism Relationship: When One into One Equals Oneness", *Journal of Personality and Social Psychology 73*, 481-494.

Clark, A.E. and A.J. Oswald (1996) "Satisfaction and Comparison Income", *Journal of Public Economics 61*, 359-381.

Colby, A. and L. Kohlberg (eds.). (1987) *The Measurement of Moral Judgment* (vol. 1-2), Cambridge: Cambridge University Press.

Cosmides L. and J. Tooby (1992) "Cognitive adaptations for social exchange", in Barkow J., Cosmides L. and Tooby J. (eds.), *The Adapted Mind,* New York: Oxford University Press, 163-228.

Cottrell, L.S. Jr. and R.F. Dymond (1949) "The Empathic Process", *Psychiatry 12*, 355-359.

Crosby, F. (1976) "A Model of Egoistical Relative Deprivation", *Psychological Review 83*, 85-113.

Damon, W. (1977), *The Social World of the Child* , San Francisco: Jossey-Bass.

Dawes, R.M. (1980) "Social Dilemmas", *Annual Review of Psychology 31*, 169-193.

Dawkins R. (1976) *The Selfish Gene*, Oxford: Oxford University Press.

Dawkins R. (1982) *The Extended Phenotype*, Oxford: Oxford University Press.

Deci, E.L. (1971) "The Effects of Externally Mediated Rewards on Intrinsic Motivation", *Journal of Personality and Social Psychology 18*, 105-115.

Dennett D.C. (1987) *The Intentional Stance*, Cambridge Mass.: The MIT Press.

Denton, K. and D. Krebs (1990) "From the Scene to the Crime: the Effect of Alcohol and Social Context on Moral Judgment", *Journal of Personality and Social Psychology 59*, 242-248.

Dolan, P. and A. Robinson (2001) "The Measurement of Preferences over the Distribution of Benefits", *European Economic Review 45*, 1697-1709.

Dovidio, J.F., Allen, J.L. and D.A. Schroeder (1990) "Specificity of Empathy-induced Helping: Evidence for Altruistic Motivation", *Journal of Personality and Social Psychology 59*, 249-260.

Dovidio, J.F. (1991), "The Empathy-Altruism hypothesis: Paradigm and Promise", *Psychological Inquiry 2*, 126-128.

Duesenberry, J.S. (1949) *Income, Saving and the Theory of Consumer Behavior*, Cambridge, Mass.: Harvard University Press.

Dufwenberg, M. and G. Kirchsteiger (1999) "A Theory of Sequential Reciprocity", Discussion paper, CentER: Tilburg University.

Dunning, D. and A.G. Hayes (1996) "Evidence for Egocentric Comparison in Social Judgment", *Journal of Personality and Social Psychology 71*, 213-229.

Eagly A.H. and M. Crowley (1986) "Gender and Helping Behavior: A Meta-analytic Review of the Social Psychological Literature", *Psychological Bulletin 100*, 283-308.

Easterlin, R.A. (1974) "Does Economic Growth Improve the Human Lot? Some Empirical Evidence", in *Nations and Households in Economic Growth*, P.A. David and M.W. Reder (eds.), New York: Academic Press, 89-125.

Eisenberg, N. (1986) *Altruistic Emotion, Cognition, and Behavior* , Hillsdale, NJ: Erlbaum.

Eisenberg, N. and P.A. Miller (1987) "The Relation of Empathy to Prosocial and Related Behaviors", *Psychological Bulletin 101*, 91-119.

Eisenberg, N. and J. Strayer (1987) "Critical Issues in the Study of Empathy", in *Empathy and Its Development*, N. Eisenberg and J. Strayer: (ed.), Cambridge: Cambridge University Press.

Eisenberg, N. and P. Mussen (1989) *The Roots of Prosocial Behavior in Children*, Cambridge: Cambridge University Press.

Eisenberg N. and R.A. Fabes (1995) "The Relation of Young Children's Vicarious Emotional Responding to Social Competence, Regulation, and Emotionality", *Cognition and Emotion 9*, 203-228.

Eisenberg, N. (1996) "Caught in a Narrow Kantian Perception of Prosocial Development: Reactions to Campbell and Christopher's Critique of Moral Development Theory", *Developmental Review 16*, 1-47.

Eisenberg N., Fabes, R.A., Karbon, M., Murphy, B.C., Carlo, G. and M. Wosinski (1996) "The Relations of Children's Disposal Comforting Behavior to Empathy-related Reactions and Shyness", *Social Development 5*, 330-351.

Eisenberg N. and R.A. Fabes (1998) "Prosocial Development", in *Handbook of child psychology: Socialization, Personality, and Social Development*, N. Eisenberg (ed.), New York: Wiley, 701-778.

Eisenberg, N., Zhou, Q. and S. Koller (2001) "Brazilian Adolescents' Prosocial Moral Judgment and Behavior: Relations to Sympathy, Perspective-taking, Gender-role Orientation, and Demographic Characteristics", *Child Development 72*, 518-534.

Falk, A. and U. Fischbacher (1999) "A Theory of Reciprocity", Working paper, Institute for Empirical Research in Economics: University of Zurich.

Falk A., Fehr E., Fischbacher U. (2001). "Driving Forces of Informal Sanctions", WP 59, Institute for Empirical Research in Economics, University of Zurich.

Fehr, E., Kirchsteiger, G. and A. Riedl (1993) "Does Fairness Prevent Market Clearing? An Experimental Investigation", *Quarterly Journal of Economics 58*, 437-460.

Fehr, E. and K.M Schmidt (1999) "A Theory of Fairness, Competition, and Cooperation", *Quarterly Journal of Economics 114*, 817-68.

Fehr, E. and S. Gächter (2000) "Cooperation and Punishment in Public Goods Experiments", *American Economic Review 90*, 980-994.

Fehr E., Gächter S. (2002) "Altruistic Punishment in Humans", *Nature 415*, 137-140.

Fehr, E and A. Falk (2002) "Psychological Foundations of Incentives", *European Economic Review 46*, 687-724.

Fehr E., Fischbacher U., Gächter S. (2002) "Strong Reciprocity, Human Cooperation and the Enforcement of Social Norms", *Human Nature 13*, 1-25.

Ferguson, C.K. and H.H. Kelley (1964) "Significant Factors in Overvaluation of Own Group's Product", *Journal of Abnormal Social Psychology 69*, 223-228.

Festinger, L. (1954) "A Theory of Social Comparison Processes", *Human Relations 7*, 117-140.

Festinger, L. (1957) *A Theory of Cognitive Dissonance*, Stanford, Cal.: Stanford University Press.

Fisher R.A. (1930) *The Genetical Theory of Natural Selection*, Oxford: Clarendon Press.

Flavell, J.H. (1992) "Cognitive Development: Past, Present, and Future", *Developmental Psychology 28*, 998-1005.

Folger, R. (1977) "Distributive and Procedural Justice: Combined Impact of "Voice" and Improvement on Experienced Inequity", *Journal of Personality and Social Psychology 35*, 108-119.

Folger, R., Rosenfield, D., Grove, J. and L. Corkran (1979) "Effects of "Voice" and Peer Opinions on Responses to Inequity", *Journal of Personality and Social Psychology 37*, 2253-2261.

Folger, R. (1986) "Rethinking Equity theory: A Referent Cognitions Model", in *Justice in Social Relations*, H.W. Bierhoff, R.L. Cohen and J. Greenberg (eds.), New York: Plenum, 145-162.

Frankenberger, K.D. (2000) "Adolescent Egocentrism: A Comparison between Adolescents and Adults", *Journal of Adolescence 23*, 343-354.

Freud, S. (1968 [1923]) "Le Moi et le Ca", in *Essais de psychanalyse*, Paris: Petite Bibliothèque Payot.

Frey, B.S. (1997) *Not Just for the Money. An Economic Theory of Personal Motivation*, Cheltenham: Edward Elgar.

Gächter, S. and E. Fehr (1999) "Collective Action as a Social Exchange", *Journal of Economic Behavior and Organization 39*, 341-369.

Gilligan, C. (1982) *In a Different Voice: Psychological Theory and Women Development*, , Cambridge, Mass: Harvard University Press.

Gintis H. (2000a) *Game Theory Evolving,* Princeton: Princeton University Press.

Gintis H. (2000b) "Strong Reciprocity and Human Sociality", *Journal of Theoretical Biology* 206, 169-170.

Gintis H. (2003) "The Hitchhiker's Guide to Altruism: Gene-Culture Coevolution and the Internalization of Norms", *Journal of Theoretical Biology 220*, 407-418.

Gintis H., Bowles S., Boyd R., Fehr E. (2003) "Explaining Altruistic Behavior in Humans", *Evolution and Human Behavior 24*, 153-172.

Glassman, M. and B. Zan (1995) "Moral Activity and Domain Theory: An Alternative Interpretation of Research with Young Children", *Developmental Review 15*, 434-457.

Gramzow, R.H., Gaertner, L. and C. Sedikides (2001) "Memory for In-Group and Out-Group Information in a Minimal Group Context: the Self as an Informational Base", *Journal of Personality and Social Psychology 80*, 188-205.

Greenberg, J. and K.S. Scott (1996) "When Do Workers Bite the Hands that Feed Them? Employee Theft as a Social Exchange Process", in *Research in Organizational Behavior* (Vol. 18), B.M. Staw and L.L. Cummings (eds.), Greenwich, CT: Jai Press, 111-156.

Grusec, J.E. and E. Redler (1980) "Attribution, Reinforcement, and Altruism: A Developmental Analysis", *Developmental Psychology 16*, 525-534.

Grusec, J.E. (1981), "Socialization Processes and the Development of Altruism", in *Altruism and Helping Behavior: Social, Personality, and Developmental Perspectives*, J.P. Rushton and R.M. Sorrentino (eds.), Hillsdale, NJ: Erlbaum, 65-90.

Güth, W., Schmittberger, R. and B. Schwarze (1982) "An Experimental Analysis of Ultimatum Bargaining", *Journal of Economic Behavior and Organization 3*, 367-88.

Güth, W. (1995) "On Ultimatum Bargaining Experiments - A Personal Review", *Journal of Economic Behavior and Organization 27*, 329-44.

Habermas, J. (1979) *Communication and the Evolution of Society*, Boston: Beacon Press.

Haldane J.B.S. (1953) "Animal populations and their regulation", *Penguin Modern Biology 15*, 9-14

Hamilton W.D. (1964) "The genetical evolution of social behavior", *Journal of Theoretical Biology* 7, 1-52.

Hampson, R.B. (1984) "Adolescent Prosocial Behavior: Peer Group and Situational Factors Associated with Helping ", *Journal of Personality and Social Psychology 46*, 153-162.

Harbaugh, W.T., Krause, K. and S.G. Linday Jr. (2002) "Children's Bargaining Behavior", Working Paper, University of Oregon.

Harder, J.W. (1992) "Play for Pay: Effects of Inequity in a Pay-for-Performance Context", *Administrative Science Quarterly 37*, 321-335.

Harsanyi, J. (1955) "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility", *Journal of Political Economy 63*, 309-321.

Henrich, J. (2000) "Does Culture Matter in Economic Behavior? Ultimatum Game Bargaining Among the Machiguenga of the Peruvian Amazon", *American Economic Review 90*, 973-979.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E. and H. Gintis (2003) *Foundations of Human Sociality: Ethnography and Experiments in Fifteen Small-scale Societies*, Oxford: Oxford University Press.

Higgins, E.T (1981) "Role-taking and Social judgment: Alternative Developmental Perspectives and Processes", in *Social Cognitive Development: Frontiers and Possible Futures*, J.H. Flavell and L. Ross (eds.), Cambridge: Cambridge University Press, 119-153.

Hoffman E., McCabe K.A. and V.L. Smith (1998) "Behavioral foundations of reciprocity: experimental economics and evolutionary psychology", *Economic Inquiry 38*, 335-352.

Hoffman, M.L. (1975a) "Developmental Synthesis of Affect and Cognition and its Implications for Altruistic Motivations", *Developmental Psychology 11*, 607-622.

Hoffman, M.L. (1975b) "Altruistic Behavior and the Parent-child Relationship", *Journal of Personality and Social Psychology 31*, 937-943.

Hoffman, M.L. (1981) "The Development of Empathy", in *Altruism and Helping Behavior: Social, Personality, and Developmental Perspectives*, J.P. Rushton and R.M. Sorrentino (eds.), Hillsdale, NJ: Erlbaum, 41-63.

Hoffman, M.L. (1988) "Moral Development", in *Developmental Psychology: An Advanced Textbook*, M.H. Bornstein and M.E. Lamb (eds.), (2nd edition), Hillsdale, NJ: Erlbaum, 497-548.

Homans, G.C. (1953) "Status among Clerical Workers", *Human Organization 12*, 5-10.

Homans, G.C. (1954) "The Cash Posters", *American Sociological Review 19*, 724-733.

Hornstein, H.A. (1991) "Empathic Distress and Altruism: Still Inseparable", *Psychological Inquiry 2*, 133-135.

Hume D. (1969) *A Treatise of Human Nature*, London: Penguin Books Ltd.

Ingold T. (1986) *Evolution and Social Life*, Cambridge: Cambridge University Press.

Jellal, M. and F.C. Wolff (2002) "Dynamique des transferts intergénérationnels et effet de démonstration", Mimeographed, University of Nantes.

Jetten, J., Spears, R. and A.S.R. Manstead (1996) "Intergroup Norms and Intergroup Discrimination: Distinctive Self-Categorization and Social Identity Effects", *Journal of Personality and Social Psychology 71*, 1222-1233.

Kahneman, D., Knetsch, J.L. and R. Thaler (1986) "Fairness as a Constraint on Profit-Seeking: Entitlements in the Market", *American Economic Review 76*, 728-741.

Kahneman, D., P.P. Wakker and R. Sarin (1997) "Back to Bentham? Explorations of Experienced Utility", *Quarterly Journal of Economics 112*, 325-405.

Kandel, E. and E.P. Lazear (1992) "Peer Pressure and Partnerships", *Journal of Political Economy 100*, 801-817.

Karniol, R. and D. Shomroni (1999) "What Being Empathic Means: Applying the Transformation Rule Approach to Individual Differences in Predicting the Thoughts and Feelings of Prototypic and Non Prototypic Others", *European Journal of Social Psychology 29*, 147-160.

Kerr, N.L. (1995) "Norms in Social Dilemmas", in *Social Dilemmas: Social Psychological Perspectives*, D.Schroeder (ed.), New York: Pergamon, 31-47.

Knight, G.P, Johnson, L.G., Carlo, G. and N. Eisenberg (1994) "A Multiplicative Model of the Dispositional Antecedents of a Prosocial Behavior: Predicting More of the People More of the Time", *Journal of Personality and Social Psychology 66*, 178-183.

Knudson, K.H. and S. Kagan (1982) "Differential Development of Empathy and Prosocial Behavior", *Journal of Genetic Psychology 140*, 249-251.

Koestner, R., Franz, C. and J. Weinberger (1990) "The Family Origins of Empathic Concern: A 26-year Longitudinal Study", *Journal of Personality and Social Psychology 58*, 709-717.

Kohlberg, L. (1984) "The Psychology of Moral Development", in *Essays on Moral Development*: (vol. 2), L. Kohlberg (ed.), San Francisco: Harper and Row.

Kolm, S.C. (2000) "The Theory of Reciprocity", in *The Economics of Reciprocity, Giving and Altruism*, L.A. Gérard-Varet, S.C. Kolm and J. Mercier-Ythier (eds.), London: Mc Millan Press Ltd, 115-141.

Kolm, S.C. (2001) "Vox Populi, Vox Dei: Endogenous Social Choice and the Rational Original Position", mimeographed.

Komorita, S.S. and C.D. Parks (1995) "Interpersonal Relations: Mixed Motive Interaction", *Annual Review of Psychology 46*, 183-207.

Krebs, D.L. (1991) "Altruism and Egoism: A False Dichotomy?", *Psychological Inquiry 2*, 137-139.

Krebs, D.L. and F. Van Hesteren (1994) "The Development of Altruism: Toward an Integrative Model", *Developmental Review 14*, 103-158.

Kreps J.K. and N.B. Davies (1981) *An Introduction to behavioural Ecology*, Oxford: Blackwell

Lennon, R. and N. Eisenberg (1987) "Emotional Displays Associated with Preschoolers' Prosocial Behavior", *Child Development 58*, 992-1000.

Lepper, M.R. (1983) "Social-control Processes and the Internalization of Social Values: An Attributional Perspective", in *Social Cognition and Social Development: A Socio-cultural Perspective*, E.T. Higgins, D.N. Rubble and W.W. Hartup (eds.), Cambridge: Cambridge University Press, 294-330.

Levin, I. and R. Bekerman-Greenberg (1980) "Moral Judgment and Moral Reasoning in Sharing: A Developmental Analysis", *Genetic Psychological Monographs 101*, 215-230.

Levine, C. (1979) "Stage Acquisition and Stage Use: An Appraisal of Stage Displacement Explanations of Variation in Moral Reasoning", *Human Development 22*, 145-164.

Levine, D.K. (1998) "Modeling Altruism and Spitefulness in Experiments", *Review of Economic Dynamics 1*, 593-622.

Levins R. (1970) "Extinction", in Gerstenhaber M. (ed.) *Some Mathematical Problems in Biology* (Lectures on Mathematics in the Life Sciences, vol. 2), American Mathematical Society, Providence, 75-108.

Lévy-Garboua, L. and C. Montmarquette (1996) "Cognition in Seemingly Riskless Choices and Judgments", *Rationality and Society 8*, 167-185.

Lévy-Garboua, L. and S. Blondel (2002) "On the Rationality of Cognitive Dissonance", in *The Expansion of Economics and Other disciplines: Towards an Inclusive Social Science*, S. Grossbard-Schechtman and C. Clague (eds.), M.E. Sharpe, Inc., 227-238.

Lévy-Garboua, L. and B. Rapoport (2002) "A Theory of Social Norms, Fairness, and Competition", mimeo, TEAM: Université de Paris I.

Lévy-Garboua, L. and C. Montmarquette (2003) "Reported Job Satisfaction: What Does it Mean? ", *Journal of Socioeconomics*, forthcoming.

Lichtenstein, S. and P. Slovic (1971) "Reversals of Preferences between Bids and Choices in Gambling Decisions", *Journal of Experimental Psychology 89*, 46-55.

Lind, E.A. and T.R. Tyler (1988) *The Social Psychology of Procedural Justice*, New York: Plenum.

Lind, E.A., Kanfer, R. and P.C. Earley (1990) "Voice, Control, and Procedural Justice: Instrumental and Non-instrumental Concerns in Fairness Judgments", *Journal of Personality and Social Psychology 59*, 952-959.

Lind, E.A., Kray, L. and L. Thompson (1998) "The Social Construction of Injustice: Fairness Judgments in Response to Own and Others' Unfair Treatment by Authorities", *Organizational Behavior and Human Decision Processes 75*, 1-22.

Lind, E.A., Kray, L. and L. Thompson (2001) "Primacy Effects in Justice Judgments: Testing Predictions from Fairness Heuristic Theory", *Organizational Behavior and Human Decision Processes 85*, 189-210.

Lind, E.A. (2001) "Fairness Heuristic Theory: Justice Judgments as Pivotal Cognitions in Organizational Relations", in *Advances in Organizational Justice*, J. Greenberg and R. Cropanzano (eds.), Stanford, Cal.: Stanford University Press, 56-88.

Litvack-Miller, McDougall, W.D. and R. M. Romney (1997) "The Structure of Empathy During Middle Childhood and its Relationship to Prosocial Behavior", *Genetic, Social, and General Psychology Monographs 123*, 303-324.

Loewenstein, G.F., L. Thompson and M.H. Bazerman (1989) "Social Utility and Decision Making in Interpersonal Contexts", *Journal of Personality and Social Psychology 57*, 426-441.

Lumsden C.J. and E.O. Wilson (1981) *Genes, Mind and Culture*, Cambridge Mass: Harvard University Press.

Macauley, J.R. and L. Berkowitz (Eds.) (1970) *Altruism and Helping Behavior*, New-York: Academic Press.

Manski, C.F. (2000) "Economic Analysis of Social Interactions", *Journal of Economic Perspectives* 14:3, 115-136.

Marks, G. and N. Miller (1987) "Ten Years of Research on the False Consensus Effect: An Empirical and Theoretical Review", *Psychological Bulletin 102*, 72-90.

Masclet, D., Noussair, C., Tucker, S. and M.C. Villeval (2003) "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism", *American Economic Review 93*, 366-380.

Mauss, M. (1954) *The Gift: Forms and Functions of Exchange in Archaic Societies*, London: Cohen and West (first edition in French, 1925).

Maynard-Smith J. (1964) "Group Selection and Kin Selection", *Nature*, London, 201, 1145-7.

Maynard-Smith J. (1982) *Evolution and the Theory of games*, Cambridge: Cambridge University Press.

Maynard-Smith J. (1985) "Game Theory and the Evolution of Cooperation", in Bendall D.S. (ed.) *Evolution From Molecules to Men*, Cambridge: Cambridge University Press.

McFarland D. (1985) *Animal Behavior*, London: Pitman Publishing Limited.

Mead, G.H. (1934) *Mind, Self, and Society*, Chicago: Chicago University Press.

Mellers, B.A. (1982) "Equity Judgments: A Revision of Aristotelian Views", *Journal of Experimental Psychology: General 111*, 242-270.

Mellers, B.A. (1986) " 'Fair' Allocations of Salaries and Taxes", *Journal of Experimental Psychology 12*, 80-91.

Messick, D.M. and M.B. Brewer (1983) "Solving Social Dilemmas: A Review", in *Review of Personality and Social Psychology* (Vol. 4), L.Wheeler and P. Shaver (eds.), Beverly Hills, Ca: Sage, 11-44.

Michener C.D. (1974) *The Social Behavior of the Bees*, Harvard: Belknap Press.

Miller, J.G, Bersoff, D.M. and R.L. Hartwood (1990) "Perceptions of Social Responsibilities in India and in the United States: Moral Imperatives or Personal Decisions?", *Journal of Personality and Social Psychology 58*, 33-47.

Miller, J.G and D.M. Bersoff (1992) "Culture and Moral Judgment: How are Conflicts Between Justice and Interpersonal Responsibilities Resolved?", *Journal of Personality and Social Psychology 62*, 541-554.

Miller, P.A., Eisenberg, N., Fabes, R.A. and R. Shell (1996) "Relations of Moral Reasoning and Vicarious Emotion to Young Children's Prosocial Behavior Toward Peers and Adults", *Developmental Psychology 32*, 210-219.

Miller, P.H., Kessel, F.S. and J.H. Flavell (1970), "Thinking about People Thinking about People Thinking about…: A Study of Social Cognitive Development", *Child Development 41*, 613-623.

Mitchell, G., Tetlock, P.E., Mellers B.A. and L.D. Ordonez (1993) "Judgments of Social Justice: Compromises between Equality and Efficiency", *Journal of Personality and Social Psychology 65*, 629-639.

Mullen, B., Atkins, J.L., Champion, D.S., Edwards, C., Hardy, D., Story, J.E., and M. Vanderklok (1985) "The False Consensus Effect: A Meta-analysis of 155 Hypothesis Tests", *Journal of Experimental Social Psychology 21*, 262-283.

Neuberg, S.L., Cialdini, R.B., Brown, S.L., Luce, C., Sagarin, B.J. and B.P. Lewis (1997) "Does Empathy Lead to Anything More than Superficial Helping? Comment on Batson et al. (1997)", *Journal of Personality and Social Psychology 73*, 510-516.

Nishida T. (1986) "Local traditions and cultural transmissions", in Smuts B.B. and al. (eds.) *Primate Societies*, Chicago: The University of Chicago Press, 462-474.

Oliner, S.P. and P.M. Oliner (1988) *The Altruistic Personality: Rescuers of Jews in Nazi Europe*, New-York: Free Press.

Offerman, T. (2002) "Hurting Hurts More than Helping Helps", *European Economic Review 46*, 1423-1437.

Otten, S. and D. Wentura (2001) "Self-Anchoring and In-group Favoritism: An Individual Profiles Analysis", *Journal of Experimental Social Psychology 37*, 1-8.

Parducci A. (1965) "Category Judgment: A Range-frequency Model", *Psychological Review 72*, 407-418.

Park, B. and C.M. Judd (1990) "Measures and Models of Perceived Group Variability", *Journal of Personality and Social Psychology 59*, 173-191.

Payne, J.W., Bettman, J.R. and E.J. Johnson (1992) "Behavioral Decision Research: A Constructive Processing Perspective", *Annual Review of Psychology 43*, 87-131.

Piaget J. (1973 [1932]) *Le jugement moral chez l'enfant*, Paris: Presses Universitaires de France (English translation (1997) *The Moral Judgment of the Child*, New York: Free Press Paperbacks).

Rabbie, J.M. and M. Horwitz (1969) "Arousal of Ingroup-Outgroup Bias by a Chance Win or Loss", *Journal of Personality and Social Psychology 13*, 269-277.

Rabbie, J.M., Schot, J.C., and L. Visser (1989) "Social Identity Theory: A Conceptual and Empirical Critique from the Perspective of a Behavioural Interaction Model", *European Journal of Social Psychology 19*, 171-202.

Rabin, M. **(**1993) "Incorporating Fairness into Game Theory and Economics", *American Economic Review 83*, 1281-302.

Rawls, J. (1971) *A Theory of Justice*, Cambridge, Mass.: Harvard University Press.

Robson A.J. (2001) "The biological basis of economic behavior", *Journal of Economic Literature 39*, 11-33.

Roch, S.G., Lane, J.A.S., Samuelson, C.P., Allison, S.T. and J.L. Dent (2000) "Cognitive Load and the Equality Heuristic: A Two-stage Model of Resource Overconsumption in Small Groups", *Organizational Behavior and Human Decision Processes 82*, 185-212.

Rohrbaugh, J., McClelland, G. and R. Quinn (1986) "Measuring the Relative Importance of Utilitarian and Egalitarian Values: A Study of Individual Differences about Fair Distribution", in *Judgment and Decision Making: An Interdisciplinary Reader*, H.R. Arkes and K. R. Hammond (eds.), Cambridge: Cambridge University Press, 613-624.

Rosenhan, D.L. (1970) "The natural socialization of altruistic autonomy", in *Altruism and helping behavio*r, J. Macauley and L. Berkowitz (eds.), New York: Academic Press, 251-268.

Ross, L., Greene, D. and P. House (1977) "The 'False Consensus Effect': An Egocentric Bias in Social Perception and Attribution Processes", *Journal of Personality and Social Psychology 13*, 279-301.

Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M. and S. Zamir (1991) "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study", *American Economic Review 81*, 1068-95.

Runciman, W.G. (1966) *Relative Deprivation and Social Justice*, Berkeley: University of California Press.

Rushton, J.P. (1980) *Altruism, Socialization, and Society*, Englewood Cliffs, NJ: Prentice Hall.

Rushton, J.P. (1981) "Television as a Socializer", in *Altruism and Helping Behavior: Social, Personality, and Developmental Perspectives*, J.P. Rushton and R.M. Sorrentino (eds.), Hillsdale, NJ: Erlbaum, 91-107.

Rushton, J.P. (1991) "Is Altruism Innate?", *Psychological Inquiry 2*: 141-143.

Samuelson, C.D. and S.T. Allison (1994) "Cognitive Factors Affecting the Use of Social Decision Heuristics in Resource-sharing Tasks", *Organizational Behavior and Human Decision Processes 58*, 1-27.

Schachter, S. (1951) "Deviation, Rejection, and Communication", *Journal of Abnormal and Social Psychology 46*, 190-207.

Segal, U. and J. Sobel (1999) "Tit-for-tat: Foundations of Preference for Reciprocity in Strategic Settings", Economics discussion paper, University of California at San Diego.

Sibicky, M.E, Schroeder, D.A. and J.F. Dovidio (1995), "Empathy and Helping: Considering the Consequences of Intervention", *Basic and Applied Social Psychology 16*, 435-453.

Silk J.B. (1986) "Social Behavior in Evolutionary Perspective", in Smuts B.B. and al. (eds.) *Primate Societies*, Chicago: The University of Chicago Press, 318-329.

Simner L. (1971) "Newborns' Response to the Cry of Another Infant", *Developmental Psychology 5*, 136-150.

Smith, A. (1759) *The Theory of Moral Sentiments*, reprinted in 1982 from the Oxford University Press edition (1976), Indianapolis: Liberty Classics.

Smith, A. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, New York: Modern Library (1937).

Smith, K.D., Keating, J.P. and E. Stotland (1989) "Altruism Reconsidered: The Effect of Denying Feedback on a Victim's Status to Empathic Witnesses", *Journal of Personality and Social Psychology 57*, 641-650.

Smuts B.B., D.L. Cheney, R.M. Seyfarth, R.W. Wrangham and T.T. Struhsaker (eds.) (1986) *Primate Societies*, Chicago: The University of Chicago Press.

Sober E (1984) *The nature of Selection*, Cambridge, Mass: The MIT Press.

Sober E. and D.S. Wilson (1998) *Unto Others. The Evolution and Psychology of Unselfish Behavior*, Cambridge, Mass.: Harvard University Press.

Sorrentino, R.M. (1991) "Evidence for Altruism: The Lady is Still In Waiting", *Psychological Inquiry 2*, 147-150.

Stark, O. (1995) *Altruism and Beyond: An Economic Analysis of Transfers and Exchanges Within Families and Groups* (Oscar Morgenstern Memorial Lectures), Cambridge Mass.: Cambridge University Press.

Staub, E. (1971) "Helping a Person in Distress: The Influence of Implicit and Explicit "Rules" of Conduct on Children and Adults", *Journal of Personality and Social Psychology 17*, 137-144.

Staub, E. (1974) "Helping a Distressed Person: Social, Personality, and Stimulus Determinants", in *Advances in Experimental Social Psychology* (Vol. 7), L. Berkowitz (ed.), 293-341.

Staub, E. (1981) "Promotive Positive Behavior in Schools, in Other Educational Settings, and in the Home", in *Altruism and Helping Behavior: Social, Personality, and Developmental Perspectives*, J.P. Rushton and R.M. Sorrentino (eds.), Hillsdale, NJ: Erlbaum, 109-133.

Staub, E. (1992) "The Origin of Caring, Helping and Non Aggression: Parental Socialization, the Family System, Schools, and Cultural Influence", in *Embracing the Other: Philosophical, Psychological, and Historical Perspectives on Altruism*, P.M. Oliner, L. Baron, L.A. Blum, D.L. Krebs and M.Z. Smolenska, (eds.), New York: New York University Press, 390-412.

Stinson, L. and W. Ickes (1992) "Empathic Accuracy in the Interactions of Male Friends Versus Male Strrangers", *Journal of Personality and Social Psychology 62*, 787-797.

Tajfel, H., Billig, M., Bundy, R.P. and Flament**,** C. (1971) "Social Categorization and Intergroup Behavior", *European Journal of Social Psychology 1*, 149-178.

Tajfel, H. and J.C. Turner (1986) "The Social Identity Theory of Intergroup Behavior", in S. Worchel and W.G. Austin (eds.), *Psychology of Intergroup Relations*, Chicago: Nelson Hall, 7-24.

Trivers R.L. (1971) "The evolution of reciprocal altruism", *Quarterly Review of Biology 46*, 35-57.

Turiel, E. (1983) *The Development of Social Knowledge: Morality and Convention*, Cambridge: Cambridge University Press.

Turiel, E. (1998) "Prosocial development", in *Handbook of Child Psychology: Socialization, Personality, and Social Development*, N. Eisenberg (ed.), New York: Wiley, 863-932.

Tyler, T.R. and E.A. Lind (1992) "A Relational Model of Authority in Groups", in M. Zanna (ed.), *Advances in Experimental Social Psychology* (Vol. 25, 115-191), San Diego, CA: Academic Press.

Underwood, B. and B. Moore (1982) "Perspective-taking and Altruism", *Psychological Bulletin 91*: 143-173.

Van Boven, L., Dunning, D. and G. Loewenstein (2000) "Egocentric Empathy Gaps between Owners and Buyers", *Journal of Personality and Social Psychology 7*, 66-76.

Van den Bos, K., Lind, E.A., Vermunt, R. and H.A.M. Wilke (1997) "How do I Judge my Outcome when I do not Know the Outcome of Others? The Psychology of the Fair Process Effect", *Journal of Personality and Social Psychology 72*, 1034-1046.

Van den Bos, K., Vermunt, R. and H.A.M. Wilke (1997) "Procedural and Distributive Justice: What is Fair Depends More on What Comes First than on What Comes Next?", *Journal of Personality and Social Psychology 72*, 95-104.

Van den Bos, K., Wilke, H.A.M., and E.A. Lind (1998) "When do we Need Procedural Fairness? The Role of Trust in Authority", *Journal of Personality and Social Psychology 75*, 1449-1458.

Van Dijke, E. and H. Wilke (1995) "Coordination Rules in Asymmetric Social Dilemmas: A Comparison between Public Good Dilemmas and Resource Dilemmas", *Journal of Experimental Social Psychology 31*, 1-27.

Van Dijke, E., Wilke, H., Wilke, M. and L. Metman (1999) "What Information Do We Use in Social Dilemmas? Environmental Uncertainty and the Employment of Coordination Rules", *Journal of Experimental Social Psychology 35*, 109-135.

Veblen, T. (1934) *The Theory of the Leisure Class*, New York: Modern Library.

Walker, L.J., P. Gustafson, and K.H. Hennig (2001) "The Consolidation/Transition Model in Moral Reasoning Development", *Developmental Psychology 37*, 187-197.

Walster, E. and J.A. Piliavin (1972) "Equity and the Innocent Bystander", *Journal of Social Issues 28*, 165-189.

Wilson E.O. (1971) *The Insect Society*, Harvard: Belknap Press.

Wilson E.O. (1975) *Sociobiology: The New Synthesis*, Harvard: Belknap Press.

Wynne-Edwards V.C. (1959) "The control of the population density through social behavior: A hypothesis", *Ibis 101*, 436-441.

Yamagishi, T., Jin, N., and T. Kiyonari (1999) "Bounded Generalized Reciprocity: In-group Boasting and In-group Favoritism", *Advances in Group Processes 16,* 161-197.

Youniss, J. (1980) *Parents and Peers in Social Development: A Sullivan-Piaget Perspective*, Chicago: Chicago University Press.

Zajonc, R.B. (1980) "Feeling and Thinking: Preferences Need No Inferences", *American Psychologist 35*, 151-175.