PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2010) **365**, 303–317
doi:10.1098/rstb.2009.0192

Review

Selfish or selfless? The role of empathy in economics

Alan Kirman[1] **and Miriam Teschl**[2,3,*]

[1]*GREQAM, Université Paul Cézanne, Ecole des Hautes Etudes en Sciences Sociales, et Institut Universitaire de France, Aix-Marseille III, France*
[2]*Robinson College, University of Cambridge, Cambridge, UK*
[3]*Department of Economics, University of Vienna, Austria*

Empathy is a longstanding issue in economics, especially for welfare economics, but one which has faded from the scene in recent years. However, with the rise of neuroeconomics, there is now a renewed interest in this subject. Some economists have even gone so far as to suggest that neuroscientific experiments reveal heterogeneous empathy levels across individuals. If this were the case, this would be in line with economists' usual assumption of stable and given preferences and would greatly facilitate the study of prosocial behaviour with which empathy is often associated. After reviewing some neuroscientific psychological and neuroeconomic evidence on empathy, we will, however, criticize the notion of a given empathy distribution in the population by referring to recent experiments on a public goods game that suggest that, on the contrary, the degree of empathy that individuals exhibit is very much dependent on context and social interaction.

Keywords: empathy; rationality; other-regarding preferences; identity; public goods game; social interaction

1. SETTING THE BACKGROUND OF EMPATHY IN ECONOMICS

Both sympathy and empathy are concepts that have been used in economics at least since the eighteenth century (Fontaine 1997, 2001). Sympathy is generally taken as the concern for the welfare of others, whereas empathy is conceived of as the capacity to put oneself in someone else's shoes and thus to share the sentiments or thoughts of that person. Despite the early interest of Smith (1759 [1976]) and Hume (1740 [2000]), it would not be true to say that both concepts have a particularly developed tradition in economics, especially insofar as empathy is concerned. It is standard to characterize the historical interest in empathy by referring to Smith's view that we can place ourselves in another person's situation by imagination and thus understand what it is like to be the other person in particular circumstances, as explained in his *Theory of moral sentiments*. However, as Fontaine (1997) shows, empathy plays a much more general role in Smith's work and also figures in his *Wealth of nations*, simply because it is conducive to a successful trade between economic agents. Indeed, if one wanted to make an offer, one must do so in a way that appeals to the self-love of the trading partner. For this to happen, the traders must put themselves in the shoes of the other person and see how they would react in those circumstances. Yet, there is little evidence that subsequent economists took up this point about an advantageous change in perspectives (Fontaine 1997).

One might well ask why considerations of empathy disappeared for so long from the economics literature. One answer is that as economic theory developed and became formalized in the twentieth century, almost all of the emphasis was put on the idea of anonymous individuals satisfying specific axioms of rationality and interacting only through the market. In such a view, there was no place for the idea that individuals might want, or need, to put themselves in the place of others. However, with the development of game theory, such an idea became central. Here, the idea is that individuals interact directly and consciously with each other. Indeed, the basis of game theory is that this interaction is strategic. In that framework, unlike the standard economic model, there is what is usually referred to as the 'common knowledge' assumption (Aumann 1976; Binmore 1990), which means that the individuals involved reflect on the actions of the others with whom they interact and know that the others do the same. It is clear, of course, that this is a different idea from the more standard idea of empathy as an affective understanding of the emotions of another. Singer & Fehr (2005) make a distinction between the two ideas and refer to this conscious taking into account of the other as 'mentalizing'. But, if, as authors such as Binmore suggest, we are to build our model of society and its functioning on the basis of the strategic interaction between individuals, then we must keep this aspect of empathy as a central consideration. As Binmore (1994, p. 289) points out, empathy must not be

* Author for correspondence (miriam.teschl@univie.ac.at).

One contribution of 12 to a Theme Issue 'Rationality and emotions'.

considered as 'some auxiliary phenomenon to be mentioned only in passing', but rather as something basic to humanity which can enable us to understand the nature of strategic interactions between individuals. Hence, '*Homo economicus* must be empathetic to some degree' (Binmore 1994, p. 28).[1] This empathy, if it is to condition the way in which individuals anticipate each other's actions and coordinate, must be based on experiences with other people. It is these experiences that allow the empathizing person to better understand the position in which the other person may be. The notion of empathy in this mentalizing sense provides an important reason as to why economists have recently become more interested in this capacity.

Yet, this idea has a long history and it was only as the problem of the coordination of individual activities became less central that it fell out of sight. In fact, this problem goes back at least to Hobbes and was discussed by Hume (1740 [2000]) who was perhaps the first to make explicit reference to the role of mutual knowledge in coordination. In his account of convention in *A treatise of human nature*, Hume argued that a necessary condition for coordinated activity was that agents all know what behaviour to expect from one another. But he does not argue that by so doing they will reach a good outcome. Note here, however, an important distinction with the usual idea of affective empathy. What Binmore holds is that empathy can trigger an identification with another person, without the person in question *ceasing to be able to clearly separate which preferences belong to which person*. This 'empathetic identification' (Binmore 1994, p. 288) is crucial for human societies because without it people would be unable to arrive at and sustain equilibria in the games that are played with others throughout their lives and which Binmore refers to as the 'Game of Life'.

Yet, despite the resurgence of interest in empathy as a tool to gather information about another's preferences, beliefs and intentions, the only persistent use of empathy in economics has been as an instrument of interpersonal comparison of utility with the aim of constructing social preference orderings. Here, it is important to be clear about whose preferences are involved. In an analysis of this social choice literature, two different concepts of empathy have been proposed (Fontaine 2001): partial empathetic identification, which refers to an imagined change of circumstances with another, and complete empathetic identification, which is the imagined change of circumstances together with subjective features. That means that while in the first case, the individual keeps his or her individual preferences[2], in the second case the individual actually adopts the other person's preferences. The latter approach, however, causes problems for the continuity of personal identity (in economics often defined as being the preference ordering) of the empathizing person and renders the construction of a social preference ordering on the basis of individual preferences difficult. Sen (2006) makes this point in relating the question to that of identity. As he says,

> There are two quite different uses of identity, namely an 'epistemic' use, in trying to know what others feel and what they see by placing oneself in the position of others, and an 'ethical' use, in counting them as if they were the same as oneself. The epistemic use is inescapably important, since our knowledge of other people's minds has to be derivative, in one way or another, on our placing ourselves in the position of others.

Partial empathetic identification has its own difficulties, however, and has been shown to lead to Pareto inferior social states with respect to what individuals would really prefer. Moreover, partial empathetic identification has been criticized precisely because it omits the subjective features of individuals in the imagined change of positions. What an empathetic person is interested in should be the feelings of the other and not their own. To overcome these problems, propositions have been made to consider complete empathetic identification without the loss of personal identity. This has been achieved, for example, by assuming not simply the adoption of another's preferences, but some form of deductive understanding of the causal variables that determine these preferences. Thus, not only the desires of the other are taken on board, but also the factors that produced them are taken into account. This is not without its own problems since it implies a deeper identification with the other.

Another source of interest in empathy can be associated with recent developments in behavioural and experimental economics, which have now started to include empathy among a number of other emotions attributed to the economic agent that are considered to foster prosocial behaviour. Bowles & Gintis (2003, pp. 432–433), for example, write:

> *Prosocial emotions* are physiological and psychological reactions that induce agents to engage in cooperative behaviours [...]. Some prosocial emotions, including shame, guilt, empathy, and sensitivity to social sanction, induce agents to undertake constructive social interactions; others, such as the desire to punish norm violators, reduce free riding when the prosocial emotions fail to induce sufficiently cooperative behaviour in some fraction of members of the social group.

If we take this literally, empathy is not clearly defined and is rather taken to be some form of visceral reaction, which makes it difficult to understand whether empathy is really referring to the process of changing perspective and looking at the world from another person's point of view, or is simply taken to be an experience of communal feelings that promotes concern for the other person's welfare. In any case, the inclusion of those emotional aspects of the economic agent has given a new stimulus to the discussion as to whether 'homo economicus' really is the selfish or self-regarding individual who is usually depicted in economics, or whether the nature of homo economicus is fundamentally different, e.g. by having an in-wired other-regarding component, which may have evolutionarily advantageous origins.

Neuroeconomists have now started to look at empathy in more detail (Singer & Fehr 2005; Singer *et al.* 2006). Roughly speaking, neuroeconomics tries to analyse the nature of the activation of the human brain while individuals are carrying out economic

decisions, often in the context of typical economic experiments such as the ultimatum game (Fehr & Camerer 2007; Sanfey 2007; Oullier *et al*. 2008). Neuroeconomics is a recent but growing field; yet the specific research on empathy is only in its infancy. Nevertheless, neuroeconomists consider that the study of empathy will help shed light on two different issues. One is an epistemological question (similar to Binmore's suggestion): understanding the process of how people put themselves in the shoes of others will help to clarify what knowledge people can reasonably have about other people's beliefs, intentions and motives. This refers back to the common knowledge assumption mentioned above. This capacity is certainly important from a self-interested point of view: if one can better predict what another person is going to do based on an empathetic understanding, it will help to make better decisions for oneself. On the other hand, the study of empathy may also be important for motivational reasons: if empathy leads not only to a better understanding of the other person's beliefs, intentions and motivations, but to a feeling of shared experience with the other person's sensations and emotions, this may well undermine purely self-interested choices and instead promote other-regarding behaviour. That is, empathy may be the basis for 'social'[3] preferences and lead to altruistic and other-regarding behaviour. If this were the case, it would again provide a good reason for reconsidering the conception of the economic agent as a purely self-interested or self-regarding individual.

Thus, broadly speaking, we can say that empathy concerns three key areas: (i) game theory and the common knowledge assumption, (ii) welfare economics and interpersonal utility comparisons, and (iii) behavioural economics/neuroeconomics, testing on the one hand how people can come to know about the preferences of others and on the other investigating the idea that other-regarding preferences may or may not lead to non-selfish behaviour.

The 'third' economic area of empathy, behavioural and neuroeconomics, is of course a particularly interesting field of research since it combines economic theory with results from neuroscientific research and, in that sense, promises to make substantial contributions to the other two areas in which empathy enters into economics, i.e. game theory and welfare economics. In addition, some neuroeconomists are eager to show the existence of other-regarding preferences as explained above. Before proceeding it is worth making a cautionary remark. Some economists (see Gul & Pesendorfer 2008) have argued that this field has little or nothing to tell us about the nature of economic decision-making. Furthermore, the existence, relevance and importance of other-regarding preferences are currently quite a contentious domain in economics (Binmore & Shaked in press *a*,*b*). If we assume such preferences do exist, then the next important issue for an economist to clarify is the stability and distribution of other-regarding preferences. The question is, are there any particular 'other-regarding preference types' in the population? Some neuroscientific experiments have been claimed to show that there exists a heterogeneous expression of empathy across experimental subjects (Singer *et al*. 2006). The idea is that each person is endowed with a certain degree of empathy or other-regarding preference. This could therefore be translated into a particular distribution of other-regarding behaviour based on these intrinsic other-regarding preferences, assuming that it is this type of empathy that leads to such behaviour. This is good news for standard economists as they generally assume preferences to be given and stable, and it is on the basis of these that they are then able to construct models that lead to specific behavioural predictions over time. Behavioural and neuroeconomists who are trying to introduce new insights into an existing formalized background of economic theory have therefore started to develop models that incorporate a given distribution of other-regarding preferences (see Fehr & Schmidt 1999). While this is clearly consistent with standard procedures in economics, we will show, using data from a specific economic experiment (Hichri & Kirman 2007), that people reveal anything but stable other-regarding preferences. The claim of some neuroeconomists to be able to identify different personal capacities for mind reading and empathy should therefore be accepted with caution, precisely because there is so little experimental evidence to date. The conclusion that we can draw from our discussion is that individuals are indeed concerned about aspects of outcomes other than their personal advantage, but they do not have consistent attitudes in this respect over time. It is not yet clear what the reasons for such inconsistency are. This raises a number of important questions that need to be clarified at this point, namely whether and how empathetic experiences are translated into other-regarding preferences that are then acted upon and whether empathy can be considered to be a particular trait or simply as a transitory product of particular circumstances and, in particular, of social interactions.

Hence, the broad question that we will try to address in the rest of this article is to what extent cognitive and neurological empathy research will help us to clarify certain conceptual issues in economics such as processes of interpersonal utility comparison or the common knowledge assumption. The more specific question, however, that we will look at is just how far empathy research undermines the standard conception of the economic agent as a self-interested individual if it can be shown that empathy does lead to other-regarding behaviour. For this we shall explore in more detail the still very young neuroeconomic literature on empathy and mind reading, and will analyse how it feeds back into standard economics.

To answer the sort of questions that we have posed so far, we have to understand the social psychological and neuroscientific findings and investigations on empathy. We will not try to give a review of this literature—there are some very comprehensive review papers on this available already (e.g. Decety & Lamm 2006; Singer 2006; de Waal 2008; Singer & Lamm 2009). What we shall do is to read some of this literature through the lenses of an economist keeping in mind, all the time, the conceptual issues raised in the three 'economic' areas of empathy mentioned above. In order to do so, we will try to give a short

presentation as to what cognitive and neuroscientific research has to say on (i) how we can access and understand the minds of others, (ii) whether empathetic experience maintains the distinction between self and other, and (iii) whether these experiences lead to selfish or selfless acts.

In the following section we will look more specifically into the empathy literature from neuroeconomics. In §3, we present a critique of the specific neuroeconomic literature on empathy and mind reading, which concerns the relation of empathy to other-regarding preferences. We do so by briefly discussing the issues raised in a debate between two pairs of economists who have very conflicting views about the value of experimental and behavioural economics. On the one hand, there are two, Binmore & Shaked (in press *a*,*b*), who are very reluctant to replace the standard conception of the economic agent as a self-regarding individual, and who claim that this conception is sufficient to predict much of observed behaviour, even that in apparent contradiction with standard rationality issues. On the other hand, there are two behavioural and neuroeconomists, Fehr & Schmidt (1999, 2009), who have presented a model of an economic agent with particular other-regarding preferences, namely those exhibiting 'inequity aversion'. In this model, the agent is supposed to care about the difference between others' monetary outcomes and his own to the extent that any difference between the agent and someone else affects the agent negatively. One of the key issues of their debate is Fehr and Schmidt's proposed distribution of intrinsic other-regarding preferences that Binmore and Shaked basically criticize as not being derived according to well-defined scientific standards. Should such a distribution exist, then according to Binmore and Shaked, it should be properly specified in order to make valid predictions. We take this point up to show the importance that some economists attribute to the stability of a given distribution of other-regarding preference types. Once again this allows one to fall back on the standard idea of fixed and immutable preferences. It is in response to this that we present the results of Hichri and Kirman's experiment which basically shows that while there is some stability in terms of the degree of other-regarding behaviour at an average level, no such stability is observed at the individual level. In our last section, we present some ideas as to what may trigger this observed inconsistency and conclude by emphasizing the importance of continuing to look more closely into the decision processes that may involve mind-reading and empathetic components. The time is surely ripe for economists to study empathy, in all its different forms, from the understanding of another person's perspective to the sharing of emotional feelings with others.

2. AN INTERPRETATION FROM AN ECONOMICS POINT OF VIEW OF NEUROSCIENTIFIC AND SOCIAL PSYCHOLOGICAL FINDINGS ON EMPATHY

In this section, we give an overview of some psychological and neuroscientific findings that are of direct relevance to the three empathy areas in economics outlined above.

Social neuroscience generally makes the distinction that we have mentioned between two types of empathic processes: a cognitive version of empathy and an affective version of empathy (Singer 2006). The former is often called the 'theory of mind' or mentalizing and refers to the capacity to represent other people's intentions and beliefs, which is the idea that game theorists often have in mind. The affective aspect of empathy (what is then referred to as 'empathy' proper) refers to the capacity to share the feelings of others. This distinction can hardly be clear-cut because representing another person's beliefs and intentions may also lead to the affective sharing of a particular emotion or sensation, and experiencing the same feelings may lead to a better understanding of that person's perspective.

(a) *How can we access and understand the minds of others?*

One of the reasons for the interest in the mentalizing type of empathy has been the discovery of so-called mirror neurons (di Pellegrino *et al*. 1992; Gallese *et al*. 1996). Up to now mirror neurons have been described in the premotor cortex and the inferior parietal lobe of monkeys (Rizzolatti 2005). The evidence for some similar system, often referred to as the mirror system, in the human brain is only indirect as there is no single-neuron study for healthy brains,[4] although functional imaging studies have revealed similar neural activation during action observation (Rizzolatti & Craighero 2004). To be more precise, although some consider that it is not appropriate to say that mirror neurons have a particular functional role (Rizzolatti 2005), there is an argument for the whole brain having a mirror modus operandi (see Wicker *et al*. 2003) and this may play a role in action understanding, intention understanding, imitation and also the emotional aspect of empathy.

This means that with the help of mirror neurons, motor acts executed by other people and, in particular, other people's facial expressions, can be understood and corresponding emotional experiences felt (e.g. of disgust; Wicker *et al*. 2003). For example, experiments have been conducted in which monkeys were tested on grasping behaviour in two conditions (Fogassi *et al*. 2005): one involved the monkey grasping for a piece of food located in front of it and bringing it to its mouth (grasping-to-eat condition); in the second condition, the monkey grasped for an object in front of it and placed it in a container (grasping-to-place). It was observed that the same neurons were fired in the observer's brain during the grasping act, but not during the subsequent motor act, i.e. bringing the food to the mouth and eating act or the placing act. Indeed, what this experiment showed is that the respective discharge intensity of the neurons depended on the goal of the action. It is often found in this kind of study that the motor act of one action is influenced by the subsequent action, which also means that one knows before one does something what is going to happen. Even more, there is a potential in the brain

that can be measured prior to an action which can be translated as 'the brain knows before the mind does'. The interesting aspect of this experiment is that even if these monkeys observed another monkey performing the same kind of actions, again, neurons were activated differently depending on the grasping-to-eat and the grasping-to-place condition. This indicates that they not only understand the goal of the observed motor act, but also discriminate between identical motor acts according to the action in which these acts are embedded. It is as if they could 'read' the intention of the monkey they were observing. This capacity to understand intentionality has sometimes been claimed to be unique to humans.

The explanation for this sort of anticipation in humans is that the brain stores internal representations of dynamic information, which can then be used in order to anticipate future movements and thus to infer intentions from other people by observing their movements (Blakemore & Decety 2001; Iacoboni 2009). That is, a person may have stored the consequences of his or her own actions based on particular intentions, and that may be used to understand the intentions of other people. It is therefore argued that the brain is a 'powerful simulating machine, designed to detect biological motion in order to extract intentions from the motion and to predict the future actions of other animate beings' (Blakemore & Decety 2001, p. 566). It is interesting to note, in this context, that a part of the mirror system is the superior temporal sulcus which is not premotor at all but fires when biological motion is detected. This creates 'shared representations' (p. 566) of motions that may be used for subsequent understanding of somebody else's actions and intentions.

The step to mimicry is not difficult, at least for humans who are considered to be the 'best' imitators, for as the philosopher Hoffer (1955, p. 21) said,

> When people are free to do as they please they usually imitate each other.

Indeed, it is widely recognized that humans seem to have a strong tendency to imitate and align their behaviour with other people during social interactions (see Lieberman 2007). Indeed, imitation facilitates social interaction and influences people's liking of others (for an overview, see Iacoboni 2009). It has also been shown that imitative behaviour fosters empathy and that empathic persons are more inclined to unconscious mimicry (Chartrand & Bargh 1999; see also de Waal (2008) for an overview). There are even functional magnetic resonance imaging (fMRI) studies that neurally connect motor mimicry, such as contagious yawning, with empathic modelling (Platek *et al.* 2005). This all suggests that personal experience and social closeness, together with motor actions that lead to motor cognition (see Jeannerod 2006; Oullier & Basso 2010), are important factors that foster the understanding of other people's actions and intentions. Through imitation and mimicry people learn to feel what other people feel. However, there are still some critics of this research who call for caution. 'The observation of similar neural activations during the firsthand vs. the vicarious experience of various sensations and emotions (e.g. disgust, taste, pain) raises the question whether these activations can indeed be interpreted as shared representations' (Singer & Lamm 2009, p. 87). This is because none of the current neuroimaging methods can measure the activity of single neurons or neural networks directly. Usually, one compares two fMRI activation maps with overlapping clusters—but the activation could potentially still result from differing neural activity. Extensive research on these issues is currently being undertaken.[5]

(b) *The self–other distinction*

The neuroscientific literature insists on a clear separation between self and other, and this is, of course, crucial for an understanding of empathy. However, it is not entirely clear that such a distinction can always be made. For such a distinction to be made one might argue that one needs a clear definition of what constitutes self. This has been a subject of debate in philosophy at least since the classical Greeks. Sorabji (2006) gives a detailed account of the role of self in Western classical and modern philosophy as well as in the Hindu, Buddhist and Islamic traditions. Indeed, the denial of self stemming from antiquity was embraced by Hume and Nietzsche, and later by many analytical philosophers such as Wittgenstein, Anscombe, Kenny and Dennett. But, to avoid what could become an endless detour here, it may be just as well to adopt Sorabji's (2006, p. 20) more pragmatic approach when he says,

> One reason why the notion of self comes in is that humans and animals could not cope with the world at all unless they saw things in terms of 'I'.

Again, the subject of this article, empathy, has no meaning if we cannot make a meaningful distinction between self and other. In particular, the assumption that such a distinction can be made lies at the heart of the literature on neuroscientific experiments on empathy.

Consider Singer *et al.*'s (2004) experiment. They recruited couples and both partners were brought into the same scanner environment. Brain activity was then measured in the female partner while painful stimulation was applied either to her own or to her partner's right hand. The results suggested that some parts, but not all, of the pain-related brain activation were activated when the female partner observed pain being inflicted on her partner, in contrast to the neural activity when she experienced pain herself. In particular, knowing that the partner experienced pain activated the same affective pain circuits, but not the primary somatosensory cortex. In addition, the higher the activation of the affective pain circuits, the higher the individuals scored in social psychological empathy questionnaires that were distributed to assess the individual level of empathy. The result was that people experienced different levels of empathy and this led the authors to the conclusion that not everybody is endowed with the same level of empathetic experience. But most importantly, the

level of self-reported empathy was strongly correlated with increased brain activity.

Now, it has been argued that 'close relationships' (such as couples) not only lead to behavioural similarities, but also have cognitive consequences (Aron *et al*. 1991). In a close relationship, it has been observed that a person acts as if some or all aspects of his or her partner are partially the person's own. This concerns, for example, the distribution of resources, where it has been found in experiments with couples that little difference is usually made between personal and partner's resources and both are considered beneficial to the person. The vicarious sharing of particular characteristics of the other person has also been observed. Hence, self and other apparently overlap, they become one, and this 'oneness' can be measured in several different ways. One measure is, for example, the Inclusion of Other in Self (IOS) scale, which asks subjects to select from among seven different diagrams of overlapping circles, representing total overlap to no overlap, to describe their relationship with a particular person. It is clear that if there is such an experience of oneness, then empathetic experiences may trigger responses, many of which are neuromotor in nature, that are not strictly speaking directed towards the other, but towards oneself, which makes them fundamentally self-interested.[6]

Cialdini *et al*. (1997) conducted a number of experiments to show that any helping behaviour which follows empathetic experiences can be attributed to the experience of perceived oneness. Subjects were asked to imagine the situation that a stranger, an acquaintance, a good friend or a family member (the four experimental conditions indicating different degrees of closeness) was evicted from their apartment. They were then asked to indicate how much help they would be willing to provide (from nothing, to giving an apartment guide, driving the person through town to visit apartments, inviting the person to stay for few nights in one's own apartment, etc.). They had also to indicate the level of empathetic concern and oneness (measured with the IOS scale among others) the respective situation caused. By empathetic concern, the study meant the emotional reaction characterized by feelings such as compassion, tenderness, soft-heartedness and sympathy, brought about by the act of perspective taking, that is the subject taking the point of view of the evicted person. This concept followed Batson (1991) and colleagues, who proposed the alternative theory to the oneness hypothesis (the 'empathy-altruism' hypothesis), which presupposes a clear distinction between self and other, and stipulates that (altruistic) helping behaviour is caused by the prior experience of empathic concern for the other. Cialdini *et al*.'s (1997) results are that helping increases with both empathetic concern and experience of oneness, but empathetic concern was no longer predictive of helping behaviour once the influence of oneness had been extracted. Hence for Cialdini *et al*., empathic concern is only a signal of unity with another and any helping is thus not an act of promoting anybody else's welfare, but one's own.

Obviously, Batson *et al*. (1997) contested this and ran their own experiments in which they induced in people a low-empathic state and a high-empathic state with respect to the story of Katie, a young student in her final year who had recently lost her parents in an accident and who now had to look after her two younger siblings. In the low-empathic state, subjects were told to be as objective as possible in the assessment of Katie's situation and not to get caught trying to imagine what the person had been through. In the high-empathic state on the other hand, subjects were encouraged to try to imagine how Katie may have felt. As part of the experiment, subjects were also given the possibility of helping Katie cope with her difficult situation (helping here meant participating in sending letters to raise money for Katie and her siblings). Subjects also had to fill in a number of questionnaires, among them the IOS scale to measure any self–other merging. It turned out that the subjects in the high-empathic condition did help more than those in the low-empathic condition. Hence, perspective-taking manipulation led to increased helping. Moreover, Batson *et al*.'s hypothesis was that even if the effect of self–other merging was taken into account, a direct effect of empathy on helping would remain. And indeed, while there was a slightly higher IOS score in the high-empathy condition, the self–other merging could not account for the empathy-helping relationship. What Batson *et al*. seemed to show with this experiment is that IOS may be a valuable measure of self–other merging in close relationships, but that in the case in question, subjects did not necessarily need to 'merge' with Katie in order to feel some degree of care for her. Again, the thrust of their argument is that perspective manipulation can lead to higher empathy and this is associated with increased helping behaviour.

Decety & Lamm (2006) give evidence for the self–other distinction at a neurological level which came from experiments that tested the effect of perspective taking: either the subject had to imagine how it is for another person to be in everyday-life situations that elicit, for example, painful experiences, or how it is for themselves to be in such a person's shoes. The results show that the right inferior parietal cortex, at the junction with the posterior temporal cortex (temporoparietal junction), is specifically involved when participants imagine how another person would feel, but not when they imagine these situations for themselves (Decety & Lamm 2006, p. 1155). With respect to Singer *et al*.'s (2004) experiment described above, they comment that, while there are similarities in the neural networks when one person observes another person involved in receiving pain, it is also the case that not exactly the same neural networks are activated in the observation case and in the real experience case, which excludes a complete overlap of self and other.

(c) *Selfish or selfless?*

This distinction between the 'imagine-self' and 'imagine-other' perspective is important since it leads to different behavioural consequences. If a person observes another person in difficulties and experiences

personal distress, they may try to escape helping this person in order to alleviate their own negative emotional state. If this person experiences empathetic concern, he or she would engage in actions that try to soothe the distress of the other person. Given that we often observe that people do help, this would seem to confirm the idea that people are aware of the difference between self and other.

Decety & Lamm (2006) hold, rather 'romantically', that for the experience of empathy, individuals need to be able to disentangle their own feelings form those of others. Empathy's 'goal' is not to diminish personal distress but to soothe the distress of another person through helping acts (p. 1154). Maintaining a self–other distinction is also a requisite for self-awareness and the sense of being an agent. While self-awareness is important in order to be able to make inferences about another's mental states, the sense of having a personal identity, as an agent (see Jeannerod 2006), is crucial to the idea that one is able to have a selfless regard for the other.

Also related to the agent view is the fact that cognitive and neuroscientific models of empathy suggest that observing or imaging another person in a particular emotional state automatically triggers a representation of that state in the observer (e.g. Preston & de Waal 2002). This is referred to as a bottom-up process. Some researchers, however, point out that 'automatic' means that this process does not require conscious or effortful processing, but this does not mean that it cannot be inhibited and controlled. This is reflected by the fact that deficient empathy is possible—people can selectively disengage from empathy. This empathy regulation is referred to as top-down processes. One can, for example, selectively focus on specific sensory cues (such as facial expressions) that convey the emotional state of some person and that would trigger unpleasant or distressing emotional reactions in the observer. This evokes the early work of Darwin (1872) based on that of Clive Bell, on the expression of emotions and the selective interpretation of that expression by others. Distraction is another mechanism for controlling empathetic reactions (see Decety & Lamm 2006 for an overview). Indeed, empathy regulation may become an important factor if empathy becomes too costly in everyday life. There is therefore a limit as to how much empathy a person can 'bear'. This, however, suggests that people are capable of reducing or even augmenting empathetic experiences, which may trigger particular behavioural responses—such as other-regarding or prosocial behaviour. And yet, while the close link between empathy and prosocial behaviour is repeatedly stressed, the exact mechanism governing the passage from empathy to prosocial behaviour is not clear at all. In fact, Singer & Lamm (2009, p. 84) admit that 'a clearcut empirical demonstration of a link between empathy and prosocial behaviour is still missing'.

Nevertheless, it should be clear from this discussion that social and cognitive psychology as well as neuroscience analyse many issues that seem to be particularly relevant to economic decision-making and economic concepts in general. Here, we will not try to make any direct connections between, for example, mind-reading capacities and the game theoretic assumption of common knowledge. This would be too rudimentary and needs to be the focus of more intensive investigation. What we shall do, however, in the next sections is to concentrate on certain neuroeconomic experiments and the conclusions that have been drawn from them, and analyse the extent to which they have already, rightly or wrongly, influenced the conception and representation of the economic agent.

3. THE NEUROECONOMICS OF MIND READING AND EMPATHY

The title of this section is taken from that of an article by Singer & Fehr (2005). This is one of the rather few articles to date on mind reading and empathy in economics, but it describes enthusiastically what one may expect from such research. First, we will summarize some of the limited number of experiments on empathy.

Singer *et al.* (2006) conducted an experiment to test empathy-related responses to observing pain induced in people with whom subjects previously interacted. More specifically, subjects played a sequential prisoners' dilemma game in which they could decide how much money to give to another player. The other player could then respond, either fairly or unfairly, by returning a low or high amount of money.[7] The purpose of this game was to make subjects like the fair players and dislike the unfair players. In a second part of the experiment, they used fMRI to investigate whether the liking or disliking acquired during the preceding game modulated empathic responses for pain. All subjects of both sexes exhibited empathy-related activation in pain-related brain areas (fronto-insular and anterior cingulate cortices) towards fair players. However, only women had real empathy for unfair players. Men on, the other hand, had significantly reduced empathy-related activation, but experienced increased activation in reward-related areas, correlated with an expressed desire for revenge. These results (at least for men) are consistent with other results that show that people derive satisfaction from norm violations. For example, de Quervain *et al.* (2004) showed that if subjects had the possibility to effectively punish when they heard about a defector's abuse in trust, the dorsal striatum was activated. This part of the brain has been shown to be involved in the processing of rewards that accrue as a result of goal-directed actions. Additionally, subjects with stronger activations in the dorsal striatum were willing to incur greater costs in order to punish. Singer *et al.* (2006) take their results as providing the neural basis for theories of what they refer to as 'social preferences' but which we have called 'other-regarding' preferences,[8] which suggest that people value others' gains positively if they were perceived as being fair, but negatively if they are considered to be unfair. People thus value 'fairness' in addition to their own personal advantage. Moreover, fair behaviour of another person increases empathic experiences of the other person who observes this behaviour, while selfish behaviour reduces it.

Knoch & Fehr (2007) let subjects play an ultimatum game. In the ultimatum game, a proposer has to make an offer of an assigned amount of money to a responder, who has the possibility either to accept the proposed offer or to reject it, in which case both get nothing. A Nash equilibrium of such a game is a choice of strategies by both players such that given the strategy of the other neither has an incentive to deviate. If the total amount in play is €10, for example, one such equilibrium would be for the proposer to propose €10 for himself and for the responder to refuse any offer that gave a positive amount to the proposer. This equilibrium does not satisfy a more reasonable condition that of 'sub-game perfection'. Here, the idea is that the proposer works backwards from the last step, and at each step sees what is the best strategy for the other. In the last step, the responder should accept any positive amount that is proposed since something is better than nothing. Since he understands this, the proposer should propose the minimum positive amount for the responder and keep the rest. It is worth explaining this because it captures the essence of the mentalizing approach, the proposer anticipating what the responder will do when faced with each possible proposition. Now, the question is, do people actually play like this?

In stark contrast to the theory, what has been shown in many different experimental settings of the ultimatum game (see Oosterbeek *et al*. 2004) is that a majority of proposers offer between 30 and 50 per cent of their assigned money and that a vast majority of responders reject offers that are below 20 per cent of the assigned money. These results have also been interpreted as being a sign for a concern of fairness or equity. In Knoch & Fehr's (2007) experiment, one of the hypotheses that they examined was that reduced self-control has an effect on accepting unfair offers. Accepting such offers is considered here to be consistent with more immediate satisfaction of self-interest. It has been observed that patients with lesions, particularly in the right prefrontal cortex, are less resistant to the satisfaction of self-interest. To test this hypothesis, Knoch and Fehr relied on individuals who were, in fact, healthy, but they induced a 'virtual lesion', which meant that they applied a low-frequency repetitive transcranial magnetic stimulation over the course of several minutes, which allowed a transient disruption of cortical functions.

Results showed that subjects who received right prefrontal transcranial magnetic stimulation were less able to resist the economic temptation to accept unfair offers, despite the fact that in an evaluation after the experiment, they considered low offers as being unfair. Knoch & Fehr (2007, p. 129) add that their '[...] findings are also congruent with the observation of empathy deficits in patients with predominantly right frontal lesions, as an inhibitory component is required to regulate and tone down the prepotent self-perspective to allow the perception and evaluation of others' perspectives'. Hence, this shows that reduced self-control has similar implications to those of empathy impairment. This again might seem to suggest that empathy leads to more other-regarding behaviour. However, a less ambitious, alternative explanation is that the dichotomy between emotions and rationality no longer holds and that the complex neural activation during these decisions is interfered with by the lesions.

These results would seem to lend support to Singer & Fehr (2005) when they advocated the idea that neuroscientific research on mentalizing and empathizing may help to explain how individuals assess other players' preferences. Given the evidence from empathy research so far (and again this was reinforced by Singer *et al*.'s (2006) results explained above), they presented the following 'testable' hypothesis—but have not tested it yet (Singer & Fehr 2005, p. 343): first, people with stronger capacities to empathize are better predictors of other's motives and actions. Second, people who exhibit more affective concern are more likely to display altruistic behaviour. This clearly means that more empathetic people should be more aware of what others want and should also be more other-regarding.

Singer & Fehr (2005) also referred to the fact that economists usually attribute a common prior distribution of types in the population (e.g. of other-regarding preference types) in games with incomplete information, but emphasized that this assumption lacks empirical foundations. But 'neuroeconomic research (in mind reading and empathy) may help us to understand what is going on in this black box' (p. 344). And if empathy is related to other-regarding preferences as suggested above, then we may also learn something from observed behaviour about the distribution of other-regarding preferences in a given population. Given the heterogeneity of empathetic reactions as shown by Singer *et al*. (2004), this may well be taken to reveal the existence of a particular distribution of other-regarding preferences. And, as Fehr & Schmidt (in press) put it, it is of prime importance for the empirical knowledge about other-regarding preferences 'to have a parsimonious empirical characterization of the distribution of social preference types'.

4. STABILITY OF TYPES

This distribution of other-regarding preferences has become an issue in the economics literature lately. Fehr, together with Schmidt, has proposed a particular type of other-regarding preference called inequity aversion. Deviations from equality may be in two directions, either the person receives less than somebody else (disadvantageous inequality), or the person receives more than somebody else (advantageous inequality).[9] In both cases, the person is supposed, according to this approach, to suffer a loss. The inequalities are weighted, respectively, by, what Binmore & Shaked (in press *a*,*b*) call in a critique of Fehr and Schmidt's paper, an 'envy parameter' α and a 'discomfort parameter' β. These parameters measure the extent to which inequity aversion weighs on the preferences of each person. These two parameters summarize an individual's attitude towards inequality. Fehr & Schmidt (in press) assume that individuals do not know exactly what other people's value of these parameters are, but that the joint

distribution of α_i and β_i in the population is common knowledge for all players. The polemic that has arisen is about the distribution of these parameters that Fehr and Schmidt claim to deduce from the experimental evidence. Indeed, as Binmore & Shaked (in press *a*,*b*) rightly assert, the proposed distribution (Table III of Fehr and Schmidt's paper) is far from their aim of a 'parsimonious empirical characterization of social preference types'. They propose a simple discrete distribution of α_i and β_i, a distribution chosen because it is consistent with the substantial experimental evidence they have on the ultimatum game. The same distribution of the envy and discomfort parameters is used by Fehr and Schmidt in order to explain other experiments in their paper, among them the public good game without and with punishment (see again Fehr & Schmidt (1999) for references). But as Binmore and Shaked show, if Fehr and Schmidt had used slightly different parameters, their theory of inequity aversion would no longer be able to predict observed behaviour in the various games they consider in their original 1999 paper. The second critique Binmore and Shaked make is that in later follow-up papers, Fehr and co-authors (Fehr & Schmidt 2004; Fehr *et al*. 2007, 2008) do not keep the same distribution, but use what they call a 'simplified' distribution of their 1999 paper because it makes calculation less tedious. Indeed, they anticipate that 'a complete characterization of the distribution of different social preference types in the population may introduce so much complexity at the individual level that models that attempt to capture this complexity may become analytically intractable. For this reason, a simple model such as the theory of inequity aversion may still be useful, even though there is evidence that it does not provide a full description of other-regarding preferences' (Fehr & Schmidt in press).

The question of whether there is a fixed distribution of degrees of other-regarding behaviour in the population is important, and the exchange of opinion between Fehr and Schmidt and Binmore and Shaked is interesting for at least two reasons. Those who claim that neuroeconomic research, and in particular research on mentalizing and empathizing, will be able to shed light into the 'black box' of assumed distributions of types are relatively unconcerned about making quantitative estimates of the distribution of the respective preferences appropriately, that is econometrically. Yet, as Binmore & Shaked (in press *a*,*b*) argue, to find an appropriate and given distribution of types is considered to be crucial for the economic discipline. This view is also reflected in the fact that economists usually assume given and stable preferences. The common understanding among economists is that 'for preferences to have explanatory power they must be sufficiently persistent to explain behaviours over time and across situations' (Bowles 1998, p. 79). Hence, the ideal situation would, from this perspective, be that research on empathy reveals a given empathy distribution in a population, which can then be correlated with stable other-regarding preferences that serve as the basis of economic accurate predictions of behaviour. All of this, of course, rests on the fundamental assumption in economics that behaviour in general and choices in particular can be assimilated to fixed and immutable preferences. Choices and preferences are, in this view, synonymous, and this poses problems because the consistency of choice, and particularly intertemporal choice, that is imposed by the usual axioms is frequently violated in experiments. It would, of course, be very convenient if individuals had a certain fixed level of empathy which was independent of the context in which they found themselves. Indeed, if this were the case, one could design experiments that would permit each individual to reveal their empathetic predisposition.

However, the idea that any such stable distribution exists is likely, as we will show, to be 'wishful thinking'. In what follows, we will use the evidence from a series of public goods experiments run by Hichri & Kirman (2007). The basic idea of public goods experiments is that each individual is in a group and has an initial allotment of money. Each person can split this money into a private share and into the contribution to a public good. Once individuals have made their contributions, the total production of the public good which is consumed by all the individuals and the payoffs to each individual are determined. In some experiments, this game is repeated for several periods.[10] The payoff to the participants of their choices, depends on and varies with the experimental design, but in most experiments it is taken to be linear (Andreoni 1995). This linear case gives rise to a corner solution. This means that, assuming that it is common knowledge that players are rational payoff maximizers, such a function gives a Nash equilibrium in which every player contributes nothing for the one shot as well as the finite repeated game. On the other hand, at the social optimum everybody contributes all of their wealth. Nevertheless, experimental studies show that, at least in the early stages of such games, there is a large fraction of people who over-contribute in a public goods game with respect to the Nash equilibrium.

Hichri & Kirman (2007) were particularly interested in finding a situation in which, even at the Nash equilibrium, total contributions can be positive. For this, they conceived a theoretical model with a concave payoff from the contributions to the public good instead of a linear payoff. In other words, the return from a contribution to the public good diminishes as the total contributions rise. The result of this is that the Nash equilibrium is any situation in which the total contributions of the individuals adds up to a certain positive amount. Thus, even at the non-cooperative solution one can distinguish between the levels of payment of different individuals. Indeed, since many experiments focused on a linear payoff function which implied that the Nash equilibrium is for all players to contribute zero, little attention has been paid to individual differences in payments. With the model Hichri and Kirman propose, the Nash equilibrium implies positive contributions and one can see whether at the Nash equilibrium it may be the case that some individuals systematically contribute more than others. The idea is that if individuals can be distinguished in this way, one may be able to deduce that some are more 'socially

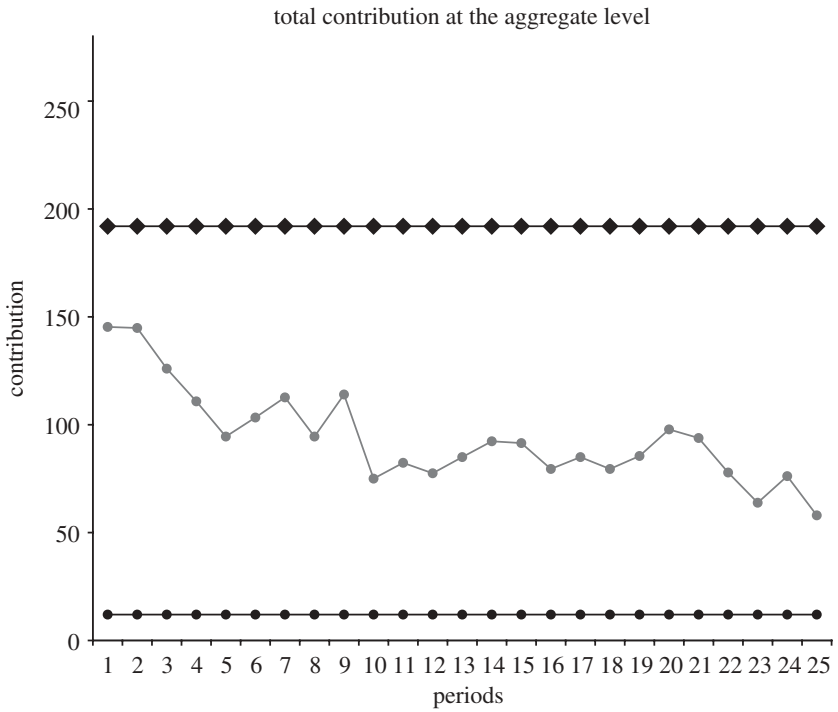

Figure 1. Total contributions in the public goods game. Adapted from Hichri & Kirman (2007). Grey circled line, contribution; filled diamond, collective optimum and black circled line, Nash equilibrium.

conscious' than others, they empathize more or, to use Fehr & Schmidt's (1999) term, they are more 'inequality averse'.

The Nash equilibrium and the social optimum corresponding to this payoff structure are not trivial solutions but are in the interior of the set of the possible choices. With a concave payoff function, the respective Nash equilibrium can easily be calculated. Hence, it can be said that in this game there is a unique Nash equilibrium in the sense that for any Nash equilibrium the contribution of the group, as a whole, is the same. However, a contribution to that equilibrium can be obtained by several combinations of individual contributions. In that sense, we could also say that many Nash equilibria are possible. Since there are many Nash equilibria for the one-shot game, one might ask precisely what constitutes an equilibrium for the repeated game. The answer is that a Nash equilibrium for the game repeated n times will be a sequence of n equilibria for the one-shot game. This means that, in such an equilibrium, the same player may make very different contributions at each step. The only requirement is that the total contribution of the group should remain constant. This means that the following sort of arrangement could be an equilibrium. In the first period, one player contributes a lot and the others very little, and then in each subsequent period another takes over the role of leader and makes the large contribution. How, or why, a group would coordinate on such a solution is a different question. Much simpler, of course, would be a situation in which the different players converge to constant contributions but which are different from each other. In this case, we can characterize people's empathy by their ultimate generosity in the Nash equilibrium. That solution would be consistent with the debate between Binmore and Shaked and Fehr and Schmidt outlined above, but it is not what Hichri and Kirman have found.

In most public goods experiments, total contributions decline towards zero but do not reach that level. In Hichri and Kirman's experiments, total contributions do decline but towards the positive level associated with the Nash equilibrium as is shown in figure 1.

The question then remains, is it true that certain individuals contribute more than others even when they may have 'learned to play Nash'?[11] If this is so then those that pay more can be characterized as having more empathy than those who contribute less. This would correspond to the idea that each individual has a certain level of empathy that is essentially stable. However, the experiments in question do not show this. Looking at the payments of the different groups who make up the total, emphasizing that in the experiments the groups have no interaction with the other groups, one can observe that different groups organize themselves so that different totals are obtained, as can be seen in figure 2.

Hence, this does not suggest that certain groups must contain more empathetic or other-regarding members. This idea that this might be so is undermined by the fact that the contributions are far from monotone and that as time progresses groups' contributions fluctuate considerably. But the most striking observation comes from the contributions of the individuals themselves. The contributions of each of the individuals in several groups do not converge to a particular level, rather individuals modify their payments as others do so, as can be seen in figure 3. Yet, the remarkable aspect is that they seem somehow to coordinate their contributions on a solution that is

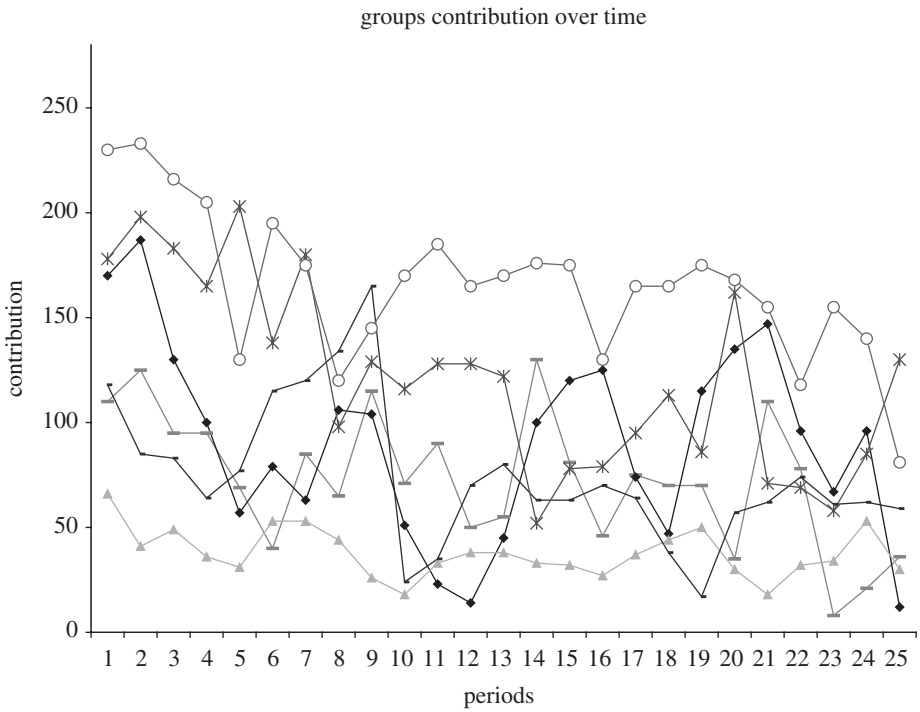

Figure 2. The contributions of six different groups in the same treatment. Adapted from Hichri & Kirman (2007). Lines with dash, group 1; diagonals, group 2; triangles, group 3; solid line, group 4; crosses, group 5; and circles, group 6.

not far from a Nash equilibrium according to standard rationality.

Here, it can clearly be seen that individuals offer different amounts at different moments and, with few exceptions, cannot be classified as low or high contributors reflecting their intrinsic empathy. This does not mean that empathy does not play a role in determining the generosity of the individuals but, instead of having a constant level of empathy, as is suggested by Singer & Fehr (2005), the level of empathy may be context dependent and influenced by the interaction with others.

5. CONCLUSION

In this paper, we have argued that empathy is actually 'pervasive' in economic theory, and understanding empathetic and mind-reading processes better is of crucial importance both from an empirical and a theoretical point of view in economics. Empathy is, of course, a subject that has interested many disciplines, such as philosophy and psychology, and it has recently attracted a great deal of attention in the neurosciences. The research taken up is neuroeconomics, a rapidly growing area of research that uses neuroscientific insights to elucidate the theory of economic decision-making. We have discussed this very recent literature on empathy and mind reading in some detail and have shown that one of the main conclusions claimed by some economists for these studies is the existence of other-regarding preferences. Such preferences are said to be responsible for the altruistic or prosocial behaviour of the economic agent. Positing the existence of these preferences has been considered quite a radical departure from the description of the economic agent in standard economic theory. There they are taken to be, in the 'worst' case, a purely selfish or, in the best case, self-regarding individuals. In the latter case, agents may care about the betterment of 'others' as well as themselves but their consumption or state is simply another argument of their own preferences. Other-regarding preferences are sometimes presented not simply as being concerned with the welfare of specific others, but may incorporate more general concerns, such as inequity aversion (Fehr & Schmidt 1999). However, it should be noted that, in a sense, this idea is at odds with empathy as generally conceived. For, in that view, one is not concerned with the situation of another individual but the state of the population with respect to oneself. Can this really be assimilated with the original idea of empathy?

Indeed, our argument is that even if there is such a component governing people's choices and, as Fehr and Singer in a later contribution argue, the level of such inequity aversion is intrinsic to each individual, there is still a problem. For this would mean that one would observe a stable distribution of inequity aversion across players. Each person would behave systematically in the same way.

In some experiments we referred to earlier (Singer *et al*. 2004, 2006), individual empathic character traits have been assessed by empathy questionnaires after having been scanned, and this revealed a positive correlation between the self-reported empathy level and actually measured brain activation. This seems to suggest that there is indeed a particular disposition or trait in individuals to experience empathy. However, if people watch distressed people on the news, they may not experience any empathy, while they would do so if they saw people in their local community suffering in the same way. This therefore suggests

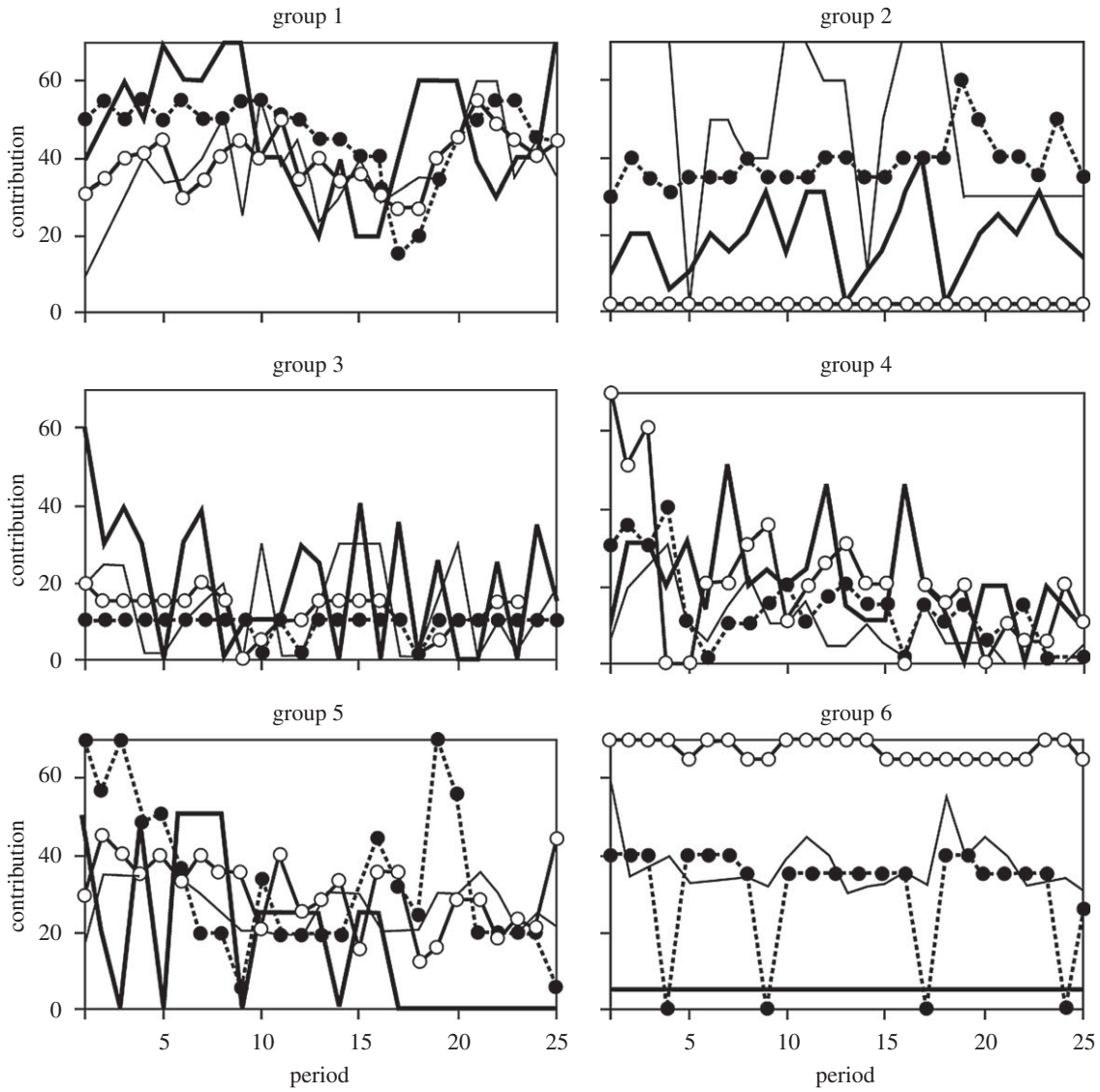

Figure 3. Individual contributions in six groups. Adapted from Hichri & Kirman (2007). Open circles, subject 1; filled circles, subject 2; thick line, subject 3 and thin line, subject 4.

that empathy is not simply an automatic 'bottom-up process', but that individuals may actually engage in 'empathy regulation' through 'top-down processes'. Indeed, there is quite a lot of evidence that empathy can be modulated by a number of social-cognitive factors such as emotion regulation, selective attention, but also by the kind of experiences individuals have lived through and their closeness to certain people. This would suggest that empathy is situation-specific or at least context dependent. The question of whether empathy can be seen as a disposition or trait, or whether it is context dependent, is currently a matter of debate (Decety & Lamm 2006; Singer & Lamm 2009).

In the public goods game we presented in some detail above, Hichri & Kirman (2007), contrary to what is done in other analyses of public goods games, focused on individual donations to a public good and found that these were not only rather dispersed but also volatile over time. This sort of behaviour could not be explained simply in terms of an empathy trait. Alternative explanations are needed. One possibility is that 'mind reading' and empathizing, together with the reaction to other people's contributions to the public good, enable players in one particular group to coordinate on something close to an equilibrium value of total contributions. Another explanation could be the following: it has been shown that people seem to have a particular self-concept of their moral self-worth (Sachdeva *et al.* 2009). This implies that people do not always have to behave in a particular way according to their self-concept, but rather tend to use this as a reference point around which they can move. If they perceive themselves to have acted very morally, they feel 'licensed' to act immorally (e.g. act in a non-cooperative way), while if they perceive themselves as not having been very moral, they need to 'cleanse' themselves by behaving morally (e.g. act cooperatively). Given that many individuals in each of these groups, in many cases, alternate between giving a lot and giving very little, this might be because they engage in this kind of moral self-regulation. It remains to be seen, of course, to what extent moral self-worth is connected with people's self-assessed empathetic character traits. If it is, which might seem plausible

given that empathy is often associated with prosocial behaviour, then it could be that people consider themselves to be high- or low-empathy people (suggesting a particular disposition), but that they smooth out selfish and altruistic actions around that particular level of empathy 'disposition'. In this case, individuals would have to be observed over a very long period in order to ascertain their natural level of empathy, if it exists. Yet, even this would not put paid to the idea that what is important is social interaction and a constant evaluation and revaluation of the social situation, as represented by the behaviour of others.

All of this suggests that we are still far from understanding the role and nature of empathy, and its consequences for altruistic or general other-regarding behaviour. Nevertheless, it also seems to be the case that recent developments in the neurosciences in general and neuroeconomics in particular may contribute to understanding the mechanisms involved. However, existing work in that field has concentrated on rather specific and simple examples from game theory such as the ultimatum, public goods or prisoners' dilemma game. It is important to see whether the results available extend to more general cases of economic decision-making and coordination. However, it is equally important that neuroscientists become fully aware of the nature and underlying assumptions of the game they want subjects to play, including the payoff structure, as these can crucially influence the results that are obtained. This would mean that by running, for example, public goods games with different levels of concavity of the payoff function, one may change the particular empathy levels that individuals apparently reveal. This of course distorts the results of the experiments. One way to test this would be to run the different experiments with the same pool of subjects, grouped in the same way. However, as is well known this can also produce distortions. But, what is important is that the members of the teams from different disciplines who run these experiments should acquire a common pool of knowledge. Or to put it more clearly, while it is certainly desirable that economists have more cognitive and neuroscientific knowledge, it is equally important that neuroscientists know more about economics.

But it would be wrong to conclude by restricting our attention purely to neuroscience and neuroeconomics. Empathy and its role put the relations between individuals at the centre of the stage. Taking up the earlier discussions in economics on this subject and linking them with recent progress in other disciplines means changing our basic economic thinking in a fundamental way. If concern for others and anticipation of each others' intentions are basic ingredients of individuals in a society, then the standard view of homo economicus is not only inadequate but also misleading.

ENDNOTES

[1]Binmore's idea of homo economicus is, of course, far from the traditional isolated maximizing individual since he interacts consciously with others. This leads him to contemplate the reactions of others and for this he has to be aware of what their utilities or payoffs are.

[2]This is similar to Binmore's concept of empathetic identification.

[3]We will refer to other-regarding preferences instead of social preferences because social preferences are traditionally attributed to preferences of a group of individuals or even of a whole society and are assumed to rank different social states.

[4]The sort of direct interference with the brain that is permitted for experiments with monkeys is not allowed for human beings.

[5]More criticism is soon to appear in a number of journals. See, e.g. the 'Talking brains' blog by Greg Hickok, a professor of cognitive neuroscience, and David Poepel, professor of linguistics and biology, and their entry on 'Eight problems for the mirror neuron theory of action understanding': http://talkingbrains.blogspot.com/2008/08/eight-problems-for-mirror-neuron-theory.html.

[6]It has been pointed out to us, by Gay Meeks, that if one pursued this argument to the limit and imagined someone perfectly capable of identifying with any member of the human race then this person would become a selfless self, incapable of self-interest!

[7]We are not quite sure why Singer *et al.* refer to this game as prisoners' dilemma game, even though this type of game usually refers to the investment game, trust game or centipede game if repeated several times.

[8]Singer, in her articles, usually refers to Rabin (1993), Fehr & Schmidt (1999) and Camerer (2003) as being representatives of theories of social preferences.

[9]The theory of inequity aversion (Fehr & Schmidt 1999) assumes that people are maximizing a utility function that takes account of a person's payoff, but also of unequal and hence inequitable outcomes. Although the two terms inequity and inequality are used interchangeably in the literature, it is important that one carries a welfare judgement (inequity), whereas the other (inequality), if it is measured by incomes or payoffs, can be measured objectively.

[10]Economists take it that the situation in repeated games is not the same as that in 'one-shot' games. In a game played repeatedly with the same players, each player can have a strategy that might 'punish' those who contribute little and therefore more cooperation would be observed than in the one-shot game. However, experiments can only involve a finite number of rounds and the previous argument only holds for games with infinite horizons. Endowed with the concept of backward induction, it is easy to see why. In the last round everyone should play Nash since there is no possibility of subsequent punishment. But, if this is so, people should do so in the penultimate round and so forth back to the start. So, in such a repeated game, people who are good game theorists, or rather act as good game theorists expect them to, should immediately settle to the equilibrium of the one-shot game. The question is again, what if they do not do so? Perhaps, since people play repeatedly they 'learn how to play Nash'. So, they do not start with a common knowledge and work everything out from the beginning, but do move towards equilibrium. Thus, after a while differences between behaviour will be ironed out and in the end something close to equilibrium would be attained. Alternatively, and this is of course Fehr and Schmidt's (and others') argument, they may be motivated by something else such as inequity aversion or other emotions and might empathize with those who receive little and, as a result, contribute more than non-cooperative game theory would suggest.

[11]That players learn to play Nash does not mean that they are even aware of this notion; it is simply based on the idea that agents reinforce the probability of playing what seem to be successful actions and diminish the probability of choosing others.

REFERENCES

Andreoni, J. 1995 Cooperation in public goods experiments: kindness or confusion? *Am. Econ. Rev.* **85**, 891–904.

Aron, A., Aron, E. N., Tudor, M. & Nelson, G. 1991 Close relationship as including other in the self. *J. Pers. Soc. Psychol.* **60**, 241–253. (doi:10.1037/0022-3514.60.2.241)

Aumann, R. 1976 Agreeing to disagree. *Ann. Stat.* **4**, 1236–1239. (doi:10.1214/aos/1176343654)

Batson, C. D. 1991 *The altruism question: toward a social psychological answer*. Hillsdale, NJ: Erlbaum.

Batson, C. D., Sager, K., Garst, E., Kang, M., Rubchinsky, K. & Dawson, K. 1997 Is empathy-induced helping due to self-other merging? *J. Pers. Soc. Psychol.* 73, 495–509. (doi:10.1037/0022-3514.73.3.495)

Binmore, K. 1990 *Essays on the foundations of game theory.* Oxford, UK: Blackwell.

Binmore, K. 1994 *Playing fair. Vol. 1 of game theory and the social contract*. Cambridge, MA: MIT Press.

Binmore, K. & Shaked, A. In press *a*. Experimental economics: where next? *J. Econ. Behav. Organ*.

Binmore, K. & Shaked, A. In press *b*. Experimental economics: where next? Rejoinder. *J. Econ. Behav. Organ*.

Blakemore, S. J. & Decety, J. 2001 From the perception of action to the understanding of intention. *Nat. Rev. Neurosci.* **2**, 561–567.

Bowles, S. 1998 Endogenous preferences: the cultural consequences of markets and other economic institutions. *J. Econ. Lit.* **36**, 75–111.

Bowles, S. & Gintis, H. 2003 The origin of human cooperation. In *The genetic and cultural origins of cooperation* (ed. P. Hammerstein). Cambridge, MA: MIT Press.

Camerer, C. F. 2003 *Behavioral game theory: experiments in strategic interaction*. Princeton, NJ: Princeton University Press.

Chartrand, T. L. & Bargh, J. A. 1999 The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.* **76**, 893–910. (doi:10.1037/0022-3514.76.6.893)

Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C. & Neuberg, S. L. 1997 Reinterpreting the empathy-altruism relationship: when one into one equals oneness. *J. Pers. Soc. Psychol.* **73**, 481–494. (doi:10.1037/0022-3514.73.3.481)

Darwin, C. 1872 *The expression of the emotions in man and animals*. London, UK: John Murray.

de Quervain, D. J. F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A. & Fehr, E. 2004 The neural basis of altruistic punishment. *Science* **305**, 1254–1258. (doi:10.1126/science.1100735)

de Waal, F. B. M. 2008 Putting the altruism back in to altruism: the evolution of empathy. *Annu. Rev. Psychol.* **59**, 279–300. (doi:10.1146/annurev.psych.59.103006.093625)

Decety, J. & Lamm, C. 2006 Human empathy through the lens of social neuroscience. *Sci. World J.* **6**, 1146–1163. (doi:10.1100/tsw.2006.221)

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V. & Rizzolatti, G. 1992 Understanding motor events: a neurophysiological study. *Exp. Brain Res.* **91**, 176–180.

Fehr, E. & Camerer, C. 2007 Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* **11**, 419–427. (doi:10.1016/j.tics.2007.09.002)

Fehr, E. & Schmidt, K. 1999 A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114**, 817–868. (doi:10.1162/003355399556151)

Fehr, E. & Schmidt, K. 2004 Fairness and incentives in a multi-task principal-agent model. *Scand. J. Econ.* **106**, 453–474. (doi:10.1111/j.0347-0520.2004.00372.x)

Fehr, E. & Schmidt, K. In press. On inequity aversion: a reply to Binmore and Shaked. *J. Econ. Behav. Organ.*

Fehr, E., Klein, A. & Schmidt, K. 2007 Fairness and contract design. *Econometrica* **75**, 121–154. (doi:10.1111/j.1468-0262.2007.00734.x)

Fehr, E., Kremhelmer, S. & Schmidt, K. 2008 Fairness and the optimal allocation of property rights. *Econ. J.* **118**, 1262–1284. (doi:10.1111/j.1468-0297.2008.02169.x)

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F. & Rizzolatti, G. 2005 Parietal lobe: from action organization to intention understanding. *Science* **308**, 662–667. (doi:10.1126/science.1106138)

Fontaine, P. 1997 Identification and economic behavior: sympathy and empathy in historical perspective. *Econ. Philos.* **13**, 261–280. (doi:10.1017/S026626710000451X)

Fontaine, P. 2001 The changing place of empathy in welfare economics. *Hist. Polit. Econ.* **33**, 387–409. (doi:10.1215/00182702-33-3-387)

Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. 1996 Action recognition in the premotor cortex. *Brain* **119**, 593–609. (doi:10.1093/brain/119.2.593)

Gul, F. & Pesendorfer, W. 2008 The case for mindless economics. In *The foundations of positive and normative economics* (eds A. Caplin & A. Schotter), pp. 3–42. Oxford, UK: Oxford University Press.

Hichri, W. & Kirman, A. 2007 The emergence of coordination in public good games. *Eur. Phys. J. B Condens. Matter Complex Syst.* **55**, 149–159. (doi:10.1140/epjb/e2007-00032-8)

Hoffer, E. 1955 *Passionate state of mind*. New York, NY: Harper Collins.

Hume, D. 1740 [2000] *A treatise of human nature* (eds D. F. Norton & M. J. Norton). Oxford, UK: Oxford University Press.

Iacoboni, M. 2009 Imitation, empathy and mirror neurons. *Annu. Rev. Psychol.* **60**, 653–670. (doi:10.1146/annurev.psych.60.110707.163604)

Jeannerod, M. 2006 *Motor cognition*. Oxford, UK: Oxford University Press.

Knoch, D. & Fehr, E. 2007 Resisting the power of temptations: the right prefrontal cortex and self-control. *Ann. NY Acad. Sci.* **1104**, 123–134. (doi:10.1196/annals.1390.004)

Lieberman, M. D. 2007 Social cognitive neuroscience: a review of core processes. *Annu. Rev. Psychol.* **58**, 259–289. (doi:10.1146/annurev.psych.58.110405.085654)

Oosterbeek, H., Sloof, R. & van de Kuilen, G. 2004 Differences in ultimatum game experiments: evidence from a meta-analysis. *Exp. Econ.* **7**, 171–188. (doi:10.1023/B:EXEC.0000026978.14316.74)

Oullier, O. & Basso, F. 2010 Embodied economics: how bodily information shapes the social coordination dynamics of decision-making. *Phil. Trans. R. Soc. B* **365**, 291–301. (doi:10.1098/rstb.2009.0168)

Oullier, O., Kirman, A. & Kelso, J. 2008 The coordination dynamics of economic decision making: a multilevel approach to social neuroeconomics. *IIEEE Trans. Neural Syst. Rehabil. Eng.* **16**, 557–571. (doi:10.1109/TNSRE.2008.2009960)

Platek, S. M., Mohamed, F. B. & Gallup, G. G. 2005 Contagious yawning and the brain. *Cogn. Brain Res.* **23**, 448–452. (doi:10.1016/j.cogbrainres.2004.11.011)

Preston, S. D. & de Waal, F. B. M. 2002 Empathy: its ultimate and proximate bases. *Behav. Brain Sci.* **25**, 1–72.

Rabin, M. 1993 Incorporating fairness into game theory and economics. *Am. Econ. Rev.* **83**, 1281–1302.

Rizzolatti, G. 2005 The mirror neuron system and its function in humans. *Anat. Embryol.* **210**, 419–421. (doi:10.1007/s00429-005-0039-z)

Rizzolatti, G. & Craighero, L. 2004 The mirror-neuron system. *Ann. Rev. Neurosci.* **27**, 169–192. (doi:10.1146/annurev.neuro.27.070203.144230)

Sachdeva, S., Iliev, R. & Medin, D. 2009 Sinning saints and saintly sinners. *Psychol. Rev.* **20**, 523–528.

Sanfey, A. G. 2007 Social decision-making: insights from game theory and neuroscience. *Science* **318**, 598–602. (doi:10.1126/science.1142996)

Sen, A. 2006 *Identity and violence*. New York, NY: Norton.

Singer, T. 2006 The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. *Neurosci. Biobehav. Rev.* **30**, 855–863. (doi:10.1016/j.neubiorev.2006.06.011)

Singer, T. & Fehr, E. 2005 The neuroeconomics of mind reading and empathy. *Am. Econ. Rev.* **95**, 340–345. (doi:10.1257/000282805774670103)

Singer, T. & Lamm, C. 2009 The social neuroscience of empathy. *Ann. NY Acad. Sci.* **1156**, 81–96. (doi:10.1111/j.1749-6632.2009.04418.x)

Singer, T., Seymour, B., O'Doherty, J. P., Kaube, H., Dolan, R. & Frith, C. 2004 Empathy for pain involves the affective but not sensory components of pain. *Science* **303**, 1157–1162. (doi:10.1126/science.1093535)

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. & Frith, C. 2006 Empathetic neural responses are modulated by perceived fairness of others. *Nature* **439**, 466–469. (doi:10.1038/nature04271)

Smith, A. 1759 [1976] *The theory of moral sentiments* (eds D. D. Raphael & A. L. Macfie). Oxford, UK: Oxford University Press.

Sorabji, R. 2006 In *Self: ancient and modern insights about individuality, life, and deaths*. Chicago, IL: University of Chicago Press.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V. & Rizzolatti, G. 2003 Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron* **40**, 655–664. (doi:10.1016/S0896-6273(03)00679-2)