What can Neuroscience offer to Economics?

Matteo Colombo[*]
M.Colombo-2@sms.ed.ac.uk

ABSTRACT

The specific regions in the brain that are active when some behaviour is observed is a kind of information that may be interesting for neuroscientists, but how could it be fruitful for economic theory? The thesis defended in the essay is that the brain matters to prediction. By using the Ultimatum Game as a benchmark, it is argued that if the goal of a model of human behaviour is to yield good predictions about important classes of choices, then models that incorporate neurobiological variables may have some advantages over alternative models. The essay comprises two parts. Part I first analyses the Ultimatum Game and illustrates some of its experimental results. Then, it evaluates in detail the merits and shortcomings of Cristina Bicchieri's model based on social norms. It centres on the predictive power of the model, and articulates some challenges it faces. Part II begins with a review of some neurobiological findings which suggest a different approach to construct predictive models of human behaviour. Drawing upon these findings, it gives some reasons why the predictions of a neurobiologically-informed model seems to have some advantage over those of Bicchieri's model. A critical discussion of the thesis against possible objections terminates the essay. The conclusion follows that one way in which the study of the neurobiological foundations of decision-making might be fruitful for economic modelling is in enhancing the predictive quality of its models.

INTRODUCTION

The specific regions in the brain that are active when some behaviour is observed is a kind of information that may be very interesting for neuroscientists. But what does it add to economic theory? The claim defended here is that the type of knowledge of brain processes offered by neurosciences matters to prediction (see also Camerer 2007). To the extent I am right, the attempt to integrate evidence, concepts and tools from the fields of economics, psychology and neuroscience within the new domain of neuroeconomics will turn out to be a realization of the methodological ideal described by Milton Friedman in "The Methodology of Positive Economics". There, Friedman advocates a requirement of predictive success for judging a "positive" scientific theory: "The ultimate goal of a positive science is the development of a "theory" or "hypothesis" that yields valid and meaningful (i.e. non truistic) predictions about phenomena not yet observed." (Friedman 1953, p.7)

An important cautionary note right from the beginning: It may sound that I focus on prediction in order to avoid "the real" problems – e.g. issues about explanation, and understanding. I would like to reply to this charge by giving it a different twist. One way to read my claim is that *however* you want to explain, or understand human behaviour, a model that takes into account neurobiological parameters seems to have an advantage over competitors for predictions. If we consider the importance that predictions have in our lives, it may be easier to acknowledge that the choice to limit a research on prediction is still worthwhile. Prediction serves at least two crucial goals: one pragmatic, the other epistemic

[*] Department of Philosophy, University of Edinburgh, UK

(see Rescher 1997). On the one hand, predictions are necessary to interact successfully with our environment. E.g., predicting that if we jump off the top of the Cathedral of Milan, nasty consequences for our health will be extremely likely to ensue, we may prefer to take the stairs to go down. On the other hand, prediction is a test of scientific theories. It is a common view that 'in assessing the confirmation or evidential support of a hypothesis, we must take into account especially (and perhaps exclusively) the predictive success or failure of its *predictions*' (Musgrave 1974, p. 2).

Before spelling out the two criteria that will orient my assessment of the comparative merits and problems of the predictions given by concurrent models, it is worth making clear what it is meant by 'prediction'. Following Forster (2008), '[t]he term prediction is always used to refer to the "diction" of past, present, and future events', the "diction" of a claim that previously we had no reason to believe. Hence, a good prediction is not necessarily to refer to the future. I take a good prediction to be one that is both *secure* and *informative*. A secure prediction is based on reliable, well-evidentiated grounds. Because of this, it is likely that the prediction turns out to be correct. E.g., the prediction that there will be a full moon within the next thirty days is secure. The more adjustable parameters a model has, the more secure its predictions, but the greater the risk of accommodation. *Accommodation* is one of the risks of a secure prediction. A model accommodates the data when it is merely consistent with them. The typical case of accommodation is when a set of data is deduced from the model (hence, model and data are consistent), *and* the same set of data was used in the construction of the model. E.g., a model of the form (H&Y) does not genuinely predict that Y: It accommodates Y. A good prediction is not trivial, is *informative*. Informative predictions are not vague, preferably they are quantitatively accurate. E.g., the prediction that the comet Hale-Bopp had its closest approach (at a distance of 1.315 AU) to Earth on March 22, 1997 is informative.

With these conceptual tools at hand I can now make clear the structure of my argument. The Ultimatum Game (UG) is my benchmark for evaluating the predictive quality of a model. I begin by detecting an "anomaly"[1] in the UG. Standard game-theoretic prediction is at loss in the face of the behaviours of people playing the UG. Thus, if we want to stick to the predictive success requirement, we must revise our model. Several enriched models have been proposed. Some of them, because of their flexibility, seem promising to predict well. However, I argue that this flexibility comes at some cost: Because it is problematic to obtain a reliable measure of the adjustable parameters they contain, these models are particularly subjected to merely *accommodate* the evidence. That's the main reason why we should try to develop a model with adjustable parameters flexible enough to predict the variety of "anomalies" observed in the UG, *and* specific enough to figure out how to reliably measure them so that the model enables us to give predictions that are both informative and secure. I argue that a model that incorporates neurobiological variables seems to satisfy these requirements. However, neurosciences offer no panacea. I discuss some of the limits of the type of neuroscientific results that back my argument, and I provide positive suggestions on how these shortcomings might be tackled.

If my argument is correct, it follows that *if* we want to take Friedman's advice seriously, then we have good reasons to try to incorporate neurobiological parameters in our models of decision-making in economic environments.

[1] Borrowing Richard Thaler's words: an anomaly is 'an empirical result … if implausible assumptions are necessary to explain it within the [rational choice] paradigm' (Thaler 1988, p. 195).

The essay is organized in two parts.

The first part is divided into two sections. Section *I* explains why I believe that the UG is interesting and worth studying. I describe the UG, its game theoretic analysis, and how people play the game. Section *II* introduces the concept of "enriched models". These models try to account for the actual behaviour of the people who play the UG by appealing to such concepts as "fairness", "warm glow", "envy", "social norms". I restrict my analysis to one of these proposals. After having motivated my choice, I analyze in detail how the norm-based approach defended by Bicchieri (2006) works when it's called for accounting for the UG results. I argue that this kind of account risks to merely accommodate evidence and predictions because of the nature of its adjustable parameters.

The *second* part of the essay is shorter. The reason is simple, and it is important to be clear: Currently, there is no neurobiologically-informed model. Drawing on both the detailed analysis of the first part and on some recent neurobiological findings, my goal is to give some reasons why *if* we want good predictions, then we should try to point to quantifiable biological variables which have a large influence on behaviour and are underweighted or ignored both in game theoretic, and enriched models such as Bicchieri's. After having described one study on the neural basis of economic decision-making in the UG, I critically discuss possible objections to the potential predictive significance of neurobiological variables.

I. THE ULTIMATUM BARGAINING

Game theory is a collection of models attempting to understand situations in which decision-makers interact with one another. Game theoretic analyses predict that *rational*, *self-interested* players will make decisions to reach outcomes, known as *Nash equilibria*, from which no player can increase her own payoff unilaterally. Strategic bargaining behaviour is one of the concerns of game theory.

To see the role played by the assumptions of rationality and self-interest in game theory, let us consider the Ultimatum (or "take-it-or-leave-it") Game which is one of the simplest form of bargaining. This two-stage, two-person game is defined as follows. A sum of money m is provided. Player 1 proposes that x units of the money ($x \leq m$) be offered to player 2. Player 1 would retain ($m - x$). Player 2 responds by either accepting or rejecting the offer x. If player 2 accepts, player 1 is paid ($m - x$) and player 2 is paid x; if she rejects, each player receives nothing (0, 0). In either case the game is over.

Two features of the UG are worth emphasizing. First, the UG is a non-cooperative game: Players cannot make binding agreements about what to do. They have to form expectations about other players' action without communicating. Second, the UG is very simple in its game-theoretic analysis: It requires only two assumptions to make a prediction. It is simple in its instructions: Players understand quickly and without effort the rules of the game.

'The UG is one of the most successful experimental designs in the history of the social sciences' (Guala 2008). According to Guala, the success of the UG can be explained if we focus on some of the epistemic features that qualify it as a "paradigmatic experiment". One of these features is its versatility: A paradigmatic experiment enables us to make comparisons and draw a variety of inferences in different contexts of scientific inquiry. This versatility makes the UG suitable for my argument, where I evaluate the comparative merits and problems of the predictions given by two models built in two different contexts.

The game theoretic analysis of the game yields a precise prediction about what players will do. The prediction of a restrictive concept in game theory, *the subgame perfect equilibrium*, is that for any positive amount offered by Player 1 (the proposer), Player 2 (the responder)

knows that she faces a choice between gaining nothing (if she refuses the offer) or something (if she accepts). If the responder maximizes her own payoff, she will accept any positive amount. If the proposer maximizes her own payoff *and* expects the responder to maximize, she will offer the smallest amount possible. Hence, the proposer will offer the minimum possible split to the responder, who will accept.

The assumptions of *rationality* and *self-interest* entail this prediction. According to the self-interest assumption, players prefer more money to less, and don't care about the outcomes or preferences of others. Notice that this is a narrow conception of "self-interest", concerned with money alone. Rationality has to be understood as *practical, instrumental, rationality*. Given the agent's perfect knowledge of the outcomes of the alternatives open to her and to the other player, and given that she can identify the best of them, she will be practically rational in choosing action *A* if *A* is the action she believes will lead to the consequences she prefers.

The experimental literature on the UG indicates a robust behavioural pattern at odds with the game theoretic prediction. Since the first experiment which studied the UG (Güth, Schmittberger, & Schwarze 1982), the UG has been studied in many diverse settings where different parameters of the game were modified. Supposing that the total sum of money is 10, the split offered is typically around (6,4). And low offers, namely offers around 2 or less out of 10, are very likely to be refused (Camerer 2003).[2] The UG is typically *anonymous* and *one-shot*: Players don't know the identity of the other, and play only once. The rationale for these two characteristics is to abstract away from the possibility of incentives for reciprocity and cooperativeness - which would be involved in repeated games with the same partner - and to keep players' behaviour insulated from such influences as the desire to please the experimenter, or the fear of ruining a friendship, or the incentive to build a good reputation – which would be involved if the identity of the player was known.

The behavioural pattern displayed by people playing the UG immediately raises the well-known Duhem problem: Why do people tend *not* to play the subgame perfect equilibrium, and instead tend to coordinate on 50-50 or 60-40 splits? What part of the standard game-theoretic model has to be blamed for the anomaly? And, once we have identified the culprit, what modification has to be done so as to enable the model to predict well?

II. BICCHIERI'S SOCIAL NORM MODEL. AN ANALYSIS

By virtue of the simplicity of the UG, we have reason to maintain that the players are rational in the minimal sense specified above: their actions follow from their preferences and beliefs. What seems to be revised is the assumption of strict self-interest that individual preferences are concerned with money alone. Accordingly, the target of most of the theoretical developments in the game-theoretic literature of the last two decades has been the assumption of strictly self-interested preferences. The most common move has been to build models with non-standard utility functions, according to which individuals have "other-regarding" preferences. This kind of models allows a player's utility function to take into account the outcomes, preferences, and expectations of *other* players. Notice that a common feature of these "enriched" models is the preservation of the logical framework of expected

[2] There is however some variation in the findings. The most remarkable variations are found either across cultures or across subjects with a neurological condition such as autism. (see Roth *et al* 1991, and Henrich *et al*., eds 2004 on cross-cultural variations; Sally & Hill 2006 for a research with autistics).

utility theory: They do not reject the rationality assumption, they point to the maximization of a non-classical utility function whose empirical substance is provided by the new parameters.[3]

In this section, I critically analyze the "enriched" model developed by Cristina Bicchieri in *The Grammar of Society* (2006) (henceforth, GS). I focus on one model to facilitate a more detailed discussion than it would be possible if I were to consider a wide range of current views. I focus on Bicchieri's for two reasons. First, hers is one of the most recent and promising proposals; second, and importantly, she argues that her model fares better than the alternatives when it comes to prediction. In my discussion I pay special attention to whether, and to what extent, this last claim is justified.

Every known society has a multitude of social norms that regulate the behaviours of individuals in a variety of situations. In a given social context, the same kind of social norm might produce different behavioural patterns across individuals. If most people abide by social norms, then norms can account for behavioural patterns observed in a population. And different norms can account for behavioural variations across societies. Consequently, social norms are important for predicting the behaviour of individuals. These observations call for elucidation:

1) What is it meant exactly by a 'social norm'?
2) What are the *mechanisms* that regulate the power social norms have to influence human behaviour?
3) What are the *conditions* under which individuals are likely to follow a social norm?

In GS, Bicchieri attempts to account for these three problems by providing the foundations for a new model of human behaviour based on a precise characterization of a *social norm*.

Definition According to Bicchieri, a rule is a social norm in a population if and only if a sufficient number of people in that population:

(i) know that the rule exists and applies in situations of a certain type S, and
(ii) prefer to conform to it, in situations of type S, on the condition that
  (a) it is believed that a sufficient number of others conform to it in situations of type S, and, either
  (b) it is believed that a sufficient number of others expect one to conform in situation of type S, or
  (b`) it is believed that a sufficient number of others expect one to conform in situation of type S, prefer conformity, and may sanction one if one does not conform (GS, p.11).

Suffice here to clarify three points about Bicchieri's definition. First, a social norm is a set of mutual expectations, and a communality of beliefs is a precondition for its existence. A social norm like "tipping for service in a restaurant" has no reality other than our expectations that others leave a tip in a restaurant, *and* that others expect us to tip in the same type of circumstance. Second, a social norm cannot be simply identified with a recurrent collective behavioural pattern. Taking a shower in the morning is not a social norm. I take a shower in the morning whether or not I expect others to do the same. Moreover, a rule can be a social norm in a population, even if compliance to it is not observed. Imagine the social norm in a population that whoever first makes a proposal that something has to be done is directly responsible for making sure that the proposal is carried out. During a seminar students may

[3] Well known "enriched" models are Rabin (1993), Fehr & Schmidt (1999), Bolton & Ockenfels (2000).

avoid suggesting a certain topic of discussion fearing that that social norm will be followed, and, hence, they would have to prepare the talk. In this situation nobody is violating the norm. Everybody is eluding it. Third, the conformity to a social norm is *conditional*: One has a preference to conform to a norm *N* in a situation of type S, under the conditions that one expects others to conform to *N* in S (*empirical expectations*), and one believes that others think one ought to conform to *N* in S (*normative expectations*). This condition is discussed in detail later. Notice, for the moment, that the conditionality of the preference to conformity to a norm lends itself to empirical testing. Were the expectations that underlie a social norm to be different, we would predict behaviour to change in determinate ways. Clearly, unless we have an account of when, how, and to what degree the expectations that constitute a social norm affect behaviour, there is little hope in drawing precise, informative predictions. We need understand the *mechanism* by which social norms influence our behaviour, and the *conditions* under which individuals are likely to follow a social norm.

Mechanism Bicchieri undertakes the first task by relying on findings from experiments in cognitive psychology (GS, Ch.2). The mechanism works as follows. Subjects interpret and categorize a given context as a function of the situational cues that spark their attention. The process of categorization relies on *spreading activation*. That is, the activation of the representation of a certain concept spreads to representations of concepts related to it. E.g., when we are presented with the stimulus word *tiger*, we retrieve not just the representation of a tiger, but also related representations like *feline*, *predator*, etc. Then, depending on how subjects categorize the context, a script of a certain type activates. *Scripts* are cognitive structures we acquire through personal experience and habit that represent stored knowledge about people, objects, events, and roles relevant to the situation at hand. Scripts prompt beliefs and expectations about social roles and sequences of actions appropriate in that situation. A "restaurant script", e.g., represents roles (waiters and diners) and sequences of appropriate actions (diners enter the restaurant, wait to be seated at a table; waiters take their order; diners eat, ask for the bill, pay, leave a tip, and leave the restaurant). '*Social norms are embedded into scripts*' (GS, p.94): Social norms are among the set of beliefs, expectations, and preferences prompted by a script. To see this mechanism in action, consider an UG framed in terms of the gains from a transaction between a buyer and a seller. Player 1 is said to be the seller; she is endowed with €10. Player 2 is said to be the buyer. A table registers the profit of the seller and of the buyer for each price (€0, €1, €2,…, €10) charged by the seller when the buyer decides to purchase. The profit of the seller is equal to the price she states; the profit of the buyer is €10 minus that price. The profit of each is zero if the buyer refuses to purchase at the price stated by the seller. In such context, the cues provided are likely to guide the interpretation of the situation in terms of a market situation. Presumably, the categorization "market exchange" activates a script that defines determinate mutual expectations underlying a social norm (if it exists) among the players.

When compared with the results of an UG with standard instructions, it is found that the "buyer-seller" manipulation elicits lowered offers, whereas the rejections rates remains unchanged (Hoffman *et al* 1994). Indeed, in western culture the right of sellers to quote a higher price is not usually questioned, nor that of the buyer to decide to purchase or not to purchase. Two points are worth noticing. First, consistently with Bicchieri's account, the example just provided unambiguously shows that context matters. Two UGs with the same logical structures, but embedded in different contexts, with different situational cues, are likely to elicit different behaviours because they are likely to prompt different social norms. Second, as acknowledged by Bicchieri, '[*t*]*he predictive power of a theory of norms therefore depends*

on knowing which situational cues trigger which norm' (GS, p.76). But it also depends on a number of other conditions. In the remaining of this section it is argued that the fulfillment of these conditions represent important challenges to the predictive success of her model.

Conditions for prediction There are two conditions under which an individual is likely to follow a social norm *N* in a context of type *S*. First, *N* is correctly identified. Second, individuals in *S* have the right kind of expectations, and therefore *N* exists in *S*.

Identifying a norm There are at least three reasons why *identifying N* in a certain context might be problematic. The first is that, if we exclusively focus on the behaviour of people e.g. in the UG, there might be alternative social norms, *N*, *N`*,…,*N*^{n}, that entail the same behavioural pattern. This is the problem of (deductive) *underdetermination*, and it would threat the informativeness of the prediction given by the model. If we want our model to give informative predictions, we need an assessment of the social norm likely to be in place in the UG *independent* of the behaviour observed in the game itself. Independent measurement is one of the remedies for limiting the bite of the underdetermination problem. With social norms, however, independent measurement is made difficult because they might be too vague, and consequently it might not be clear how to assess them independently of behavioural data. This is the second problem. The third is that individuals of a population might be unaware to be in a context where *N* applies. This has to do with the situational cues, crucial in priming *N*.

Underdetermination Consider this slightly different form of the UG. Before the responder hears the offer, she must set an acceptable offer range: she is asked whether she would accept a 100-0 split, and then whether she would accept a 90-10, a 80-20, a 70-30, etc, until a point is reached where she would accept anything higher. If player 1's offer is below her acceptable range, her response would count as a rejection. Player 1's proposal is finally revealed. The players belong to the same population, and there are no contextual cues that affect their expectations. An experiment with Gypsies in Vallecas, Madrid, shows that in this situation although 97% of proposers offered an equal split, as responders the Gypsies were willing to accept completely unfair offers: The acceptance of the *zero* offer was the modal value (Pablo *et al* 2006). In order to have predicted such behavioural pattern on the side of the respondent, we would have needed to know about the social norms existing in that population relevant to an UG-situation. One possible way to make sense of this behavioural pattern is by means of a norm of hospitality. Hospitality seems to be a social norm that primarily affects the behaviour of those who can offer: The proposer offers half and the other accepts whatever is offered. Once this norm has been recognized as obtaining in that situation, Bicchieri's model wouldn't have problems in predicting the behaviour of the Gypsies. However, the same behavioural pattern is also compatible with a different social norm like "help the needy". According to this norm one would expect that if the proposer offers zero, then she is needy, and this would motivate the respondent to accept a zero offer. With some imagination we can devise other social norms compatible with the same behavioural pattern. Because of underdetermination, it would be logically possible to find an infinite number of models that will accommodate the same behavioural pattern without genuinely predict it. One way to tackle the problem is to "measure" the expectations underlying the social norm we assume to be in place *independently* of the observed behaviour. As convincingly argued by Larry Laudan in different places (e.g. Laudan 1990), deductive underdetermination does *not* entail that the *choice* of a model is underdetermined. One of the criteria used to decide which model is better, and consequently to be preferred over rivals, is independent measuring of its assumptions. Independent evidence about a social norm in a population would facilitate us to uncover the

real expectations and motives behind the behaviour in an UG. And knowing motives and expectations would enable us to prefer a model that yields informative predictions.

Vagueness and Independent Measures How we measure a social norm independently of the observed behaviour is problematic since social norms are complex, vague, and might be impossible to define them in a precise way. A norm of hospitality can be regarded as an example of a vague social norm. I use 'vague' meaning that hospitality cannot be precisely defined because it has many diverse aspects. In a typical situation where hospitality exists we might expect that the host invites the guests in her house; the host entertains them with kindness; she offers food and drinks; the guest accepts whatever is offered with goodwill, etc. Hospitality might exist in a context even if any one of these expectations is missing as long as people in that context share some of the other expectations. Yet, there are at least three methods to measure a social norm: first, questioning people; second, inferring norms from behaviour (other than that displayed in the UG); third, in-depth ethnographic research. Each of these methods has biases that may undermine the reliability of the measure.

"Which split proposed in an UG would be fair?", or "What is the norm to follow in situation of type S?" are the kind of questions that may enable us to determine the content of a social norm of fairness in a population. Questioning people, however, may yield an unreliable measure because they may lie, may give the answer they suppose the research wants, may understand the question in different ways.

Apart from the reasons given above for why a social norm cannot be simply identified with a recurrent collective behavioural pattern, there are at least two further challenges in measuring a social norm from behaviour. First, to account for behaviour by citing a social norm may lead to circularity if the norm is first measured from the behaviour in question. Second, there is the problem of simultaneous attribution of belief and desire, namely the problem of 'discriminating the respective roles played by an agent's beliefs and desires in the production of the actions we observe her to perform' (Bradley unpublished). On the basis, e.g., of the observation that a number of people leave a tip at a restaurant, we might attribute to them the belief that there exists a social norm of tipping at a restaurant, and the desire to follow it. Alternatively, we might attribute to them a belief that the service has been really good, and the desire to reward good service regardless of there being a social norm of tipping at a restaurant. The two alternatives entail different behaviours under different circumstances: After an unsatisfactory service, according to the first belief-desire attribution one may still leave a tip following the social norm; instead, according to the second belief-desire attribution one will not. The choice as typically observed does not *alone* allow us to decide between these two alternatives, and the many other possible ones. To be sure, this is not a knock-out challenge since each hypothesis is testable in principle; that is, there are conditions under which we can determine whether one of them is probably false. Nonetheless, the problem remains to specify an underlying theory of beliefs and desires which enables us to systematically infer one's mental states from her observed behaviour.

The third method, that of ethnographic research, has been adopted by a group of economists and anthropologists who set out to study the foundations of human sociality through classic economic experiments like the UG in fifteen "small-scale societies" in South America, Asia, and Africa (Henrich *et al* eds. 2004). For each population experimental research was flanked by independent information concerning social context, political and economic structure, religious beliefs, etc. It turned out that the between-group behavioural differences in the UG were related to indicators of patterns of social and economic interaction (e.g. level of market exchange, importance of anonymity in commercial transactions) that framed the daily life of the population. Yet, there is no guarantee that a significant correlation will be

always found between behavioural patterns in the UG, and the socio-economic variables that researchers decide to study. This may be due either because there is in fact no correlation between UG decisions and variables of socio-economic patterns, or because the researchers don't collect the appropriate information, or because the information they collect from certain individuals is not representative of the whole population but of only some subgroup which has to be identified.

Context and Cues The third condition that would enable us to predict that an individual is likely to follow a norm is that the individual has to be aware to be in a context where the norm applies. Situational cues govern the mapping between context, recognized as being of a certain type courtesy of the cues, and activation of the social norm appropriate to that context. But, what drives people's attention to situational cues? The key notion to tackle this question is *salience*. A salient item stands out relative to neighbouring items, thereby sparking people's attention. A red dot surrounded by green ones is salient. Saliency is perhaps best understood in the field of visual perception where it is defined in function of such cues as colour, intensity, orientation, and motion. In the case of norms, salient cues – Bicchieri suggests (GS, p.112) - 'may involve a direct statement or reminder of the norm, observing others' behaviour, similarity of the present situation to others in which the norm was used, as well as how often or how recently one has used the norm.' Salient cues prime the expectations underlying the norm. Conditional on these expectations, certain behaviour is likely to ensue.

Bicchieri tested her model by manipulating salience. In an UG, people's attention has been cued by information about the normative expectations of others that had played the game before. The prediction obtained that more players would follow a norm of fairness, which in that context would dictate an equal split (Bicchieri & Xiao 2008).

The potential problem for prediction here stems from a tension within the model. On the one hand, Bicchieri's model makes reference to *type*-situations. On the other hand, the appeal to situational cues makes the interpretation\activation of social norms dependent on specific, *token*-situations. The interpretation of a social norm is *local*: Fairness, e.g., has different meanings in different circumstances, depending on the people, objects and environment that define the situation. A *type-situation* is a general type of situation like "football match", "bargaining", "theatre-play". A *token-situation* is a specific situation of a certain type performed in a particular context. Hamlet played at La Scala now, with a certain setting, certain actors, and a certain audience, is a token-situation of the type "Hamlet-theatre-play". We can *judge* that the individual *i* is in a situation of a certain type; e.g. we can judge that *i* is at a theatre play of Hamlet. But *i* makes decisions, has expectations, follows a social norm, always in token-situations. E.g. *i* expects to smile when the Hamlet she is attending is being performed by a company of funny comics wearing fancy dresses; instead, she expects something deep when the play is being performed by a company of continental philosophers. The two situations are of the same type, but differ in situational cues. The difference is likely to prompt different expectations, preferences and actions. Consider again the UG. It has been shown that in a common UG-token-situation, a bargaining type-situation, attractive people are treated differently by others in that they are offered more, and they are expected to give more (Solnick & Schweister 1999). Now, in a token-situation of a certain type there may be innumerable situational cues that can spark one's attention. The attractiveness of the people in a bargaining type-situation can be one. In function of our personal history, we will interpret the context in one way or another, we will find one person attractive or not. The interpretation of the context will affect our expectations and preferences, and hence may prime this or that norm we may follow in that token-situation. Even if we can judge that *i* is in a type-situation *S*, we still need careful inquiry into the token-situation. The potential problem for prediction is here with informativeness. For we would like a model that could be applied to type-situations

besides the token-ones from which it was deduced. We would need a general underlying theory that specifies some systematic functional relationship between situational cues, which are adjustable variables of the model, and experimental results. We would like something analogue to a saliency map, which represents visual saliency of a corresponding scene, made available by neurocognitive research on visual perception (e.g. Koch & Ullman 1985). As far as I know there are no such "maps" in the field of decision-making yet. The challenge is in measuring, or quantifying, the internal state of an individual, such as her personal history, the goals, beliefs and motivations she has at a time.

Having the right kind of expectations One has a *preference* to conform to a norm *N* in a situation of type S, provided *N* exists and has been correctly identified, under the conditions that she expects others to conform to *N* in S (*empirical expectations*), and she believes that others think she ought to conform to *N* in S (*normative expectations*). There are at least two problems here. First, the conditions are *not* sufficient to determine one's following a rule. This may threat the security of the predictions the model gives. Second, there is no hint as to how (quantitatively) determine the *likelihood* of observed behaviours. This is a drawback for predictive informativeness.

The insufficiency of having the right kind of expectations is apparent: We often have other personal motives for not following a norm. I may live in a society with a strong norm of revenge. I expect that most people take vengeance on those who have wronged them, and I believe that others think I ought to take revenge on those who have wronged me. However, I have also a stronger preference for not harassing others. Bicchieri acknowledges that the presence of a norm of revenge, 'and its salience in a particular situation, motivate me to act in a congruent manner, but my behaviour is ultimately explainable [and predictable] only by reference to my preferences and expectations' (GS, p.22). Suppose we ask the responder in an UG what she believes is fair in that UG-situation, and whether she expects others to follow it. This might give us information that this person believes a norm *N* applies in that situation, and this would lead us to predict her action in accordance with *N*. However, such action might still not occur, due to the presence of other motives and incentives which, unbeknownst to us, bear on her decision more than her expectations related to *N*. To remedy this problem, Bicchieri suggests (Bicchieri & Chavez 2008) that a fine-grained account of individuals' sensitivity to specific norms would give us more reliable grounds to predict their behaviour since it would enable us to compare the strength of concurrent motivations in a given situation - e.g. the preference to follow a norm versus other preferences that may overcome that preference. How does the suggestion work exactly? Consider responders and proposers in an UG. Each player can be of different types. In order to measure the expected norm-sensitivity, call it K, of a type of proposers (P) 'we may ask responders about the expected type of offer versus the offer that is fair. Then we would have indirect information about the expected value of KP: If the two values differ, the expected KP is low. If they are the same, the expected KP will be high or low depending on the (high or low) values of the expected type of offer and the offer that is fair. If we ask proposers about the distribution of normative expectations on the part of responders (R), and we then observe their offers, we can get information about the expected KR.' (Bicchieri, personal communication). I see three problems with this suggestion.

First, Bicchieri's reasoning entails that we can *only* have indirect evidence about one's K. The information we obtain following her suggestion is a measure of the type of individuals expected by the *others*. The grounds to ascribe a certain K to a responder in an UG would be the proposers' beliefs about the responders' normative expectations, seeing whether the

proposers behave according to their beliefs. The problem is whether this kind of evidence gives us grounds strong enough to reliably ascribing Ks to respondents. My suggestion is that direct evidence about one's K would do a better job. For a proposer may believe e.g. that a responder has strong normative expectations of fairness, and accordingly she makes a fair proposal; yet, the respondent's K doesn't match with the proposer's expectations since she also accepts unfair offers. The contrast between having direct evidence that the individual *i* is sensitive to a norm and having indirect evidence thereof is between *i*'s having certain expectations and preferences and *her* saying so, and *another*'s believing that *i* has certain expectations. A direct way of knowing what a respondent believes and prefers is thus her own sincere confession. But this would pose again the threat of accommodation since the model would simply describe the situation at hand.

Second, there is a problem with informativeness since Bicchieri's approach seems to have a difficulty in providing a quantitative answer to the question: *How well* the model predicts? We would like to know about the predictive strength of a model by means of probabilities that it would assign to the predicted outcomes. If our predictions depend on a parameter K, we would like to know a systematic way to determine to *what degree* one is sensitive to a norm in a context. In default of such positive proposal the predictions of our model lacks in informativeness.

The third problem is conceptually prior. It is that of determining quantifiable *individual*-level variables statistically correlated with players' decisions. Bicchieri's model takes into account *types* of individuals, defined by their beliefs about the relevant norm in a type of context, their expectations about their opponents, and their norm-sensitivity (Bicchieri, personal communication). The model seems promising with predictions *across* groups of individuals where we can take group-membership as defining a type of individual. However, it is unclear how it would predict the behaviour of specific individuals *within* a group. It is, in fact, possible that individuals of the same type behave differently. For the way Bicchieri construes a type of individual allows that they can have different motives and preferences. Although variation between types of individuals can be accounted with economic, social, or cultural differences, the same does not apply within the same type of individuals.

Such a charge may seem unfair: the social sciences (and to a great extent also psychology) are less concerned with single individuals than with statistical tendencies across individuals. Also in physics it is customary to take into consideration the behaviour of many particles rather than a single one.

My point however can be read in a weaker way. Even if the behaviour of a single individual within a group may not be that interesting, collecting appropriate individual-level information systematically correlated to one's choice may be useful to make predictions more secure and informative *across* groups of individuals after statistical generalization. As usual in science, we may examine a sample of a certain class of individuals and then generalize our findings to the class as a whole. This information may be neurobiological. The second part of this work articulates such proposal.

III. NEUROBIOLOGICALLY-INFORMED MODELS

Twenty-one years after Güth *et al.*'s seminal work on the UG, a group of psychologists and neuroscientists led by Alan Sanfey analyzed subjects with functional magnetic resonance imaging (fMRI) as they played the UG. Sanfey *et al.*(2003) compared the brains of subjects responding to 50-50, 60-40 offers (in the experiment the total pie to split was $10), and 90-10, 80-20 offers. Three brain areas were found to be differentially activated: the Dorsolateral

prefrontal cortex, the anterior cingulate, and the anterior insula. For the purpose of this essay, the crucial finding is the correlation between the activation of the anterior insula with choice-behaviour. Specifically: First, the activation of the anterior insula was significantly correlated with the rate of rejection; Second, the magnitude of activation was "a function of the amount of money offered to participants" (p.1756). Third, the activation was "uniquely sensitive to the context" (Ibid.): there was greater activation for a 80-20 offer from human partners than the same offer from computers. From the first finding follows that whether players reject an offer or not may be *predicted* with a certain accuracy by the level of their insula activity.

Sanfey *et al*'s experiment is quite typical in the world of so-called "neuroeconomics": From a subject who confronts a choice problem D, a pair (d, *x*) is drawn. d is the alternative in D chosen by the subject; *x* is a vector of numbers representing the activities measured in various areas of the subject's brain during the period between the moment she is presented with the choice-set of D and the moment she chooses d. After statistical analysis, a correlation between choice-behaviour d, and the activity of a specific brain area may become apparent. Building upon this kind of correlation, it may be possible to integrate neurobiological variables into behavioural models. The predictive power of such models would be bound to the significance and robustness of the neurobiological finding. The remaining of this essay is a critical assessment of the comparative merits, and limits, of the neurobiological approach to modelling I wish to suggest. Security and informativeness of the predictions that this kind of models would give orientate my discussion as with Bicchieri's.

Security Sanfey *et al*'s experiment involved 19 American subjects that completed 10 rounds playing the game with a human partner, and 10 with a computer partner. Rubinstein (2006) and Harrison (2008) crisply point out that the inferences drawn from experiments of this kind are highly problematic because of their small, culturally homogeneous samples, the methodology of pooling subjects, and the heavy statistical machinery required to make raw neurological data amenable to analysis. How can one feel comfortable about building upon such weak grounds? I would like to make two points about Rubinstein's and Harrison's critiques.

First, their charges about samples and statistical analyses are empirical in nature. As such, there seems to be no principled reason why the methodological standards that enable us to warrant the inferences from a neuroeconomic experiment couldn't be raised. Whereas statistical flaws can be solved with careful data analysis made explicit and open to scrutiny, the problem with samples size can be faced in the long run, when many *replications* will tell us about the *robustness* of regularities like that lying below Sanfey *et al*'s set of data. I use 'replication', as opposed to 'exact repetition' (see Radder 1996; Guala 2005, Ch.2). A replication involves some (slight or radical) modification of the original experimental design. A nice replication of Sanfey *et al*'s experiment might involve subjects from different cultures. At the moment we can only speculate e.g. that in Israel, where lower offers are accepted more often (Roth *et al* 1991), we may expect less insula activity for the same sized offer than in the US, and less rejection. If the regularity firstly observed survives across replications, then we have reason to conclude that it is *robust*: It doesn't depend on details of the situation or on the particular statistical assumptions used to derive the results. Therefore, by emphasizing the evidential value of replications, and by noticing that we have no reason why we couldn't expect replications of neuroeconomic results, and higher statistical standards, the bite of these charges can be limited.

Second, a better way of evaluating Rubinstein's and Harrison's critiques is to ask whether the *evidential grounds* for neurobiological predictive claims may be more reliable than the grounds for psychological ones. Analyzing Bicchieri's model we discussed some of the

evidential problems with measurement and ascription of expectations to people. I argued that: First, because of the nature of social norms, secure predictions can be gained at the risk of accommodation; Second, the conditionality of preference to follow a social norm seems to open an epistemic gap between the actual motives, in a *token*-situation, that ultimately render a prediction correct\incorrect, and the information about *type*-situations and *type* of individual, necessary to ground the predictions of Bicchieri's model.

The grounds for predictions for a model which integrates neurobiological parameters consist in the empirical finding that two quantitative factors (e.g. insula activation and rate of rejections in the UG) are correlated in such a way that the behaviour of the one foreshadows the behaviour of the other with statistical significance. The security of the predictions is bond to the stability of the functional specialization of the target neurobiological structure *across brains and situations*. On the one hand, however, the activation of the target structure may not be sufficient to warrant the prediction since it may be overcome, or blocked, by the activities of other structures. However, the problem is not insoluble: We can identify *to what degree* the insula activity is sufficient to warrant the prediction of rejection in the UG by inquiring its relation\interaction with other brain structures. On the other hand, the activation of the structure may not be necessary. As before, the problem can be solved with further, accurate experimentation. In order to assess to what extent the insula activity is necessary in predicting the rejection rate, it would be interesting, e.g, to study the behaviour in the UG of neurological subjects with lesions in the insula; but, as far as I know, such experiment has not been set up yet.

A more serious threat to the reliability of the evidential grounds for neurobiologically-based predictions is the variability of brains across individuals. The problem, here, is not with token vs. type-situations, which would be bypassed by focussing on brain activations since brain activations are *already* sensitive to the token-situation where a person acts. The problem, instead, has primarily to do with the *plasticity* of the brain: Plasticity, or neuroplasticity, is the capacity of the brain to reorganize neural pathways based on new experiences. Both structure and function of developing brains are shaped, to some extent, both by the environment and by cultural experience. Differences in neural responses to a given stimulus are also likely to exist All brains, then, are different from one another to some degree. To the extent that different cultural norms and practices, and different environments exist across social groups differences in neural responses to a given stimulus are also likely to exist (Chiao & Ambady 2007). This likely variation poses at least one substantive problem to a model neurobiologically-informed. We might expect to find differences in the *type* of neural activity correlated to the same *type* of behaviour across subjects. Were this actually so, then the basis for our neurobiologically-informed model of human behaviour would be intrinsically unstable. However, we have reason to believe that the brain variance across groups is not so dramatic.

First, it is worth noticing that despite notable progress in describing cultural variation at the behavioural (and genetic) level, relatively little is known about how the structure and function of the human brain vary across subjects and cultures. Hence, sweeping claims are not justified by current evidence. Second, although there is no arguing that there is some kind of plasticity in our brains, there are clear limits on plasticity (Gazzaniga, Ivry, & Mangun 2002, ch.15). Different types of neuroplasticity occur during certain critical periods, notably: 1. During normal brain development when the brain begins to process sensory information, when it is still immature; 2. After brain injury, especially when the lesion involves the somatosensory system, to compensate for lost function; 3. Through adulthood underlying learning and memory mechanisms. After these critical periods, the central nervous system is characterized by a relative rigidity. Moreover, when it occurs, neuroplasticity seems to be particularly

significant at micro-levels of organization, e.g. at the synaptic level (Elman *et al*. 1996; Quartz & Sejnowski 1997). Here, plasticity involves subtle changes in the strength of synaptic connections between neurons which underlie the mechanisms of learning and memory. For these reasons, it appears unlikely that we find dramatic differences in the *type* of brain activity correlated to the same type of behaviour across adult individuals who have not suffered brain lesions – the target of brain imaging experiments indeed. Brains are shaped, to some extent, both by the environment and by cultural experiences. All brains, then, are different from one another to some degree. To the extent that different cultural norms and practices, and different environments exist across social groups, Given this relative stability in the brain activity underlying a given type of behaviour, more reliable grounds for a neurobiologically-informed predictive model are to be expected when we will have a general, detailed theory of the brain, and as we come to understand better how the instruments and techniques employed by neuroscientists mediate their observations (see Bechtel forthcoming, for the epistemology of evidence and instruments in neuroscience). At the moment, as with any other immature science, there is lots of work to do in the neurosciences: We know a good deal about brain neuroanatomy, and cytoarchitecture. We know less about the physiology and functioning of the brain than we know about other organs. Little is known about the mechanisms underlying the exercise of our cognitive abilities. Currently, the level of detail and quality of our knowledge about the brain is essentially constrained by the improvement of the technological apparatus used during experimentation. How and to what extent we will be able to integrate in a comprehensive theory of the brain the huge amount of data collected during experiments is still to be determined.

Informativeness Consider the following experiment: People are asked to say aloud, sincerely, their reasoning during an UG. The experimenter records everything, and at the end of the game comes up with a model based on the notes she has written down. Her model is highly successful since it fits exceptionally well the data: It gets the behaviour of those subjects right with a degree of accuracy of 99% -- 1% being the error rate due to her disattention. Should we regard that a good predictive model? Clearly not. For it is a mere record, or description, of what was going on during the UG. The evidence that that model purports to predict was already used in the construction of the model itself. The evidence is written into the model: This is a clear case where the model accommodates the evidence without predicting it. Now, what of the objection that the situation of a brain imaging experiment during an UG is analogous to that case? In the imaging experiment, so the charge would go, we would be just *witnessing* the brain activity correlated to the thought processes leading to decision; moreover, if mapped closely enough there may be a 100% correlation between patterns of brain activity and observed behaviour. Hence, the evidence would be written into the model, and, as before, the model would merely accommodate the evidence.

There are at least three main reasons why the case of a neurobiologically-informed model is unlike the one above. All three reasons draw on the special weight of independent assessment. First, the insula, the brain structure that correlates with rejection in the UG, is known to be involved in a variety of behaviours beside the one observed in the UG. Significant activations of the insula have been observed in response to disgusting gustatory and olfactory stimuli (Small *et al*. 2003; Zald & Pardo 2000), to the sight of disgusted facial expressions of others (Phillips *et al*. 1997). Following stimulation of the anterior insula, it has also been found that subjects report feeling nauseous (Krolak-Salmon *et al.* 2003). If we compare these findings with those of Sanfey *et al.*'s fMRI experiment during the UG, "repulse" is the feature common to all the correlated observed behaviour. This commonality would lend predictive

informativeness to our neurobiologically-informed model since it could be applied to situations *beside* the UG from which it was deduced to predict rejection. To this point it may be objected that the remark by a subject 'I am going to reject the offer' is also applicable to a wide variety of situations besides the UG, e.g. selling a car or renting an apartment. So this doesn't distinguish brain imaging studies from the imagined 'write down everything they say about their reasoning' study. Therefore, the point made above seems irrelevant. However, this objection seems to me to overlook a fundamental difference between the two cases. Suppose the ideal situation where a subject is sincere in her saying 'I am going to reject the offer', there is no physical impediment to the realization of her intention, *and* she has no other intentions that may override that one. Contrast this situation with another one where a subject's insula is significantly activated *and* there is no other activity in her brain that may override that one. Now, considering the first case, it seems to me that it is not an empirical discovery that when the subject states that she is going to reject the offer, she has the intention to do so, and she will do so *regardless* of the context of her utterance. Because her behaviour is *conceptually* bound up with the meaning of the predicate 'to be going to', the 'write down everything she says' study would simply describe the situation at hand, and would not provide us with any significant empirical discovery. Consider instead the second case. There we witness a neural event correlated with a person's repulse-behaviour. Unlike the former case, the correlation is not conceptual, but *inductive*: We first observe the correlation in a certain context; then, given numerous observations of the same correlation in a variety of contexts *beside* the first one we infer that a certain brain region is correlated to a certain type of behaviour. Unlike the first study, then, we have made an empirical discovery, a discovery revisable in light of further empirical evidence.

The second main reason why the two cases differ is the following. Courtesy of the vast array of technologies used in neuroscience we might use the data from one experiment to predict *independently* what is likely to happen in another. We might activate or disrupt the insula by using e.g. rTMS (repetitive magnetic stimulation, which temporarily disrupts brain activity in a target region of interest), predict differences in rates of rejection that would follow, and see whether we get them right. That, unfortunately, has not been done.

Finally, as emphasized by Worrall (2006), the interesting cases in which some evidence follows from a model, all involve a general underlying theory which itself stands in need of confirmation from evidence. Although, as indicated earlier, there is currently no comprehensive theory of the brain, there is no reason why after accumulating evidence such theory, which would specify the functional relationship of the insula with other brain structures and other observed behaviours, cannot be articulated.

Provided the model would not simply accommodate the data, and that it would give quantitative predictions by relying on correlations whose significance can be determined statistically, to what extent would it mark a substantive improvement in predicting behaviour? I would like to answer by borrowing an analogy from Rescher (1997, p.127). Take meteorological forecasting in Britain. The easiest route to predictive success seems to stick with the hypothesis that tomorrow's weather will be the same as today. In recent years, this prediction has had a rate of success of 75%. The British meteorological Office, courtesy of its high-tech resources, has been able to push this rate up to 85%. It doesn't look like a terrific improvement. Nonetheless, it is comparatively significant and further improvable in the level of detail. The same might be said for how a neurobiologically-informed model would fare in the UG: It would give comparatively more precise predictions sensitive to potentially significant details ignored by alternative models.

IV. CONCLUSION

Ariel Rubinstein (2008) asks us to assume the following scenario: 'We are able to map all brains onto a canonical brain. The functions of the different areas of the brain are crystal clear to us. The machines used in experiments are cheap enough that thousands of subjects can be experimented on. And finally the data are unambiguous and double-checked. What would be the potential role of brain studies in economic theory?' The suggestion this essay has defended is that the brain would matter to prediction.

A neurobiologically-informed model would not be in opposition to other models such as Bicchieri's. It would provide predictions more secure and informative by incorporating parameters neglected by current models, which may significantly correlate to human behaviour. When compared to models like Bicchieri's, because of the nature of its parameters, the potential advantage of a neurobiologically-informed model is that its grounds would be more easily warranted by independent lines of research without the risk of accommodation, and its predictions would be quantitatively accurate.

One way to summarize my argument is that by observing brain area *x* activation we may bypass such problems as the "specification problem" in models like Bicchieri's. The "problem of specification" is the problem of identifying the norms that prime a certain behaviour. For example, we may have a situation such as:

NORM A -> BRAIN ACTIVATION *x* -> UG BEHAVIOUR Y
NORM B -> BRAIN ACTIVATION *x* -> UG BEHAVIOUR Y

If different norms activate the same brain areas, and these are correlated with the same types of behaviour, then by observing the brain we would have an instrument for prediction more simple and effective. The importance of making good predictions is thus the reason to explore alternative models informed by neurobiological evidence. The hope is that these models will account for anomalies and make interesting new predictions.

However, the scenario just envisaged is ideal. Neuroscience is still an immature science, and the work to do before trying to integrate neurobiological parameters into a model of economic decision-making is a lot. This essay has tried to argue that this work would be important and worthwhile.[4]

BIBLIOGRAPHY

Bechtel W. (forthcoming) The epistemology of evidence in cognitive neuroscience. In R. Skipper Jr., C. Allen, R. A. Ankeny, C. F. Craver, L. Darden, G. Mikkelson, and R. Richardson (eds.)*, Philosophy and the Life Sciences: A Reader. Cambridge,* MA: MIT Press.

Bicchieri C. (2006) *The Grammar of Society: the Nature and Dynamics of Social Norms*, Cambridge University Press.

[4] The bulk of this paper was written at the London School of Economics (UK) during my M.Sc. Alex Voorhoeve, Francesco Guala provided generous comments on previous drafts. Cristina Bicchieri helped me very kindly by answering many questions I had about her theory. At an earlier stage I discussed some of the ideas contained here with Matteo Motterlini. A sincere thank you to all of them for their help. The usual disclaimers about the remaining errors apply.

Bicchieri C. & Chavez, A. (2008) Behaving as Expected: Public information and fairness norms, unpublished. URL: < http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1082264>, July 2008.

Bicchieri C. & Xiao, E. (2008) Do the Right Thing: But Only if Others Do So, *Journal of behavioural decision making*, 21: 118.

Bolton G., & Ockenfels, A. (2000) ERC: A Theory of Equity, Reciprocity and Cooperation, *American Economic Review* 90: 16693.

Bradley R. (unpublished) 'Attributing Mental States'.

Camerer C. (2003) *Behavioral Game Theory: Experiments on Strategic Interaction,* Princeton University Press.

Camerer C. (2007) Neuroeconomics: Using Neuroscience to Make Economics Predictions. *Economic Journal*, 117: C2642.

Chiao J.Y. & Ambady N. (2007) Cultural neuroscience: Parsing universality and diversity across levels of analysis. In Kitayama, S. and Cohen, D. (Eds.) *Handbook of Cultural Psychology*, Guilford Press, NY.

Elman J. L., Bates E. A., Johnson, M. H., KarmiloffSmith, A., Parisi, D. & Plunkett, K. (1996) *Rethinking Innateness*. MIT Press.

Fehr E. & Schmidt, K. (1999) A Theory of Fairness, Competition and Cooperation, *Quarterly Journal of Economics* 114: 81768.

Forster M. (2008) Prediction, in S. Psillos & M. Curd (eds.), *Routledge Companion to Philosophy of Science*,London: Routledge.

Gazzaniga M. Ivry R., Mangun G., (2002), *Cognitive Neuroscience*, W.W. Norton & Co. Inc., New York.

Guala F. (2005) *The Methodology of Experimental Economics,* New York: Cambridge University Press.

Guala F. (2008) Paradigmatic Experiments: The Ultimatum Game. *Philosophy of Science*, forthcoming.

Güth W., Schmittberger, R., & Schwarze, B. (1982) An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization, 3*(4), 367-388.

Harrison G. (2008) Neuroeconomics - A Critical Reconsideration. *Economics and Philosophy*, 24: 303 – 344.

Henrich J., Boyd R., Bowles, S., Camerer C., Fehr E., & Gintis H. (eds. 2004) *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen SmallScale Societies.* Oxford: Oxford University Press.

Hoffman E., McCabe K., Shachat, K., & Smith, V. (1994) Preferences, property rights and anonymity in bargaining games. *Games and Economic Behavior*, 7:346-380.

Koch C. & Ullman S. (1985) Shifts in selective visual attention: towards the underlying neural circuitry, *Human Neurobiology* 4:219-227.

Krolak-Salmon P. *et al.* (2003) An attention modulated response to disgust in human ventral anterior insula. *Ann. Neurol.* 53, 446–453.

Laudan L. (1990) De-Mystifying Underdetermination, in W. Savage, ed., *Scientific Theories*. Pp. 26797. Minneapolis: University of Minnesota Press.

Pablo B.-G., Ramón, C.-R., & Almudena, D. (2006) Si él lo necesita: Gypsy fairness in Vallecas. *Experimental Economics, V9*(3), 253-264.

Phillips M.L. *et al*. (1997) A specific neural substrate for perceiving facial expressions of disgust. *Nature* 389, 495–498.

Quartz S.R. & Sejnowski, T.J. (1997) The neural basis of cognitive development: A constructivist manifesto. *Behavioural and Brain Sciences* 20: 537–596

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

Radder H. (1996) *In and about the World: Philosophical Studies of Science and Technology*, Albany: State University of New York Press.

Rescher N. (1997) *Predicting the Future*. Albany: State University of New York Press.

Roth A., Prasnikar V., Okuno-Fujiwara, M. & Zamir, S. (1991) Bargaining and Market Behavior in Jerusalem, Lubljana, Pittsburgh and Tokyo: An Experimental Study, *American Economic Review* 81: 1068-1095.

Rubinstein A. (2006).”Comments on Behavioral Economics”, in *Advances in Economic Theory* (2005 World Congress of the Econometric Society), Edited by R. Blundell, W.K. Newey and T. Persson, Cambridge University Press, 2006, vol II, 246-254

Rubinstein A. (2008) Comments on Neuroeconomics, *Economics and Philosophy*, 24: 485 – 494.

Sally D. & Hill E. (2006) The development of interpersonal strategy: Autism, theory-of-mind, cooperation and fairness, *Journal of Economic Psychology* 27(1):73-97.

Sanfey A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003) The neural basis of economic decision-making in the ultimatum game. *Science, 300*(5626), 1755-1758.

Small D.M. *et al*. (2003) Dissociation of neural representation of intensity and affective valuation in human gestation. *Neuron* 39, 701–711.

Solnick S.J. & Schweitzer, M.E. (1999) The influence of physical attractiveness and gender on ultimatum game decisions. *Organizational Behavior and Human Decision Processes* 79(3): 199-215

Thaler R. H. (1988) Anomalies: The Ultimatum Game, *Journal of Economic Perspectives*, 2, pp. 195–206.

Worrall J. (2006) History and Theory-Confirmation in J. Worrall and C. Cheyne (eds) *Rationality and Reality: Conversations with Alan Musgrave*. Pp. 31-61 Kluwer Academic Publishers, 2006.

Zald D.H. & Pardo, J.V. (2000) Functional neuroimaging of the olfactory system in humans. *Int. J. Psychophysiol*. 36, 165–181.