No Brainer Predictions in the Ultimatum Game

Matteo Colombo

MSc Philosophy and History of Science, 2008

Observations of the specific regions in the brain that are active when behaviour is observed can be very useful for a neuroscientist. But what could it add to our understanding of economic behaviour? My suggestion is that the brain matters to the prediction of economic behaviour. The goal of this essay is to argue for this claim.

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 realisation of the methodological ideal described more than fifty years ago by Milton Friedman. In his famous essay on "The Methodology of Positive Economics" (Friedman, 1953), Friedman strongly advocates predictive success as a means of judging a "positive" scientific theory. In his own words:

> 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)

In order to defend my thesis, I shall focus on the Ultimatum Game (UG). I shall try to show that a model which incorporates neurobiological variables fares way better in predicting the behaviour of the players in the UG than alternative models.

The structure of my essay is as follows. In the first section, I shall describe the UG, its game theoretic analysis, and how people have been observed to play the game. In the second section, I shall introduce the concept of "enriched models". These models try to catch the actual behaviour of players in the UG by appealing to such concepts as "fairness", "warm glow", "envy", "social norms", and so forth. I shall analyse how the model defended by Bicchieri (Bicchieri, 2006) based on social norms works when it is called to account for the UG. I shall argue that this kind of account is vague, and tends to accommodate facts and predictions *ad hoc*. In the third section I shall argue that if we want accurate predictions, and we want to appeal less to auxiliary hypotheses, then we have to point to quantifiable biological variables which have a large influence on behaviour and are underweighted or ignored in both game theoretic, and social-norms based models. I shall devote most of the section to reply to one important objection to my claim.

Game Theoretic Predictions in the Ultimatum Game

Game Theory is a collection of rigorous models attempting to understand situations in which decision-makers interact with one another. Classical 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 his or her own payoff unilaterally. Strategic bargaining behaviour is one of the concerns of Game Theory.

One of the simplest forms of bargaining in which outcomes are predictable is the Ultimatum (or "take-it-or-leave-it") Game. In the Ultimatum Game, two players interact once, anonymously. The first player proposes how to divide a sum of money between themselves. The second player has the option of accepting or rejecting. If the offer is accepted, the sum is split as proposed. If it is rejected, neither player receives anything.

The prediction of Game Theory, the subgame perfect equilibrium, is that for any positive amount offered by the proposer, the second player knows that he or she faces a choice between gaining nothing (if he or she refuses the offer) or something (if he or she accepts). The proposer should therefore always offer the minimum possible split to player two, who should always accept. Notice that since the game is one-shot and anonymous, reciprocation is not an issue.

Assumptions of rationality and self-interest underlie this prediction. Given that the preferences of the players are self-regarding, they will choose the largest payoff possible without caring for the outcomes and behaviour of the other player. Given that the players are rational, they will form correct beliefs about each other's behaviour; and, given their beliefs, the players will choose those actions that best satisfy their preferences.

Accordingly, player I, who makes the offer, infers that any positive offer is better than zero for player II, who should therefore accept it. As a consequence, player I should offer the smallest sum of money possible, in order to keep as much money as possible, and player II should accept this offer, which is better than nothing.

This game theoretic prediction is at odds with the actual observed behaviour of people playing the UG. The experimental literature on the UG indicates a robust behavioural pattern. Since the first experiment which studied the UG (Güth, Schmittberger, and Schwarze, 1982), the UG has been studied in many diverse settings in which different parameters of the game have been modified. Across cultures, sex, age and amounts of money provided, the offers are typically around 50% of the total sum. "Low" offers of around 20% out of the total are very likely to be refused (Oosterbeek et al., 2004; Samuelson, 2005). In reality, the players of the UG are generally less selfish and less rational than the game theoretic model predicts. This behavioural pattern begs two questions:

Firstly, why do people tend not to play according to the game theoretic prediction of the UG, and instead naturally coordinate on or around 50-50 splits? Secondly, is there an alternative model to the standard game theoretic one capable of yielding accurate predictions of the behaviour of the players in the UG? Obviously, an answer to the first question – which counts as a new model for accounting the behaviour of the people in the UG – will be at loss whenever the new model fails to predict well.

Bicchieri's Social Norms model

The most natural move to account for those two questions is to enrich the standard game-theoretic model by incorporating social parameters in the utility function of the players (Fehr and Schmidt, 2003, for a review). In this section, I shall analyse one of the most recent and promising enriched models, that of Bicchieri (Bicchieri, 2006). However, I shall argue that this model is not precise enough to yield "risky", falsifiable predictions.

In the face of the "anomalous"[1] behaviour of players in the UG, utility functions incorporating such parameters as 'altruism', 'equality', or 'fairness' are becoming increasingly common in the game-theoretic literature. A common feature of these enriched models is the preservation of the logical

[1] The first of a series of articles on Anomalies in the *Journal of Economic Perspectives* was on the UG by Richard Thaler in 1988. Borrowing his words: an anomaly is 'an empirical result … if implausible assumptions are necessary to explain it within the [rational choice] paradigm'.

framework of expected utility theory: they do not reject the rationality assumptions, they just point to the maximisation of a non-classical utility function whose empirical substance is provided by the new parameters.

Cristina Bicchieri (Bicchieri, 2006) develops one of the most interesting of these models, building upon the notion of social norms. According to Bicchieri, social norms are behavioural rules which exist if a sufficiently large number of people in a population know that in certain situations the behavioural rule exists, and who conform to this rule whenever they believe that:

> 1. Enough other people are following the rule in those kind of situations.
> 2. Enough other people expect them to conform to the rule, and might sanction one if one does not conform (Bicchieri 2006, ch.1).

The important point to notice here is that what makes a norm 'social' is the conditionality of preference: one follows a social norm when she is "pretty sure" that a sufficient number of members of her society will do the same. In other words, Bicchieri acknowledges in her model the crucial role played by beliefs – to which I referred by the expression "pretty sure" in the previous sentence – in sustaining social norms. Social norms, in their turn, induce one to prefer to behave in a certain way in a context of a certain type. Norms map contexts to beliefs and preferences. I shall not dig into the details of this definition, instead I shall concentrate on how the model is supposed to account for the anomaly observed in the UG.

Imagine this slightly different form of an anonymous, one-shot, UG. Before the respondent 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, and so on until a point is reached where she would accept anything higher offered to her. If player I's offer is below her acceptable-range, her response would count as a rejection. Player I's proposal is finally revealed. Suppose also that the two players belong to the same population where 'sharing' is a well-established social norm (e.g. hospitality and aiding others are strong obligations). What would Bicchieri's model predict in this situation?

We would expect that the proposer will conform to 'sharing', and therefore she will propose an equal, 50-50, split. So far, so good. But let us focus on the respondent. If she expects that the proposer will conform to 'sharing' in that context, and she is also expected to follow sharing in that context, and she doesn't have any other (contextual) reason not to prefer 'sharing', then she will be willing to reject the other's offer at a cost to herself if and only if the proposer's offer does not conform to 'sharing' (i.e. her expectation that the proposer will conform to 'sharing' is not satisfied). Therefore, we would expect that the minimum value that player II will accept will be not far from 50-50.

Contrary to this prediction, a recent study (Pablo et al, 2006) with Gypsy people in Vallecas, Madrid, showed something completely different. Although 97% of proposers offered an equal split, as responders the gypsies were willing to accept completely unfair offers: indeed, acceptance of the zero offer was the modal value!

Bicchieri's model seems therefore to fare poorly in predicting the observed behaviour of the respondents in the Gypsy experiment. When 'sharing' is broken, we might expect a shift in the expectations of the respondent. Pablo et al (2006) argues that the respondent may be willing to accept the most disadvantageous distribution because she believes that "if another subject does not respect the 'sharing rule' it is because he or she is more needy" (Pablo et al. 2006, p. 262). But why do we not see a different shift? The most natural change in the respondent's expectations may be: 'If another subject does not respect 'sharing' it is because she is stingy, so I shall sanction her for this inappropriate behaviour by refusing her offer'. The point is that, granted that a certain social norm exists in a population, once the norm is broken we don't know the direction of the shift of a subject's expectations. Inappropriate behaviours, in other words, leave space both to predictions that assume a hyper-altruistic response, and to predictions that assume a sanctioning response. In default of some constraints that lead us to the identification of the content of a social norm that is applied in a certain situation, Bicchieri's model may always accommodate facts and predictions by some ad hoc assumption about how a certain norm has mapped the context to beliefs and preferences. If I am right, then the model yields either vague predictions or precise predictions, but via *ad hoc* assumptions.

Incorporating neurobiological results

Twenty-one years later Güth et al's seminal work on the UG, a group of psychologists and neuroscientists led by Alan Sanfey analysed subjects with functional magnetic resonance imaging as they played the UG. Drawing upon the results of that experiment, I shall conclude this essay by arguing that if we want falsifiable models that give good predictions of human behaviour in the UG, then we should try to build biological models.

Sanfey et al. (2003) compared the brains of subjects responding to 50-50, 60-40 offers (in Sanfey et al's experiment the total pie split is $10), and 90-10, 80-20 offers found that three brain areas are differentially activated: the Dorsolateral Prefrontal Cortex (DLPFC), the Anterior Cingulate (ACC), and the Anterior Insula Cortex. For the purpose of this essay, the significant finding is the correlation between Anterior Insula activity and choice behaviour. They found that the activation of the Anterior Insula was positively correlated with rejection. The magnitude of activation was "a function of the amount of money offered to participant" (Sanfey et al, 2003, p. 1756), namely: there was greater activation for a 90-10 offer than an 80-20 offer. The activation was also "uniquely sensitive to the context" (Sanfey et al, 2003, p. 1756), namely: there was greater activation for a 80-20 offer from human partners than the same offer from computer partners. Crucially, whether players reject an offer or not could be predicted rather reliably by the level of their insula activity.

It seems that this is a promising beginning for a neurobiological model of the UG. The model is falsifiable since we know in which circumstances the model would give a wrong prediction, namely: when the player will accept the offer, whereas her Insula Cortex is significantly activated – it would be interesting indeed to study the behaviour in the UG of neurological subjects with a lesion to the Insula Cortex, from what I know such an experiment has not been set up yet. Of course, the reliability of the prediction can be improved by adding more variables after further experimentation – for example, both van't Wout, Sanfey et al (2005) and Knoch, Fehr et al (2006) using repetitive TMS (magnetic stimulation that temporarily disrupts brain activity) to deactivate the DLPFC when people received offers found an increase in the acceptances of low offers. The precision of the variables (and of the correlation coefficients) is essentially constrained by the improvement of the technological apparatus used during experimentation.

One natural objection to my advocacy of neurobiological models is that they do not explain why people behave the way they do. One way to rephrase this charge is to point out that a neurobiological model lacks a vocabulary that enables one to grasp what functions the brain activity is underpinning: a vocabulary that makes transparent why one player is going to reject a 90-10 offer. It is true that talking about 'Anterior Insula activity' instead of preferences, beliefs and desires, is not the familiar practice of interpreting one's behaviour. Nonetheless, this does not imply anything with regard to the predictive power of the anterior insula activity. The two issues are conceptually separate. Whether a neurobiological model, whose parameters refer to neurobiological structures, gives better predictions than a norm-based model, whose parameters refer to preferences and beliefs, is one issue. And it's (largely) an empirical one. Whether the neurobiological model also provides better explanations of why we behave the way we do than the norm-based models is a different issue, which involves complex conceptual and empirical problems concerning causality, reductionism, intentionality, and so forth. Therefore, the charge misses its target, since my claim is only about testability, and the reliability of predictions.

The critics, at this point, may appeal to Hempel's Explanation/Prediction Symmetry thesis in order to defend the conceptual connection between prediction and explanation (Hempel, 1965, pp. 366-376). This thesis states that every adequate explanation is potentially a prediction, and every adequate prediction is potentially an explanation. Hempel's argument for his thesis crucially rests on his (controversial) deductive-nomological (D-N) model of explanation. For my purpose, here, it suffices to show that the second sub-thesis – that from prediction to explanation – is false to justify why the charge misses its target. Now, Ptolemaic astronomy, where each planet moves in a small circular orbit, and the centre of the small circle moves in a large circle around the Earth as the centre, is highly successful in predicting changing positions of the stars and planets, and until recently it was still in use for purposes of navigation. Yet, it does not satisfy the conditions Hempel imposes on the explanans of a D-N explanation (Hempel & Oppenheim, 1948). For Ptolemaic astronomy rests on false assumptions (e.g. it places the Earth in the centre of the universe). So the half of the symmetry thesis from prediction to explanation is false. And, consequently, prediction and explanation are not symmetric: A model can yield good predictions without necessarily being also explanatory.

At this point, if I am right, the conclusion of my argument should be clear: If the goal of a model of human behaviour is to give good predictions about important classes of choices – of which the UG bargaining is an instance, then the best models we can build are neurobiological.

References

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

Fehr, E. & Schmidt, K. M. (2003) "Theories of Fairness and Reciprocity: Evidence and Economic Applications", in Dewatripont, M. et al. eds., *Advances in Economic Theory, Eighth World Congress of the Econometric Society,* Vol. 1, Cambridge: Cambridge University Press, pp. 208-257.

Güth, W., Schmittberger, R. & Schwarze, B. (1982) "An Experimental Analysis of Ultimatum Bargaining", *Journal of Economic Behavior and Organization*, 3(4), pp. 367-388.

Hempel, C. G. (1965). *Aspects of Scientific Explanation*, New York: Free Press.

Hempel, C. G. & Oppenheim, P. (1948) *"Studies in the Logic of Explanation"* reprinted in Hempel (1965), pp. 291-295.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V. & Fehr E. (2006) "Diminishing Reciprocal Fairness by Disrupting the Right Prefrontal Cortex", *Science*: 314 pp. 829-832.

Oosterbeek, H. & van de Kuilen, G. (2004) "Differences in Ultimatum Game Experiments: Evidence from a Meta-analysis." *Experimental Economics*, 7, pp. 171-188.

Pablo, B. G., Ramón, C. R., & Almudena, D. (2006) "Si él lo necesita: Gypsy Fairness in Vallecas." *Experimental Economics*, Vol.9(3), pp. 253-264.

Samuelson, L. (2005) "Economic Theory and Experimental Economics", *Journal of Economic Literature*, 43, pp. 65-107.

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, pp. 1755-1758.

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

Van’t Wout, M., Kahn, R. S., Sanfey, A. G. & Aleman, A. (2005) “rTMS over the Right Dorsolateral Prefrontal Cortex Affects Strategic Decision Making”, *NeuroReport,* 16, pp. 1849-1852.