Social Intelligence in Games

Comment

by

GERD GIGERENZER AND RICHARD MCELREATH

1 Introduction

In his essay "An anthropologist on Mars," the neurologist Oliver Sacks describes his encounter with Temple Grandin, a highly remarkable autistic person who holds a Ph.D. and teaches at Colorado State University. Like other autists, Temple Grandin seems to be largely devoid of social intelligence. When Temple watches the social games normal people play, the social rituals they follow, and their powerful emotions that can change from love today to hate tomorrow, she feels, as she put it, like an anthropologist on Mars. As SACKS [1995, p. 270] describes her, "Lacking it [social intelligence], she has instead to 'compute' others' intentions and states of mind, to try to make algorithmic, explicit, what for the rest of us is second nature."

Oliver Sacks studies people who have neurological disorders. Economists and cognitive psychologists usually do not; they study normal people. Nevertheless, there is a similarity between the autistic personality and the rational models that many economists and some psychologists embrace. If by "rational" we mean conforming to the classical expected-utility model, or backward induction, then we have an "autistic" conception of human rationality. Just like Temple Grandin, *homo economicus* – defined in that way – lacks social intelligence and is puzzled by the strange behavior of normal people.

Experimental economists rely on simple games to elicit behavior that is blatantly at odds with the assumption that players are attempting to maximize their own expected utility. Consensus is growing that the rational-choice model is descriptively wrong. This message has been hammered home by two groups of researchers: (i) experimental economists studying social games such as the ultimatum game, and (ii) psychologists demonstrating that individual people's judgments and decisions violate the axioms or consequences of the expected-utility calculus, the laws of probability, or logic. Real humans are capable of acting and punishing altruistically and of valuing fairness, and are guided by emotions, obligations, feelings of guilt, and other moral sentiments. What to do with a model of rational man who is socially unintelligent, that is, a model that is descriptively wrong?

One reaction is just to go on, close one's eyes, and ignore this research. A second reaction is to engage in a "repair program": One sticks with the expected-utility

Journal of Institutional and Theoretical Economics
JITE 159 (2003), 188–194 © 2003 Mohr Siebeck – ISSN 0932-4569

framework, but adds one variable that has been experimentally shown to influence choice, such as regret. Or one tinkers with the utility and probability functions, and changes the shape of the curves to accommodate other experimental results, as in prospect theory. There is a third way, a radical departure from the repair program: to dispense with the expected-utility framework entirely, and start afresh on the basis of empirically rooted knowledge about the human mind and its capabilities. This, in my view, is the program underlying the work summarized in FALK, FEHR, AND FISCHBACHER [2003].[1] For instance, in the book *Bounded Rationality: The Adaptive Toolbox*, Reinhard Selten and one of us formulated as our goal "to promote bounded rationality as the key to understanding how actual people make decisions without utilities and probabilities" (GIGERENZER AND SELTEN [2001, p. i]). Quantitative probabilities, utilities, and optimization do not play a large role in the actual processes of the human mind, whereas fast and frugal heuristic processes, such as emotions, name recognition, aspiration levels, imitation learning, limited search, stopping rules, and one-reason decision-making, do (GIGERENZER, TODD, AND THE ABC RESEARCH GROUP [1999]).

2 *Do Intentions or Only Consequences Matter?*

In three experiments, FALK, FEHR, AND FISCHBACHER [2003] use the ultimatum game as a window on the actual behaviors, motivations, and cognitions of people in bargaining situations. Specifically, they study three possible reasons why people reject or do not reject a given offer: the roles of intention, spite, and reputation. We applaud this experimental approach, because it brings social intelligence into a previously nonsocial theory of rationality in games.

Standard utility theory, so the authors write, "assumes that the utility of an action is determined solely by the consequences of an action and not by the intention behind it" (FALK, FEHR, AND FISCHBACHER [2003, p. 178]). By restricting the options of the proposer to only two possible offers, so that intentions and behavior are no longer coupled, the first experiment ingeniously demonstrates that intentions and not only outcomes count. We are quite fond of experimental economics, but as you can infer from the analogy with autism, we are impatient with the slow move towards understanding the Martians, that is, us humans.

It is both interesting and frustrating that economists still debate whether or not inferred intentions matter to people. A glance into the empirical literature on moral development in children or on notions of justice in adults would have revealed dozens or hundreds of studies that have experimentally demonstrated that even very young

[1] It was surprising to hear Armin Falk saying in the discussion that the work on fairness, intention, spite, and other forms of social intelligence is a straightforward extension of expected-utility maximization. To us, that seems to be a strategic move to reduce conflict with hard-nosed utility maximizers; Herbert Simon and Allan Newell used the same move many years ago when they claimed that their new conception of the mind as computer was in the tradition of the earlier *Gestalt* psychology (GIGERENZER [2000, ch. 2]).

children already distinguish between bad outcomes that a person caused intentionally and those that were caused unintentionally or because there was no choice. Children punish differently in these situations, as do courts in countries where the legal rules distinguish between intended and unintended harm. In Kohlberg's theory of moral development, the most influential theory on this topic, the distinction between consequences and intentions of an action arises in the human mind at Stage 2, which occurs around the ages of five to six (KEASY [1978]). Note that Stage 2 is not the highest level of moral development in his theory; there are six stages in total. Conducted in a wide variety of situations, these experimental studies in psychology have already provided the empirical evidence that Falk et al.'s first experiment found in a specific social situation, the ultimatum game. But this experimental literature is not discussed by Falk et al. This omission is part of a game played by most social scientists, and not only economists: Read and cite only what your peers in your journals have written, and ignore the rest. It is an "identification" game that is played all too frequently, and we are all guilty of participating in it.

Imagine playing the ultimatum game with a computer, and assume that you do not attribute intentions to it. The same humans who reject low offers from other humans would rarely reject low offers from computers. BLOUNT [1995] performed an ultimatum-game experiment in which responders played both against a computer making random offers and against human subjects. She found that people rejected low offers from other subjects, as is the usual result, but very rarely rejected offers from the computer. The issue is not low offers alone; it is the intentions behind them that make you feel exploited, angry, and ready to retaliate.

There are theories, such as FEHR AND SCHMIDT [1999] and BOLTON AND OCKENFELS [2000], that assume that intentions of fairness are behaviorally irrelevant. Falk et al.'s experiment using the ultimatum game provides convincing experimental evidence against that position (see also MCCABE, SMITH, AND LEPORE [2000]). This is a definite step towards helping the anthropologists understand the Martians. However, in this case the anthropologists – i.e., the economists – are themselves Martians. In addition, there is already a large amount of literature available showing that inferred intentions matter. One therefore wonders why this particular experiment needed to be conducted: Could it really be that economists did not know the answer all along?

3 On the Attribution of Spiteful Behavior

Falk et al. present inferred intentions as an explanation for why responders reject or do not reject low offers, and they add spite and reputation to this list. Spite is, interestingly, another social behavior that autists can hardly understand, just like deception and lying (GRANDIN [1995]). However, spitefulness is rarely a personality trait, but rather a behavior that emerges in a specific class of social interactions, such as competitive interactions in which coming out first, rather than maximizing one's

individual gain, matters. Moreover, it is not clear whether the experiment succeeds in demonstrating the operation of spite. The authors may infer the existence of malevolence far too quickly, or else the short description of this experiment leaves out relevant information. In a strict sense, they seem to consider spite to be any situation in which the ego lowers its own payoff to reduce the payoff of the other individual. This is the same as the definition used in evolutionary biology.

Here are two alternative interpretations. First, from the point of view of the responder, it is of advantage if the proposer offers the choice *Y* rather than *X* (see Figure 2 in FALK, FEHR, AND FISCHBACHER [2003]), because then the responder can get 15 points rather than 12 points. If the responder becomes angry when the proposer nevertheless offers choice *X*, the responder may think of punishing the proposer by responding with the option [0;10], which is interpreted as indicating spite. However, the behavior of the responder can equally be interpreted as retaliation: You did not provide me with my best outcome, so I will do the same to you. Consistent with this alternative interpretation is the observed result that when the proposer offers the choice *Y*, this kind of retaliation is rarely observed. The experiment, however, does not seem to be able to distinguish between the two interpretations. The rejoinder to this alternative explanation is that the experiment was a one-shot game and thus retaliation in order to build up reputation or change the other's behavior makes no sense. We address this rejoinder in the next section.

There is a second alternative interpretation that also does not need to assume a spiteful personality. If the responder did not consider the final 10 units of payoff to be part of the fairness calculation (and why should they, since the responder gets this portion of the payoff regardless of what the proposer does?), then rejection does not have to indicate spite. Instead, it can indicate the same motivations that led to rejections in the first experiment, namely that the proposer could have offered a fair [5;5] split, but instead offered the [8;2] distribution. Just as in Experiment 1, some responders rejected the [8;2] offer. The possible mistake here is in taking for granted that the subjects will sum up the payoffs in the same way as the experimenters, i.e. that there is only one way for a subject to understand the payoffs. The nature of payoffs can be ambiguous to the subjects in such experiments. Experimenters have a normative theory about what aspects of the payoffs matter, but the subjects often do not have such a theory, or they construct a different one. This makes the results of the second experiment much harder to interpret than the authors suggest.

4 Are One-Shot Games Always One-Shot?

The distinction between the experimenter's and the participant's theory of what matters in a given experiment can be of major importance. Georg Elwert once tried to convince a group of Ayizo peasants in Benin, West Africa, to play the ultimatum game and failed.[2] The Ayizo were concerned about features of the game: "Who is the other?" "What will people say if I behave this way?" and "I don't believe that this

[2] Personal communication, June 2002.

will be a one-shot game." For these peasants, the importance of the status and the power of the other player, as well as concern with the social acceptability of one's actions made them refuse to play a game that aims at implementing anonymity. This distinction between the experimenter's and the participant's views of what matters in an experiment is also relevant for the distinction between one-shot and repeated games with Western participants. We will make two points. The first has been made before; the second, we believe, is new.

4.1 Heuristics Adapted to the Evolutionary Past

The first point is that people have a tendency to understand a new situation by analogy with a prototype entrenched in their mind. Such a prototype can be a situation that was typical in the history of human evolution, that is, in our Pleistocene past. The outcomes of economics experiments often suggest that people behave as if they lived in a world in which the games were repeated:[3] They reject unfair offers, punish altruistically, seek reputation, and otherwise act in ways consistent with strategies that evolve in repeated games. Some evolutionary psychologists' preferred explanation is a version of the disequilibrium argument: People cannot tell the difference between finite and repeated games (at some level of cognition at least), since there were very few finite games in our Pleistocene past. A similar point is made when subjects behave as if the encounter were not really anonymous. PINKER [1997, p. 42] writes that "our brains are not wired to cope with anonymous crowds ... and other newcomers to the human experience." There was little anonymity in the environment of evolutionary adaptiveness, so the story goes, and thus people may not take advantage of anonymity.

In principle, these sorts of arguments are cogent. But it cannot be that simple. People *are* able to tell at least some situations of anonymity from others: For instance, people in restrooms, especially men, wash their hands more when other people are present (PEDERSEN, KEITHLY, AND BRADY [1986]). And, as Falk et al.'s third experiment shows, people often behave differently in repeated games than they do in one-shot games. There may have been a sufficient number of finite and anonymous interactions in the Pleistocene past, as well. What is likely happening in these experiments is that a specific set of cues are responsible for alerting subjects to the duration of the game setting. In some repeat experiments, these cues are weak or absent. In others, they are strong. The default assumption may well be a long time horizon, or perhaps the opposite. The default assumption may even vary as a function of culture, occupation, or experience. More attention to the specific cues that subjects use to select among strategies would help to demystify this issue.

[3] The crucial distinction, of course, is not between one-shot and repeated play, but between low and high probabilities of continued interaction. We will use "finite" and "repeated" to refer to variation along this dimension.

4.2 Negotiating the Rules of the Game

But there is a second phenomenon that may account for some one-shot situations in which people behave as if there were more trials to come. The argument here is not that people in the experimental situation carry over behavioral strategies that are adapted to a past world where bargaining was repeated and nonanonymous, and where reputation could be formed. The argument is that in the real world, agents can do more than follow the rules of a game; they can change the rules. That is, a one-shot game can actually be changed into a repeated game. Here is an anecdote to illustrate this point (GIGERENZER [1996, p. 324]):

> "A small town in Wales has a town fool. He once was offered the choice between a pound and a shilling, and he took the shilling. The proposer was quite amused. Someone else tried it again, and the fool again took the shilling. People came from everywhere to witness this phenomenon. Over and over, they offered him a choice between a pound and a shilling. He always took the shilling."

The town fool's choice in the original one-shot game looks irrational. But here we have a situation in which a particular choice increases the probability of getting to choose again. The fool's unusual choice changed the one-shot game into a repeated game. More generally, in a world that offers greater uncertainty than a well-defined game, the rules are not always completely fixed, and – if one is smart or lucky enough – they can even be changed. The crucial point is the existence of uncertainty and the possibility of change. As Benjamin Franklin reminded us in 1789, "in this world nothing is certain, except death and taxes." Part of social intelligence is the ability to change the rules, not only to perform well enough within the rules.

Why are attempts to change rules a point worth considering? Unlike the ultimatum game, typical real-world bargaining – with the exception of activities such as auctions – does not have fixed rules. Buyers and sellers, employers and unions often negotiate the rules in the process, along with the outcome. In the language of economic models, aspects of the rules are endogenous. The introduction of social intelligence into economic analysis has methodological consequences. If people are socially intelligent, they may behave in ways that are inappropriate to one-shot games, and even try to change the nature of the game.

4.3 Beyond Homo Autisticus

To conclude, we return to the analogy between rational choice and autism. Autism is a mental disease. People with autism have difficulties in judging the intentions of others, and they do not reckon with spiteful behavior, deception, or lying. FALK, FEHR, AND FISCHBACHER [2003] conclude that what they call "nonstandard preferences"– fairness intentions, spite, and reputation – influence people's behavior in the ultimatum game. We fully agree that these forms of social intelligence matter. But we also feel that economic theory deserves a more realistic and radical point of

departure than that of a surprised autist who studies social behavior. More courage may be needed.

References

BLOUNT, S. [1995], "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences," *Organizational Behavior and Human Decision Processes*, 63, 131–144.

BOLTON, G. E., AND A. OCKENFELS [2000], "ERC – A Theory of Equity, Reciprocity and Competition," *American Economic Review*, 90, 166–193.

FALK, A., E. FEHR, AND U. FISCHBACHER [2003], "Reasons for Conflict: Lessons from Bargaining Experiments," *Journal of Institutional and Theoretical Economics,* 159, 171–187.

FEHR, E., AND K. SCHMIDT [1999], "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114, 817–868.

GIGERENZER, G. [1996], "Rationality: Why Social Context Matters," pp. 319–346 in: P. B. Baltes, and U. M. Staudinger (eds.), *Interactive Minds: Life-Span Perspectives on the Social Foundation of Cognition*, Cambridge University Press: Cambridge.

– – [2000], *Adaptive Thinking: Rationality in the Real World*, Oxford University Press: New York.

– – AND R. SELTEN (eds.) [2001], *Bounded Rationality: The Adaptive Toolbox*, MIT Press: Cambridge, MA.

– –, P. TODD, AND THE ABC RESEARCH GROUP [1999], *Simple Heuristics that Make us Smart*, Oxford University Press: New York.

GRANDIN, T. [1995], *Thinking in Pictures*, Vintage Books: New York.

KEASEY, C. B. [1978], "Children's Developing Awareness and Usage of Intentionality and Motives," pp. 219–260 in: C. B. Keasey (ed.), *Nebraska Symposium on Motivation*, Vol. 25, University of Nebraska Press: Lincoln, NE.

MCCABE, K. A., V. L. SMITH, AND M. LEPORE [2000], "Intentionality Detection and 'Mind-reading': Why does Game Form Matter?" *Proceedings of the National Academy of Sciences of the United States of America*, 97, 4404–4409.

PEDERSEN, D. M., S. KEITHLY, AND K. BRADY [1986], "Effects of an Observer on Conformity to Handwashing Norm," *Perception and Motor Skills*, 62, 169–170.

PINKER, S. [1997], *How the Mind Works*, W. W. Norton: New York.

SACKS, O. [1995], *An Anthropologist from Mars*, Vintage Books: New York.

Gerd Gigerenzer
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94
14195 Berlin
Germany
E-mail:
gigerenzer@mpib-berlin.mpg.de

Richard McElreath
Department of Anthropology
University of California, Davis
One Shields Avenue
Davis, CA 95616-8522
USA
E-mail:
mcelreath@ucdavis.edu