

Reply to Binmore: Social Norms or Social Preferences?

Herbert Gintis

May 25, 2011

I am in agreement with many of Ken Binmore's remarks in *Social Norms and Social Preferences* (2010). Like Binmore, I welcome the breath of fresh air that behavioral economics has brought to economic theory, and along with Binmore, the themes of justice and reciprocity are particularly important for my interest in the evolution of social norms. Binmore is correct in asserting that neoclassical economics does not depend on the view that people are selfish, and as he points out, in market contexts, laboratory experiments indicate that traditional theory works well. Behavioral economic findings are thus enriching and expanding neoclassical economics rather than undermining it.

Also like Binmore, I believe social norms are an emergent property of human sociality, and exist as macrosocial structures that are not reducible to the preferences of individuals. As Binmore stresses, it is difficult to formulate a coherent theory of social preferences unless one has a systemic model of social norms existing outside of the system of individual preferences. Indeed, in Gintis (2009) I identify the endeavor to subsume social norms under the rubric of social preferences as an untenable form of methodological individualism. Given this basic agreement on the most, if not all, scientific points, this note hopes to aid the reader in understanding and evaluating the nature of our disagreements.

Ken Binmore imputes a lack of personal character and professional integrity to those with whom he disagrees. The objects of his critique "muscle in on neighboring disciplines," with a "new imperial venture," they "downplay inconvenient facts," and they have "betrayed the scientific ethos." Such assertions threaten to turn a useful dialog into unenlightened name-calling.

Binmore claims that

Experiments on the one-shot Prisoners Dilemma show that subjects (usually students) who are new to the game contribute slightly more than 50% of the time, but that their behavior changes as they gain experience. After playing repeatedly (against a new opponent each time), about 90% of subjects end up free riding after ten trials

or so... Since there is a small but vociferous school of behavioral economists who seek to downplay this inconvenient fact (Eckel and Gintis 2010), it is worth pointing out that Camerer (2003, p. 46) explicitly endorses the result as standard in his much cited Behavioral Game Theory.

There is to my knowledge only one such one-shot prisoner's dilemma experiment in which subjects have new partners on each round (Andreoni and Miller 1993). In this experiment, subjects played 200 rounds of a prisoner's dilemma, each time with a new partner, and the average rate of cooperation was approximately constant (at about 18%) over all 200 rounds, including the last 20 rounds. Certainly Eckel and I did not "seek to downplay" these findings. Nor are they "inconvenient," as they pose no challenge to, and indeed tend to validate the models of conditional cooperation that we have developed.

A careful inspection of Camerer (2003), p. 46 shows that he says nothing about such games, much less "explicitly" endorsing Binmore's claim.

Perhaps Binmore means *repeated public goods games*, not *one-shot prisoner's dilemma games*. If so, Eckel and I (2010) do briefly defend a social preferences interpretation of the evidence. A more complete defense is in Bowles and Gintis (2011), in which we make the following argument in a section entitled *Free-Riders Undermine Cooperation*, the first paragraph of which states "In a social dilemma that is repeated for a number of periods, subjects tend to start with a positive and significant level of cooperation, but unless there are very few free-riders in the group, cooperation subsequently decays to a very low level." We continue with the following analysis (I have added some comments on the most recent literature):

This decay is significant for the following reason. A supporter of the self-interest axiom is inclined to interpret other-regarding behavior in experiments as confusion... The anonymity of the laboratory may indeed be sufficiently extraordinary that subjects simply play by the prudent and self-regarding rules of everyday life. Accordingly, the... decline in contributions in the public goods game might be seen as a confirmation [that] subjects are not altruistic; rather, they are simply learning how to maximize their payoffs through game repetition.

However, were this explanation correct, if the same subjects were permitted to play a second multi-round public goods game identical to the first, they should refuse to contribute on the very first round. Andreoni (1988) and Cookson (2000) tested this prediction and found it to be wrong. When the public goods game is played with several groups and after every series of rounds group membership is reshuffled

and the game is restarted, subjects begin each new series by contributing about half, but each time cooperation decays in the later rounds. If one believes that the decay in contributions within a game is due to learning how to maximize payoffs in the context of anonymity, one would also have to believe that subjects unlearn the money-maximizing behavior between series! In fact, the only reasonable explanation for the decay of cooperation is that public-spirited contributors want to retaliate against free-riders, and the only way available to them in the game is by not contributing themselves. Subjects often report this reason for the unraveling of cooperation retrospectively. For a clear explanation of the “restart” phenomenon in terms of the rational actor model, see Fischbacher and Gächter (2010) and Ambrus and Pathak (2011).

Another indication that free-riding and the retaliation against free-riders is the cause of the unraveling of cooperation comes from an experiment by Page et al. (2005). The experimenters compared four baseline sessions, each of which included 16 subjects in a 20-round public goods game, with four sessions in which, after three 20-round games, subjects were given a list of the average contributions of the other players in all four groups, and were permitted to rank their preference for playing with one or more of these subjects. Subjects who ranked each other highly were assigned to the same group, and subjects who were not ranked highly by others were also assigned to the same group.

In baseline treatments with random assignment to groups, contributions began at an average of 60% of the maximum possible and declined to 9% in last round, for an average contribution rate of 38% of the endowment over the 20 rounds. Where subjects could choose their partners, cooperation did not decay over time, and the average contribution rate was 70% of the endowment. Note that this high average cooperation rate includes the performance of low contributors, who were obliged to play with one another.

To understand this result, note that when subjects could choose their partners, there was a strong tendency for subjects to play with others who approximately share their level of contribution. This is because the experimenters would always satisfy the request of two players who preferred to be together before the request of a pair only one of whose members preferred to associate with the other. Thus the top of the four elective groups maintained an average contribution rate of over 90% with no tendency to decay, except for an end-game effect

in the last three rounds that brought contributions down to about 60%. The second most preferred group maintained an 80% average, with a similar end-round effect, while the third group averaged about 65%, with a relatively weak tendency to decay, from about 75% in the first rounds to 60% in rounds 12 to 16, and then to about 50% in the final three rounds. The lowest group showed the usual decay from 75% contribution in the first three rounds to 10% in the final round, for an average of 45%.

These results are consistent with the idea that the decay of cooperation is due to relatively high contributors reacting to low contributors by lowering their own contribution. When subjects in the same group are relatively uniform in their contributing behavior, this decay mechanism is attenuated. In particular, when those predisposed to cooperate can associate preferentially with like-minded people, cooperation is not difficult to sustain.

I suggest, in light of these considerations, that we have not “downplayed” the issue of the decay of cooperation in the public goods game and related social dilemmas.

Binmore comments that

The introductory chapters of Henrich et al. (2004) provide a typical summary of the claims that these behavioralists make. The particular claims that are relevant here concern the extent to which social or other-regarding utility functions can usefully be said to... be inconsistent with the neoclassical paradigm.

As I explain in great detail in Gintis (2009), social preferences enrich rather than undermining the rational actor model, game theory, or more broadly, the neoclassical paradigm. The authors referred to above (I am one), however, do not claim this. Rather, they claim that the evidence undermines the notion that preferences are self-regarding and amoral.

Binmore asserts:

... a vocal minority of behavioralists have succeeded in outflanking the skepticism of mainstream economists by appealing directly to a wide range of social and life scientists with the aid of publications in journals like *Science* and *Nature* that do not normally publish the work of economists (for example, Fehr and Gächter 2002; Bowles and Gintis 2002). They have thereby been successful in discrediting psycho-sociological explanations of the data (in terms of cultural

norms) in favor of traditional economic explanations (in terms of the optimization of utility functions) while simultaneously decrying the very economic methodology that they seek to persuade other to adopt.

In fact, we have published extensively in first-ranked economics journals, to the point where behavioral economics, including the analysis of social preferences, is doubtless among the most well-accepted additions to economic theory of the past few decades. However, we are more generally behavioral scientists, not just economists, and we have published in first ranked professional journals in biology, anthropology, psychology, sociology, and political science. Our results do not “belong” to economics, and indeed can be evaluated profitably by psychologists, sociologists, and indeed any professional acquainted with the scientific method. The papers published in *Science* and *Nature* are extensions of research published widely in all the behavioral sciences. Rather than attempting to discredit neoclassical economics, we expect our findings to be added to its repertoire.

Far from having “discredited” psycho-sociological explanations of the data, our findings have strongly reinforced such explanations by adding the methodological rigor of behavioral game theory and the rational actor model..

Our transdisciplinary perspective reflects the nature of the objects of our research. Disciplinary boundaries in the behavioral sciences were set up many decades ago and do not reflect contemporary insights into the nature human sociality. We behavioral game theorists have simply followed questions concerning human behavior and strategic interaction wherever they take us in terms of disciplinary specialties. Rather than being “muscling in” or “imperialism,” our demeanor is a rather exemplary exercise of the scientific method. Moreover, we are just as insistent on bringing ideas from psychology, sociology, and the other social sciences in to economics as we are to use economic theory to enrich psychology, sociology, anthropology, and the other behavior disciplines.

REFERENCES

- Ambrus, Attila and Parag A. Pathak, “Cooperation over Finite Horizons: A Theory and Experiments,” *Journal of Public Economics* 95 (2011):500–512.
- Andreoni, James, “Why Free Ride? Strategies and Learning in Public Good Experiments,” *Journal of Public Economics* 37 (1988):291–304.
- and John H. Miller, “Rational Cooperation in the Finitely Repeated Prisoner’s Dilemma: Experimental Evidence,” *Economic Journal* 103 (May 1993):570–585.
- Binmore, Kenneth G., “Social Norms or Social Preferences?,” *Mind and Society* 9 (2010):139–158.

- Bowles, Samuel and Herbert Gintis, *A Cooperative Species: Human Reciprocity and its Evolution* (Princeton: Princeton University Press, 2011).
- Camerer, Colin, *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton: Princeton University Press, 2003).
- Cookson, R., "Framing Effects in Public Goods Experiments," *Experimental Economics* 3 (2000):55–79.
- Eckel, Catherine and Herbert Gintis, "Blaming the Messenger: Notes on the Current State of Experimental Economics," *Journal of Economic Behavior and Organization* 73,1 (January 2010):109–119.
- Fischbacher, Urs and Simon Gächter, "Social Preferences, Beliefs, and the Dynamics of Free-Riding in Public Good Experiments," *American Economic Review* 100 (2010):541–556.
- Gintis, Herbert, *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences* (Princeton: Princeton University Press, 2009).
- Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis, *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies* (Oxford: Oxford University Press, 2004).
- Page, Talbot, Louis Putterman, and Bulent Unel, "Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry, and Efficiency," *Economic Journal* 115 (October 2005):1032–1053.