

67

The selfishness-altruism debate: In defense of agnosticism

Philip E. Tetlock

Institute of Personality Assessment and Research, University of California, Berkeley, CA 94720

Caporaël et al. launch a frontal assault on egoistic incentive theory – a loose intellectual alliance of reinforcement theorists, economists, sociobiologists, and political philosophers who share a belief in the basic selfishness of human nature. Caporaël et al. claim that egoistic incentive theory rests on flawed evolutionary premises and is hard pressed to explain potentially disconfirmatory evidence without resorting to “just-so” stories. The assault is partially successful – a substantial achievement, given the well-entrenched historical status of the foe.

The proper theoretical response to the biological critique is to concede that it is indeed possible to weave a number of plausible evolutionary scenarios around our emergence as a species. Caporaël et al. may well be right that the small face-to-face group (15 to 30 individuals) was more likely to have been the primary locus of natural selection pressures than the solitary individual. They may also be right that affective and cognitive mechanisms supportive of group living are genetically linked attributes that were highly conducive to the survival of early hominids wandering the savanna plains of Africa. Empathy, a predisposition to share, and punitiveness toward cheaters are likely natural consequences of this group life-style. But they are not the only possible consequences. It was probably no more “adaptive for ancestral humans to identify automatically with an in-group and to accept its goals as their own” (sect. 4, para. 6) than it is now for modern humans. Fanatics – those most eager

to leap to the defense of the group against predators or to lead the attack against rival groups – often have life expectancies too short to pass their DNA-encoded commitments onto the next generation. I suspect that there is an Aristotelian golden mean lurking here: Nature has probably smiled especially kindly on those who recognize that their long-term self-interest hinges on the viability of their group, but who also recognize that serious conflicts can arise between individual and collective interests and who appreciate the importance of knowing when to “defect” and how to do so without incurring the wrath of the collectivity. In short, human nature is complex and embodies many, often conflicting motives; evolutionary arguments are necessarily speculative; and agnosticism on the “selfishness” question probably remains the most prudent position.

Caporael et al. not only question the evolutionary hard core of the selfishness research program, they also raise an intriguing empirical challenge. In a fascinating series of public-goods experiments, they claim to have falsified the hypothesis that people will not contribute to group welfare in the absence of egoistic incentives to do so. Subjects in these experiments were strangers who made a single decision under conditions of anonymity and in the absence of interaction either before or after the experimental session. The basic design renders implausible many of the classic egoistic incentives (e.g., reciprocity, coercion) but not all of them. As the authors are well aware, advocates of egoistic incentive theory have numerous potential lines of theoretical defense. In various experiments, substantial percentages of subjects may have cooperated out of a desire to avoid guilt, to bolster their self-esteem, to help their team win, or just to put themselves in a good mood. Indeed, there is independent evidence that each of these motives promotes prosocial behavior (Krebs & Miller 1985). There are also potential methodological defenses. For example, subjects may not have believed the assurances of anonymity (why would the experimenter bother to conduct a study in which it is impossible to monitor what I’ve done?)

The key question is, of course, whether the results reported, taken as a whole, force egoistic incentive theory to invoke such an array of ad hoc hypotheses that, far from providing a parsimonious integrative framework, the theory begins to look like a rather tattered patchwork quilt. My view is that egoistic incentive theory is indeed worse for the wear but can still be rescued with a measure of dignity by forging new intellectual alliances – in particular, with social identity and role theorists who assume that people strive to create the most favorable identities for themselves within the constraints of social situations (cf., Hogan 1982; Schlenker 1982). *Homo economicus* must meet *Homo sociologicus* at least halfway. People are trying (not always successfully) to maximize socially defined or constructed utility functions. From this standpoint, independent variables can shape the willingness to contribute to public goods in two basic ways: (1) by affecting the identity implications of response options (people do not want to appear selfish or foolish but do want to appear loyal to the group, trustworthy, and generous); (2) by affecting the importance of different identity objectives or of different audiences (self vs. other, in-groups vs. out-group). Manipulations such as anonymity, no-free-ride, opportunity for group discussion, the minimum number of contributors, and group membership status of potential beneficiaries must influence willingness to contribute through one or the other of these mediational processes – a sweeping hypothesis that can only be tested by painstakingly measuring the self and social identity implications that participants and observers feel it is reasonable to draw from different response options in the original experimental situations (for examples, see the “interpersonal simulations” by Alexander & Knight 1971; Tetlock 1980).

In closing, it is reasonable to ask whether the proposed intellectual merger would negate most of what was provocative and interesting about the original versions of egoistic incentive theory. It is one thing to claim, *pace* Hobbes, that people are

selfish to the core and will cooperate only under threat of centralized coercion; it is quite another to claim, *pace* Goffman, that people are so thoroughly socialized that they will readily make monetary sacrifices to avoid appearing to be the “wrong” type of person, either in their own eyes or the eyes of others. We can choose to call these two positions variants of egoistic incentive theory, but they are really very different. From a meta-theoretical point of view, the key question concerns where we should draw boundaries around the powerful organizing metaphors – *homo economicus*, the actor, the scientist, the generalizer – that still give intellectual impetus and direction to so much work in the behavioral and social sciences. As I have argued from the controversies over the merits of cognitive versus motivational explanations of social judgment (Tetlock & Levi 1982) and intrapsychic versus impression management explanations of interpersonal behavior (Tetlock & Manstead 1985), there is no nonarbitrary way to demarcate the explanatory range of research programs organized around competing metaphors. It is hard to tell the difference between a social-identity-theory variant of the egoistic incentive research program and the evolutionary-bases-of-group-altruism argument advanced by Caporael et al. In the end there will probably be no dramatic crucial experiment. The two positions may just gradually blur into each other. This theoretical convergence should, moreover, be taken as an encouraging sign that, notwithstanding the claims of the radical subjectivists, research programs anchored in starkly different first premises can be responsive to a common body of evidence.