gies such as tit-for-tat and Pavlov that enable individuals to cut their losses in exchanges with selfish players and reap the benefits of cooperating with cooperators may pay off in iterated games in the long run. Note that these strategies are commonly considered cooperative, not altruistic.

With respect to commitment, the more one has invested in a strategy, the greater the potential costs of switching. This said, it is in individuals' interest to abandon losing strategies. Thus, for example, we might expect a reformed alcoholic to revert to drinking if sobriety fostered depression.

The author argues that "the crucial issue is whether or not altruism is a subcategory of self-control . . . there is no need to postulate an innate altruistic mechanism; the job can be done by . . . an innate learning mechanism" (sect. 8). I disagree for two reasons. First, an evolved strategy could give rise to both self-control and altruism as easily as an innate learning mechanism could. It doesn't really matter what you call the mechanism. Second, although it is possible to define altruism in a way that requires self-control, I am not sure what is gained by viewing it as a subcategory of the process. The key to the selection of altruism is not, in my view, self-control; the key is the adaptive benefits of the overriding strategy in question.

Why cooperate? Social projection as a cognitive mechanism that helps us do good

Joachim I. Krueger and Melissa Acevedo

Department of Psychology, Brown University, Providence, RI 02912.
Joachim_Krueger@brown.edu Melissa_Acevedo@brown.edu
http://www.brown.edu/Departments/Psychology/faculty/krueger.html

Abstract: The mother sacrificing herself while rescuing someone else's child is a red herring. Neither behaviorism nor cognitivism can explain it. Unlike behaviorism, however, the cognitive process of projection can explain cooperation in one-shot social dilemmas.

Making the case for teleological behaviorism as an explanatory framework for altruism and other forms of selfless cooperation, Rachlin "does not deny the existence of [other, cognitive] mechanisms," but he considers it "unnecessary to postulate the existence of such a general mechanism" (target article, sect. 1.1). This view is unremarkable unless one takes it to mean that teleological behaviorism is the better explanation because only its mechanisms offer a *necessary and sufficient* explanation of altruism. We think that the case for the sufficiency of teleological behaviorism has not yet been made, and we offer an example of a sufficient cognitive mechanism.

The Mother running into a Burning House (MBH) to save somebody else's child while risking her own life is the paradigm of altruism throughout the article. Any ambitious theory of altruism must attempt to explain such extraordinary behavior because everyday acts of altruism are readily explained away by some lurking self-interest. It is only fair to ask whether teleological behaviorism rises to the challenge. The explanatory tale is that some people have been collecting delayed or long-term rewards for altruism or other forms of self-controlled behavior. As a result, they have formed an enduring commitment, motive, or habit of extending this pattern of behavior into the future.

The case of the MBH poses a problem. One must assume that the individual differences in habit strength or commitment are highly reliable and transferable to new situations. Unfortunately, individual differences in personality, of the type assumed here, emerge as usable predictors only after massive aggregation across situations. Psychometricians consider predicting individual acts a near-hopeless enterprise. Darley and Batson's (1973) study of Good-Samaritanism is a classic example of how psychometrics failed to predict who would help. Beyond its rarity, the case of the MBH is complicated by its extremity. It is difficult to find a class of acts with which it can be categorized. What are the charitable behaviors that shaped the habit that is now being activated? Suppose the woman had a routine of taking the neighbors' kids to the bus stop. This habit may well have been shaped by mutually reciprocated cooperation over time, but can it now be considered the cause for the woman's self-sacrifice? To suggest that it can puts credulity to the test, especially when no theoretical, empirical, or quantitative lever is offered as a guide. While teleological behaviorists and the parents of the saved child may respectively see a good habit and saintliness at work, the woman's own family may feel rather differently. Indeed, the perspective of the woman's family would probably best predict how the woman herself would feel when confronted with the existential challenge of a burning house. By casting the self-sacrificial rescue as an act of self-control, teleological behaviorism must ask which base and self-defeating impulse is being kept at bay. It would appear to be fear of death, which begs the question of what kind of learning history prepares one to scoff at death. Perhaps there is none, and that's why women with little children are particularly hesitant to die for the children of others, whatever their altruistic commitments might be otherwise.

Cooperation in the prisoner's dilemma game (PD) is far more common. Although its prevalence makes it more tractable psychometrically, cooperation depends at least as much on the perceived personality of the opponent than on the player's own personality (de Bruin & van Lange 1999). Most disturbing is the finding that once players learn *that* their opponent has either cooperated or defected, almost all defect. They cooperate only as long as they do not know *whether* the opponent cooperates (Shafir & Tversky 1992). If habit and commitment were such strong forces, why should uncertainty matter?

One answer lies in the cognitive mechanism of projection, which Quattrone and Tversky (1984) first applied to the social dilemma of voting, and which Baker and Rachlin (2001) introduced to the PD. Projection is a generalized expectation that others will reciprocate whichever course of action one chooses. Thus, cooperation increases with the perceived probability of reciprocation. When projection is perfect, the PD devolves into a choice between the payoffs for mutual cooperation and the payoffs for mutual defection. The dilemma disappears, and the player can cooperate out of self-interest. Projection can be learned, but such learning is not necessary. The expectation that others will act as we do may well be an adaptive default handed down by evolution. If anything, gradual learning about how others actually behave reduces rather than enhances perceptions of similarity (Krueger 1998).

Neither teleological behaviorism nor projection can explain the MBH. Projection can, however, parsimoniously explain why many people cooperate even in the one-shot PD when they do not know what the opponent will do, but defect when they know what the opponent did. Teleological behaviorism would have to appeal to commitments that are conditional on uncertainty, in which case they would not be terribly sincere as commitments go.

Teleological behaviorism and altruism

Hugh Lacey

Department of Philosophy, Swarthmore College, Swarthmore, PA 19081.
hlacey1@swarthmore.edu

Abstract: Rachlin shows that experiments about social cooperation may fruitfully be grouped with experiments on self-control, and that this suggests interesting possibilities for practical behavioral controls. The concepts of *selfishness* and *altruism*, however, that inform his theorizing about these experiments, do not serve to provide understanding of the behavior that commonly is referred to, derogatorily, as selfish.

A core thesis of Rachlin's teleological behaviorism is that "mental terms" – these include common value and intentional terms – re-