# Putting Altruism in Context*

Joel Sobel †

September 11, 2002

**Abstract**

I argue that Rachlin's notion of self control is imprecise and not well suited to the discussion of altruism. Rachlin's broader agenda, to improve collective welfare by identifying behavioral mechanisms that increase altruism, neglects the fact that altruism is neither necessary nor sufficient for desirable social outcomes.

*I thank Philip Babcock, Vincent Crawford, Jim Moore, and Joel Watson for useful comments and NSF for financial support.

†University of California, San Diego. Email: jsobel@ucsd.edu.

Altruistic behavior presents a puzzle for both biology and economics. Biology must explain why traits that do not maximize an organism's fitness survive. Economics, which starts from the assumptions that individual actors are selfish and rational, must explain why individuals take actions that do not appear to maximize short-term self interest.

Evolutionary biology provides powerful ways of understanding unselfish behavior to closely related individuals (Hamilton [5]) or to reciprocity in long-term relationships (Trivers [12]), Rachlin is concerned with instances of human altruism that are hard to explain using these theories. Evolutionary mechanisms that rely on group selection (Boehm [3] or Sober and Wilson [11]) or community enforced morality (Alexander [2]) provide explanations for unselfish behavior in human communities. These approaches teach us that to understand altruistic actions, we should examine the individual in connection with the composition and norms of the society in which he lives.

Economic theory assumes that agents select from their available choices only an action that maximizes utility. Unselfish behavior does not occur. Economics adapts its methodology in the face of apparently contradictory evidence by broadening the definition of self interest, for example by assuming that individuals obtain utility from the act of giving or through the consumption of others, or by recognizing that economic relationships are dynamic and that an individual's long-run selfish best interest is best served by doing things that are not consistent with short-term selfish goals.[1]

The approaches of both biology and economics illustrate that the context of actions is important and require careful attention to the definition of altruism. Rachlin [7] argues that one must consider altruism in the context of patterns of behavior and provides a definition that makes the mechanism supporting altruism a special case of the mechanism that determines self control.

Rachlin should be commended for pointing out the importance of patterns of behavior. Altruism generated by a preference for establishing a pattern of "good" behavior is internally motivated. It does not rely on generating reciprocal responses from others that are vital to the dynamic arguments in biology and economics. Rachlin's emphasis on patterns should motivate behavioral researchers to widen the context of their experiments. Choice models, at least as they are used in applications, may need to be broadened

[1]Sobel [10] provides an overview.

to permit the study of consumption patterns rather than instantaneous flows of consumption. Rachlin's argument has three weaknesses, however. First, he takes as an axiom the most puzzling aspect of altruism. Second, he fails to provide complete and coherent definitions of his terms, making it difficult to evaluate the implications of his analysis. Third, his focus on the relationship between altruism and patterned behavior is artificial. An argument for paying more attention to patterns would be much more powerful in another setting. The remainder of the comment elaborates on these criticisms.

An essential assumption for Rachlin's approach is that individuals value a lifetime of altruistic behavior more than a lifetime of selfishness. This assumption begs the central question of evolutionary biology because it does not explain why such behavior should have a fitness advantage over individual selfish behavior. Rachlin can shift the discussion of altruism to another category of behavior, but still must provide a reason why natural selection favors individuals who have a preference for maintaining altruistic patterns.

Rachlin's goal is to identify a behavioral mechanism by which altruism can be developed over a lifetime. This task is an ambitious one, so let us grant him the premise and see how well he does with it. Consider his definition of self control. The first part posits that an individual prefers a long activity to $n$ repetitions of a brief activity. The second part states that the brief activity is preferred to a $t$ length fraction of the longer activity. This definition is incomplete for two reasons.

First, the definition is incomplete because it does not explain how to divide up the long activity. Rachlin's own examples demonstrate that decompositions are problematic. Consider a few more examples. Suppose that the short activity is touching your nose with your hand. Assume that this action for 30 seconds is preferred to listening to a 30 second segment of a symphony. Is a preference for listening to the entire symphony for an hour over touching your nose evidence of self control? You (or your next of kin) would be extremely unhappy if your six hour flight from New York to Paris ended after four and a half hours. For purposes of the definition of self control, what is a fraction of a transatlantic plane flight? Does one exercise self control by staying on the plane for the entire six hours? These examples, and those provided by Rachlin, warn that careful definitions of the objects of choice must come before a discussion of self control.

The second, related, sense in which Rachlin's definition is incomplete is that it does not define the domain of preferences. This weakness seriously

interferes with an understanding of Rachlin's hypothesis. In order to satisfy the first condition of Rachlin's definition of self control, an individual must be able to rank a pair of activities performed over a $T$ period horizon. The second condition compares one-period activities. In order to say whether one brief activity is preferred to another, however, one must take into account the entire interval of length $T$. An individual may prefer to have a drink in the first five minutes of a party followed by abstinence than to have no drink at all, while the same individual may prefer to completely abstain from drinking to having eight drinks in the evening. Does this person prefer the brief activity of having a drink to a short interval of abstinence? It depends on whether the person expects the brief interval to be followed by more drinking. In order to talk about the individual's preferences over the first five minutes, one must be explicit about what the individual will do for the rest of the evening. Rachlin does not do this and, consequently, one is left with several different explanations for the self control problem. One explanation is simple impatience. Much (but not all) of Rachlin's discussion is consistent with the notion that self control is the result of placing low weight on future utility. Another explanation is based on time inconsistency. An individual could enter the party with the idea that optimal behavior is to have a drink in the first five minutes and then abstain, but the individual may be aware that she'll feel differently after she has the first drink. This individual may try to postpone – or avoid – her first drink in order to exercise control over her "future selves." This idea is similar to Ainslie's [1], whereas Rachlin plainly is after something else.[2] It is impossible to support or refute Rachlin's hypothesis until he defines his choice environment more carefully.

The third problem with Rachlin's approach is that altruistic patterns of behavior are abstract, while the desire to maintain patterns is stronger when patterns are concrete. Rachlin's article (and several of the essays in Schelling [8]) provide examples of intuitive ways in which people follow simple routines in order obtain desirable long-term outcomes. While the ideal may be to have one or two drinks at a party, with the amount of drinking determined by context (who is at the party, the quality of the liquor, what is planned for the subsequent day, ...), it is easy to identify a pattern

[2]Recent work by Gul and Pesendorfer [4] provides an elegant formulation of self control in terms of preferences over sets of choices. Although essentially static, their framework is a coherent alternative to that offered by Rachlin.

of complete abstinence. The external mechanisms that we use to control ourselves, for example diets, automatic savings plans, and religious rituals, often demand rigid adherence to clearly patterned behavior. This suggests that when following a pattern of behavior is a goal in itself, the pattern should be transparent.

Altruism is different. There are too many opportunities to give to others for us to follow a uniform pattern of goodness. We all are part Shen Te (Brecht's Good Woman of Setzuan) and part Shui Ta (her selfish alter ego). Rachlin's altruistic woman may be willing to die to maintain a pattern of good behavior, but she quickly forgets that she did not place a dollar into a homeless person's outstretched hand. A major challenge to Rachlin's experimental research agenda is to understand better what can and cannot become a pattern.[3]

Rachlin's article has a hopeful subtheme: His behavioral view of altruism suggests techniques for increasing altruistic behavior, which would then lead to good collective outcomes. This position requires closer examination. Selfless actions aggregated across individuals need not lead to good collective outcomes.[4] Selfish actions taken in the context of well designed institutions may lead to good collective outcomes.[5] One can be skeptical about Rachlin's argument or even the importance of human altruism,[6] and still believe strongly that humans can identify and construct stable institutions that lead to good outcomes even in the face of self-interested behavior.

[3]There is scope for both behavioral and evolutionary explanations of pattern formation. Nesse [6] points out how emotions can serve as commitment devices and that psychiatric conditions like depression may have selective advantages in some environments. Obsessive-compulsive disorders provide examples in which the need to create and follow patterns is excessive. These phenomena may be exaggerated versions of the mechanisms that give us pleasure from establishing and following patterns of behavior.

[4]Some would argue that both characters in O'Henry's "Gift of the Magi" would have been better off if at least one of them had acted selfishly.

[5]Economic theory's fundamental theorems of welfare economics provide conditions under which a consequence of rational self-interested behavior is economic efficiency.

[6]In the same way, Smuts [9, page 323] criticizes Sober and Wilson [11] for "their apparent assumption that a more benevolent view of human nature depends on the existence of altruism."

# References


[1] Ainslie, G. (1992). *Picoeconomics*. Cambridge University Press.

[2] Alexander, R. (1987). *The Biology of Moral Systems*. New York: Aldine de Gruyer.

[3] Boehm, C. (1997). Impact of the Human Egalitarian Syndrome on Darwinian Selection Mechanics. *The American Naturalist*, 34, S100-S121.

[4] Gul, F., & W. Pesendorfer (2001). Temptation and Self-Control. *Econometrica*, 69 (6), 1403-1435.

[5] Hamilton, W. (1964). The Genetical Evolution of Social Behavior. *International Journal of Theoretical Biology*, 7, 1-17.

[6] Nesse, R. (2001). The Evolution of Subjective Commitment. In *Evolution and the Capacity for Commitment*. New York: Russell Sage.

[7] Rachlin, H. (2002). Altruism and Selfishness. *Behavioral and Brain Sciences*.

[8] Schelling, T. (1984). *Choice and Consequence*. Cambridge: Harvard University Press.

[9] Smuts, B. (1999). Multilevel Selection, Cooperation, and Altruism. *Human Nature*, 10 (3), 311-327.

[10] Sobel, J. (2001). Interdependent Preferences and Reciprocity.

[11] Sober, E., & D. Wilson (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge: Harvard University Press.

[12] Trivers, R. (1971). The Evolution of Reciprocal Altruism. *Quarterly Journal of Biology*, 46, 35-58.