

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Publications, Department of Psychology

Psychology, Department of

June 2002

The need for proximal mechanisms to understand individual differences in altruism

Gustavo Carlo

University of Nebraska-Lincoln, carlog@missouri.edu

Rick A. Bevins

University of Nebraska-Lincoln, rbevins1@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/psychfacpub>



Part of the [Psychiatry and Psychology Commons](#)

Carlo, Gustavo and Bevins, Rick A., "The need for proximal mechanisms to understand individual differences in altruism" (2002). *Faculty Publications, Department of Psychology*. 79.

<https://digitalcommons.unl.edu/psychfacpub/79>

This Article is brought to you for free and open access by the Psychology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, Department of Psychology by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

theory as the “dogma of the ghost in the machine,” but arguably language actually does function in some ways like a ghost in a machine. Linguistic control systems enable human beings to transcend individual experience and allow the contemplation of possibilities that never have been, and never could be, experienced. It also allows the symbolic sharing of experiences with others, including others long dead: Plato, Aristotle, Aquinas. Ghosts, of a kind.

The Western tradition has often viewed logic and reason as somehow superior to the passions we share with animals: Indeed, “right conduct” is often viewed as involving the control of animal passions. However, reason, logic, and organized social rules have been at the core of some of the most violent and destructive of human behaviors, including the official directives, chains of command, and orderly bureaucratic procedures of the Holocaust. An alternative is to view prosocial emotions of attachment and bonding as being the truly effective counter to aggression and violence (Buck 1988; 1999).

There is evidence of the importance of emotional bonds in mediating a variety of behaviors with moral implications: fostering cooperation and altruism and reducing aggression and conflict. Examples of such emotional controls of behavior have been found in conflict resolution among monkeys and apes, as observed by de Waal and colleagues (e.g., de Waal 1996; de Waal & Aureli 1997). In human beings, there is considerable evidence that feeling empathy for a needy person leads to altruism, that is, to unselfish tendencies to help that person (Eisenberg & Fabes 1991; Eisenberg & Miller 1987; Hoffman 1975; 1976). C. Daniel Batson and colleagues (Batson & Oleson 1991; Batson & Shaw 1991) reviewed evidence for the role of selfish motives in altruistic behavior and advanced the empathy-altruism hypothesis (EAH): that the expression of needs by the other naturally evokes empathic emotions of sympathy and compassion that motivate altruistic responses.

In a larger sense, the gulf between the approach of Rachlin and that of Preston and de Waal reflects the gap between the Piaget-Kohlberg analysis of moral judgment and the morality of caring emphasized by Carol Gilligan. Gilligan and colleagues argued that there are two fundamental moral orientations. The justice perspective “holds up an ideal of reciprocity and equal respect,” whereas the care perspective “holds up an ideal of attention and response to need” (Gilligan & Attanucci 1988, p. 73). Either or both of these perspectives can be active during moral choice. Moreover, there are suggestions that choice behavior per se is more emotional than previously believed (Lowenstein et al. 2001).

Human behavior is multiply determined, and it arguably is an error to dismiss another point of view because it is “not necessary” to explain a phenomenon. All aspects of human empathy and altruism cannot be explained by the principles developed by Preston and de Waal (2002). However, an exclusive focus on choice can leave out the emotional controls that may set the basic agenda for human morality in general and altruism in particular.

The need for proximal mechanisms to understand individual differences in altruism

Gustavo Carlo and Rick A. Bevins

Department of Psychology, University of Nebraska-Lincoln, Lincoln, NB
68588-0308. gcarlo@unl.edu

Abstract: There are three concerns regarding Rachlin’s altruism model. First, proximal causal mechanisms such as those identified by cognitive neuroscientists and behavioral neuropharmacologists are not emphasized. Second, there is a lack of clear testable hypotheses. And third, extreme forms of altruism are emphasized rather than common forms. We focus on an overarching theme – proximal mechanisms of individual differences in altruism.

Rachlin proposes a theory of altruism that focuses on self-control as the central explanatory mechanism. Other notable aspects to

this model include the emphasis on temporal patterns of altruistic behaviors and the connections made to evolutionary theory. Although Rachlin is to be commended for elevating the importance of these aspects to explain altruism, there are a number of gaps in the model that seriously limit this theoretical perspective. These limitations will likely decrease any impact his theory may have on the field. We briefly outline some of these concerns and propose avenues for future theoretical and empirical pursuit.

Psychological debate and research on altruism have often focused on the existence of altruism. Assuming the existence of altruism moves the debate to the nature of altruism. Accordingly, it is important to place the issue of altruism in its broader context. Altruism is considered to be a subset of the larger set of prosocial behaviors (i.e., behaviors that benefit others) which include behaviors that *primarily* benefit others, often incurring a cost to the self (Carlo & Randall 2001; Eisenberg & Fabes 1998). This definition requires one to acknowledge the presence of much variation in the forms of altruism. Our working definition is, of course, subject to debate; but for the purpose of the present commentary, it is necessary to make explicit. That is, our subsequent comments regarding Rachlin’s model may be partly attributed to definitional differences.

We have three recurring concerns regarding Rachlin’s altruism model. First, virtually no time is spent on proximal causal mechanisms such as those that might be provided by areas like cognitive neuroscience and behavioral neuropharmacology. Second, there is a lack of clear testable hypotheses that follow from this self-control model. Finally, the model appears to be built around extreme forms of altruism (e.g., woman saving unknown baby) rather than the more regularly occurring forms that vary widely in occurrence across and within individuals. From our perspective, these concerns are interrelated. Thus, this commentary will address these issues by focusing on an overarching theme – proximal mechanisms of individual differences in altruism.

An account of individual differences in altruism requires the consideration of more proximal causal mechanisms. According to some theorists (Carlo & Randall 2001; Eisenberg & Fabes 1998; Hoffman 1991), individual differences in altruism can stem from differences in cognitive, emotive, and social context (e.g., culture-related socialization experiences). We will focus on a specific set of emotive variables relevant to altruism. A number of investigators posit that empathy is the primary motivator associated with altruism (Batson 1998; Carlo & Randall 2001; Eisenberg & Fabes 1998; Hoffman 1991). Empathy can be defined as an other-oriented matching emotion that results from vicariously observing another’s distress. There are two processes that may stem from empathy: sympathy and personal distress. Sympathy is an other-oriented vicarious emotional response that results in feelings of sorrow or concern for a needy other. In contrast, personal distress is a self-focused vicarious emotional response that results in aversive, uncomfortable feelings. The difference between these two empathy-related responses is critical because while sympathy can result in prosocial behaviors (including altruism), personal distress often results in avoidance behaviors (but see Batson 1998).

The characteristics of individuals who exhibit these different responses in distress situations are distinct. For example, empathy and sympathy responses reflect moderate sympathetic arousal whereas personal distress reflects over-arousal. Derryberry and Rothbart (1988) proposed a temperament theory that identifies two major dimensions: physiological reactivity and self-regulation. Physiological reactivity refers primarily to affective arousal and motor activity. It includes an assessment of emotional and behavioral threshold, latency, intensity, and rise and fall time. Self-regulation refers to behavioral and emotional control. Specific aspects of this dimension include attentional processes, approach-withdrawal, soothability, and behavioral inhibition (Rothbart et al. 1994). More important, Rothbart proposes that empathic and prosocial tendencies are best explained by examining the interaction between these two internal processes. Following Rothbart’s model, personal distress responses are the joint consequences of

individuals who have difficulty modulating their physiological arousal and have overactive physiological reactivity; there is evidence to support these assertions (Rothbart et al. 1994; see also Eisenberg & Fabes 1998). Thus, Rothbart's approach represents a set of more proximal causal mechanisms that might account for individual differences in altruism, and it also respects the wide variations in altruism.

In contrast to Rothbart's model, we are unable to find in Rachlin's proposal an explicitly defined set of analogous proximal causal agents. This situation is unfortunate in that such proximal mechanisms would help account for the wide variation seen in the forms of altruistic behaviors, the individual differences seen between and within individuals, and, arguably the most important point, provide researchers with clear and testable hypotheses that avoid circularity (Panksepp 1998). For example, Rothbart's notion of distress as a functional interaction between physiological reactivity (e.g., arousal) and self-regulation (e.g., emotional control) suggests distinct neurophysiological processes (e.g., hypothalamic pituitary adrenal axis versus frontal cortical areas, respectively). Notably, individual difference in distress, and hence likelihood of behaving altruistically, becomes a product of these processes that vary with evolutionary and individual history. The task of identifying all neurobiological factors that mediate the presence or absence of an altruistic behavior at time x is daunting. However, it is an obtainable goal that already has a basis from which to start. Powerful animal models exist that could be used to explore the processes posited to mediate altruism (e.g., distress). There are numerous papers concerning rodent models, indicating that different experiences in early development (e.g., naturally occurring maternal care) can differentially impact later sensitivity to distress (e.g., alteration in hypothalamic pituitary adrenal axis, see Liu et al. 1997; for other examples, see Boksa et al. 1998; Dellsu et al. 1996; Kehoe et al. 1998). An exciting possibility would be to merge the work on individual difference with an animal model of self-control that attempts to measure choice that includes prosocial options (see Poulos et al. 1998 for an example of assessing individual differences using a self-control [impulsivity] preparation).

ACKNOWLEDGMENTS

Funding support to the first author was provided by a grant from the John Templeton Foundation and the Mayerson Foundation. The second author was partially supported by USPHS grant DA11893 while preparing this commentary. Correspondence may be addressed to Gustavo Carlo, Department of Psychology, University of Nebraska-Lincoln, Lincoln, NB, 68588-0308, e-mail: gcarlo@unl.edu.

Learning to cooperate: Reciprocity and self-control

Peter Danielson

Centre for Applied Ethics, University of British Columbia, Vancouver, V6T 1Z2, Canada. pad@ethics.ubc.ca <http://www.ethics.ubc.ca/~pad>

Abstract: Using a simple learning agent, we show that learning self-control in the primrose path experiment does parallel learning cooperation in the prisoner's dilemma. But Rachlin's claim that "there is no essential difference between self-control and altruism" is too strong. Only iterated prisoner's dilemmas played against reciprocators are reduced to self-control problems. There is more to cooperation than self-control and even altruism in a strong sense.

As Rachlin points out, the analogy between self-control and altruism has a long history. For example, Sidgwick (1893) and Nagel (1970) use the analogy as a bridge from the natural appeal of prudential motivation to a justification of altruism. These authors argued that since we stand to others as to our own future selves, we should not treat the two cases differently. Rachlin moves the normative discussion from this weak appeal to rationality as symme-

try, to a productive blend of formal and empirical modeling methods.

Rachlin makes two significant contributions to the study of social cooperation. First, in a literature that has emphasized rationality and evolution, he focuses on agents learning to cooperate. Second, he develops the analogy between learning self-control and learning social cooperation by showing that two problems in these areas share a common structure. The primrose path (PP) to addiction is a problem for self-control and the prisoner's dilemma (PD) is a problem for cooperation because immediate reinforcement is higher for the addictive choice, X, or the noncooperative choice, D, respectively. Thus, each presents learners with a local optimum trap. Rachlin argues that learners who structure reinforcement for patterns of acts rather than single acts can avoid these traps.¹

Although he offers much to build on, Rachlin's claim that "there is no essential difference between self-control and altruism" (target article, sect. 7, last para.) is too strong. Self-control may be necessary for cooperation, but it is not sufficient; he underemphasizes the crucial factor of reciprocity. That is, although Rachlin puts great weight on reciprocity, the complexities of dealing with human self-control and cooperation obscure some issues. We propose to clarify the discussion by focusing on simple artificial agents and games.

The main problem is that whereas a temporally extended agent is the benefactor from her own self-control, this need not be the case for cooperation. In the best case, cooperators share the benefits with others; in the worse case, unreciprocated cooperation harms the agent compared to her opponent. So we suspect that self-control might be learnable by reinforcement under weaker conditions. To test our intuition, we construct the simplest learner, who tries two alternative acts and selects the one with the higher immediate reinforcement; it will choose addictive X in the PP. Even if we allow reinforcement to be remembered over trials, the result is the same (Fig. 1). But if we constrain choice (and therefore reinforcement) to the simplest patterns, that is, to sets such as XXX and YYY, Y is chosen (Fig. 2).²

Turning to the PD, we get parallel results. The act chooser will learn to defect and the pattern chooser to cooperate. In Figure 3 the dark bars show Rachlin's PD game iterated for 40 rounds with a Learner playing against Tit-For-Tat. Performance increases with pattern length up to 3, then decreases (because the longer defection trial is costly). But we will get this parallel only under restrictive conditions. First, the game must be iterated; in the one-shot PD, D remains dominant. Rachlin notes that his subjects "are making only one in a series of choices extending to their lives outside of the lecture hall" and claims that "for the teleological behaviorist there can be no social trap without repetition. All prisoner's dilemmas are repeated" (sect. 3, last para.). Be this as it may for humans, we can impose the restriction explicitly for automata and focus on repeated PD games. Second, learning to cooperate

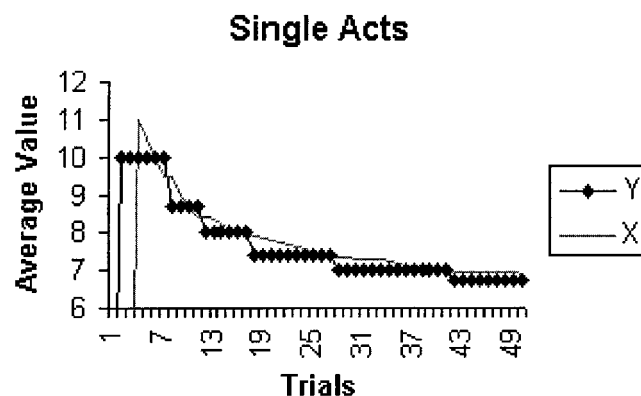


Figure 1 (Danielson).