Altruism, collective rationality, and extreme self-sacrifice

Andrew M. Colman and Briony D. Pulford

School of Psychology, University of Leicester, Leicester LE1 7RH, UK

Andrew M. Colman: amc@le.ac.uk (0)116 229 7197; Briony D. Pulford: bdp5@le.ac.uk

Andrew M. Colman: https://www2.le.ac.uk/departments/npb/people/amc
Briony D. Pulford: https://www2.le.ac.uk/departments/npb/people/bdp5

Abstract

Puzzlement about extreme self-sacrifice arises from an unarticulated assumption of psychological egoism, according to which people invariably act in their own self-interests. However, altruism and collective rationality are well established experimentally: people sometimes act to benefit others or in the interests of groups they belong to. When such social motives are sufficiently strong, extreme self-sacrifice presents no special problem of explanation and does not require outgroup threats.

Asking why people are sometimes willing to lay down their lives for the sake of their groups implies that extreme self-sacrifice is incomprehensible or perplexing. What, precisely needs explaining? The perplexity seems to arise from a deep-rooted but usually unarticulated assumption that people are inherently selfish and invariably motivated to act in their individual self-interests. Within that implicit theory of human motivation, deeply embedded in...
individualistic societies such as the United States, extreme self-sacrifice is indeed inexplicable, but the theory itself is false and misleading.

The philosophical doctrine that we are all motivated solely to maximize our individual welfare is called *psychological egoism*: (May, 2011; Slote, 1964). It asserts that everything we do, no matter how beneficial to others, is actually calculated to benefit ourselves. It allows the possibility of mistakes (acting in ways that fail to maximize one’s welfare through errors or oversights) and *akrasia* (trying to do the best for oneself but failing through weakness of will); but it rules out altruism (benefiting another person at some cost to oneself) and leaves no room for collective rationality (acting in the interests of a group to which one belongs). Psychological egoism is either tautological or false: tautological if we interpret any deliberately chosen action as selfish by definition, or false if we adopt a more intuitive interpretation of selfishness. The economist and Nobel laureate Paul Samuelson (1993) poured scorn on the tautological interpretation: “I will not waste ink on face-saving tautologies. When the governess of infants caught in a burning building reenters it unobserved in a hopeless mission of rescue, casuists may argue: ‘She did it only to get the good feeling of doing it. Because otherwise she wouldn’t have done it.’ Such argumentation (in Wolfgang Pauli’s scathing phrase) is *not even wrong*. It is just boring, irrelevant, and in the technical sense of old-fashioned logical positivism ‘meaningless’” (p. 143).

There is overwhelming experimental evidence that altruistic behavior can be elicited reliably through experimental procedures designed to arouse empathy (Batson, 2011; Batson, Dyck, Brandt, Batson, & Powell, 1988; Batson & Shaw, 1991). It is hard to escape the conclusion that
empathic emotions evoke genuinely altruistic motives that have the ultimate goal of benefiting those who elicit the empathy and not those who feel it. Furthermore, there is evidence that people sometimes choose to act in the interests of their groups rather than their individual selves (Bardsley, Mehta, Starmer, & Sugden, 2010; Bardsley & Ule, 2017; Butler, 2012; Colman, Pulford, & Lawrence, 2014; Colman, Pulford, & Rose, 2008). These findings confirm everyday observations of people acting in what they believe to be the interests of their families, companies, universities, or their religious, ethnic, or national groups, even when those interests do not coincide with their own selfish interests. When these social motives are sufficiently powerful or fanatical, they can lead to extreme self-sacrifice.

The theory of social value orientation (SVO) was introduced by Messick and McClintock (1968) and McClintock (1972) to provide a rigorous conceptual framework for interpreting altruism, collective rationality, and other social motives. In its simplest (dyadic) interpretation, an SVO represents a person’s preferences regarding the allocation of a resource between self and another individual. In terms of utility theory, suppose $t_i$ and $t_j$ are objective payoffs—for example, amounts of money—to two people $i$ and $j$. Then $i$’s utility $u_i(t_i, t_j)$ is a function of both objective payoffs. Under the individualistic SVO, $i$ is motivated to maximize $u_i = t_i$, selfishly optimizing $i$’s own objective payoff without regard to $j$’s. This is the only motivation recognized under the doctrine of psychological egoism, although from a contemporary neurobiological perspective it can be viewed as pathological (Sonne & Gas, 2018). There is also persuasive experimental evidence for an altruistic SVO, when $i$ is motivated to maximize $u_i = t_j$, hence being concerned solely to maximize another’s welfare, and a cooperative SVO, when $i$ is motivated to maximize $u_i = t_i + t_j$, the collective payoff of both individuals (for reviews of SVO theory and evidence, see...
Altruism, collective rationality

Balliet, Parkes, & Joireman, 2009; Bogaert, Boone, & Declerck, 2008; Murphy & Ackermann, 2014, who also discuss competitive and an equality-seeking SVOs that do not concern us here).

Once we acknowledge that people are sometimes motivated to perform actions intended to benefit others, the puzzle of extreme self-sacrifice evaporates, and we do not need outgroup threats to interpret social motivation. In a Stag Hunt game, for example, two hunters are motivated to act in the joint interest of the dyad in catching a stag, although each is tempted selfishly to chase a hare that can be caught without the other’s help. No external threat is involved, but collective preferences need to be supplemented with team reasoning (Bacharach, 1999; Sugden, 1993). Team-reasoners first search for an outcome that they believe would be best for the dyad or group; if such an outcome exists and is unique, they then play their parts in the joint enterprise. In these theories, orthodox decision theory is a special case of team reasoning when the team is a singleton (for a comprehensive review of team reasoning theories and experimental evidence, see Colman & Gold, 2017).

People occasionally perform acts of extreme self-sacrifice to benefit others or to act in what they believe to be the interests of the groups to which they belong. Such self-destructive actions can arise from nonselfish motives of altruism or collective rationality. How people come to hold social motives so fervently is worth investigating; in many instances of self-sacrificial mass murder, religion evidently plays a key role. According to the physicist and Nobel laureate Steven Weinberg (1999): “With or without religion, good people can behave well and bad people can do evil; but for good people to do evil—that takes religion.” However, why people practice extreme
self-sacrifice given their fervent altruistic or collectivistic social motives is hardly puzzling once we free ourselves from the debilitating misconception of psychological egoism.

Alphabetical reference list


