# BIG QUESTIONS STARTER PACK – 2018-2019

## Resolved: Humans are primarily driven by self-interest.

# Aff

## Definition

### Self-interest

#### Self-interest should be construed broadly - otherwise the debate devolves into infinite regress

Weeden and Kurzban 17 <Jason Weeden, Pennsylvania Laboratory for Experimental Evolutionary Psychology, Robert Kurzban, University of Pennsylvania, “Self-Interest Is Often a Major Determinant of Issue Attitudes,” Advances in Political Psychology, Vol. 38, Suppl. 1, 2017, https://onlinelibrary.wiley.com/doi/abs/10.1111/pops.12392>#SPS

A Wider View of Self-Interest Having addressed the usual objections to self-interest effects, we now step back and take a wider view. What justifies equating self-interest with short-term material self-interest? Is there a plausible theory of human nature that would recommend that equation? In this section, we discuss our own view of self-interest, grounded in modern perspectives on humans. These days, not even economists typically believe that human motives are reducible to short-term material considerations. Standard economic perspectives posit individuals that **maximize their preferences.** These preferences might include getting more money in the short term, to be sure, but that by no means exhausts the list. People might also seek to gain prestige, have sex, assist their children in fulfilling their own preferences, or various other goals (Becker, 1996). But a perspective on self-interest that would be useful for purposes of large-scale empirical study cannot be one that views the advancing of self-interest as anything that helps a given individual get what they want at a given moment. While the equation of self-interest with short-term material self-interest is too restrictive, an overly individualized view of self-interest would be too loose. In the middle ground, we have proposed a view of self-interest that has some fealty to key aspects of the narrower views of self-interest, but one that acknowledges that typical human motives extend beyond short-term monetary ones (Weeden & Kurzban, 2014, chap. 2). We agree, for example, that it’s preferable to ground notions of self-interest in widely shared goals rather than overly individual ones, that it’s preferable to focus on goals with tangible implications, and that an explanation of competing political views will likely be driven by goals that have competitive social implications. We also agree that short-term economic goals fit the bill—the desire for more money in the short-term is a widely shared human goal, it has tangible implications, and it’s an area where people compete over opposing outcomes. Where we part company is that we do not view short-term economic advancement as the only (or even the most important) widely shared, tangible, competitive human goal. Our own view of interests derives from our evolutionary approach, which views humans as social animals with minds designed to advance tangible, fitness-related goals (Kurzban, 2010; Petersen, 2016). These goals are genetically selfish (Dawkins, 1989)—that is, aimed at advancing the outcomes of one’s self and one’s relatives—and involve competitive aspects of social life that have been biologically relevant throughout human existence, including satisfying immediate physiological needs (e.g., eating and finding shelter), defending one’s self and valued others, establishing social ties, gaining and maintaining social status and esteem, attracting and retaining mates, and parenting (Kenrick, Griskevicius, Neuberg, & Schaller, 2010). Further, the social aspects of human life include nonrelatives sharing (to various degrees) a range of costs and benefits within coalitions and social networks. This is particularly true among close friends, who often share to a degree the benefits of each other’s positive outcomes and the burdens of each other’s negative outcomes (DeScioli & Kurzban, 2009; DeScioli, Kurzban, Koch, & Liben-Nowell, 2011) but also involves other kinds of networks (work colleagues, friends of friends, fellow church members, and so on). In short, we think that humans generally are motivated to advance outcomes across various evolutionarily relevant domains (including resources, social status, and mating lives) particularly among themselves, their relatives (in accordance with the degree of their relatedness), and members of their own social networks (in accordance with the closeness of the benefit-and-burden-sharing connection). While the narrow self-interest definition has focused on short-term economic matters—tracking political issues such immediate tax hikes or unemployment benefits for the currently unemployed—our evolutionary view expands the political terrain on which a tangible self-interest perspective can operate. As we explain below in the fifth section, we find interest-based demographic patterns involving not only issues of economic redistribution and provision of resources to the poor, but also issues of discrimination, meritocracy, and social status as well as issues affecting sexual and reproductive lifestyles. So is our view about “self-interest”? In a sense, no. Just as Dawkins (1989) discussed how (ultimately) selfish genes can produce individuals who behave at times nonselfishly, our view is one of social agents designed to behave genetically selfishly but not necessarily individually selfishly. On the other hand, as we mentioned, typical definitions of “self-interest” in political science explicitly include the interests of both one’s self and one’s family (e.g., Kinder, 1998; Sears & Funk, 1990). So a common political science usage of “self-interest” already contains a genetic expansion of self. Does our inclusion of social network members mean it’s not “self-interest”? Not really, given that we view these considerations as a kind of indirect self-interest through shared benefits and burdens among individuals. Or perhaps it’s only “self-interest” when we’re talking about economic outcomes, but something else when we’re talking about areas like social status or sexual lifestyles. But we view status and sex as tangible areas. Discrimination tangibly impacts everyday life. Restrictions on abortion and birth control tangibly impact everyday life. Thus, we have described our viewpoint as one that sees a major role for “self-interest” in political issue positions and political coalitions. We have also introduced the phrase “inclusive interests” (borrowing from the evolutionary term “inclusive fitness”) as a reminder that we’re talking about selfand-family interests across a range of evolutionarily salient social outcomes (Weeden & Kurzban, 2014, chap. 2).

### Altruism

#### Altruism must be entirely self-sacrificing – any other definition is baseless

West, Mouden, and Gardner 11 <Stuart A, Professor of Evolutionary Biology, Oxford; Claire El, Post-Doctoral Prize Research Fellow, Fellow of Nuffield College, Oxford; and Andy, Professor of Biology, St. Andrews University, “Sixteen common misconceptions about the evolution of cooperation in humans,” Evolution and Human Behavior 32 (2011) 231–262, https://ac-els-cdn-com.proxy-um.researchport.umd.edu/S1090513810000905/1-s2.0-S1090513810000905-main.pdf?\_tid=b4c44800-3b4d-475d-a299-b1c4c3cacb49&acdnat=1532202873\_69651dceecc6439f51952635b2dffbf2>#SPS

6.1.1. Misconception 1: The various redefinitions of altruism (Baschetti, 2007; Becker, 1974; Bergstrom, 1995, 2002; Bowles, 2006, 2009; Bowles & Gintis, 2004, 2008; Boyd et al., 2003; Fehr & Fischbacher, 2003; Gintis, 2000; Sober & Wilson, 1998; Trivers, 1971; Wilson, 1975a) In Section 3 we emphasized how terms such as altruism have very specific meanings, that have formal justification and convey useful information. If these terms are misused, or redefined, the result is confusion. This has been a particularly large problem with the term altruism (West et al., 2007b; p. 419–423), which has been redefined in evolutionary models in many ways, including: (a) a decrease in fitness over the short term, so that reciprocity is “reciprocal altruism” (Becker, 1974; Fehr & Fischbacher, 2003; Trivers, 1971); (b) a decrease in the fitness of the focal individual, relative to the other members of its group (relatively costly to individual, relatively beneficial to the group; sometimes termed weak altruism) (Baschetti, 2007; Bergstrom, 1995; Bowles, 2006; Bowles & Gintis, 2004; Boyd et al., 2003; Gintis, 2000; Sober & Wilson, 1998; Wilson, 1975a); (c) playing cooperate in a prisoners' dilemma game (Bergstrom, 2002); (d) a failure to harm others (Field, 2001); (f) giving up resources in order to benefit others (Pradel et al., 2009); (g) the mechanism by which one individual is motivated to help others (Axelrod, 1984); (h) the willingness to take mortal risks as a fighter (Bowles, 2009). The first problem with these redefinitions is that they lack a formal justification to use intentional language from an evolutionary or ultimate perspective. This is because they require the costs and benefits to be defined in different ways and not with respect to lifetime reproductive success. As discussed in Sections 2 and 3, natural selection produces organisms that behave intentionally, as maximizing agents, at the level of lifetime reproductive success. The second problem is that these redefinitions include scenarios where cooperation could provide a direct fitness benefit, and hence be either mutually beneficial (+/+) or altruistic (−/+). Considering a specific case, Gintis (2000) compared the relative fitness of two different strategies: “self-interested agents” who do not punish or cooperate, and altruistic “strong reciprocators” who cooperate and punish noncooperators. He labels strong reciprocators as altruistic because they “increase the fitness of unrelated individuals at a cost to themselves.” However, in this and related models, cooperation is individually costly within the social group but provides a benefit to all the members of the group through mechanisms such as increased productivity or reducing the rate of group extinction (Bowles & Gintis, 2004; Bowles et al., 2003; Boyd et al., 2003; Gintis, 2000; Gintis, Bowles, Boyd, & Fehr, 2003; Henrich & Boyd, 2001). Consequently, any individual that behaves cooperatively also gains this (direct) benefit, which can outweigh the cost of performing the behaviour (Binmore, 2005b; Burnham & Johnson, 2005; Lehmann et al., 2007c; West et al., 2007b). This leads to the confusing situation where: (a) cooperation can be favoured because it provides a direct benefit to the cooperator because it increases the chance they and the rest of their group survive, but this is defined as altruistic rather than in their self interest (West et al., 2007b); (b) a “selfish agent” (Bowles & Gintis, 2004) can have a lower direct fitness than an altruist.

## Psychology

### Strong Reciprocity - General

#### Experiments in favor of “Strong Reciprocity,” a form of altruism, are seriously flawed

West, Mouden, and Gardner 11 <Stuart A, Professor of Evolutionary Biology, Oxford; Claire El, Post-Doctoral Prize Research Fellow, Fellow of Nuffield College, Oxford; and Andy, Professor of Biology, St. Andrews University, “Sixteen common misconceptions about the evolution of cooperation in humans,” Evolution and Human Behavior 32 (2011) 231–262, https://ac-els-cdn-com.proxy-um.researchport.umd.edu/S1090513810000905/1-s2.0-S1090513810000905-main.pdf?\_tid=b4c44800-3b4d-475d-a299-b1c4c3cacb49&acdnat=1532202873\_69651dceecc6439f51952635b2dffbf2>#SPS

There is a large empirical literature showing that when humans play anonymous one-shot economic games, they cooperate more than would be expected if they were purely self-interested (Ledyard, 1995). From a proximate perspective, **it has been argued that this is because individuals value the success of others as well as their own**, showing prosocial preferences (Fehr and Schmidt, 1999). From an evolutionary perspective, it has been argued that this proximate mechanism cannot be explained by standard evolutionary explanations of cooperation, such as kin selection and reciprocity, and requires the a novel explanation of strong reciprocity (Fehr & Gächter, 2002; Fehr & Fischbacher, 2003; Fehr & Rockenbach, 2004; Fehr et al., 2002; Gintis 250 S.A. West et al. / Evolution and Human Behavior 32 (2011) 231–262 et al., 2003, 2005b). For example, it has been claimed that human behaviour “cannot be rationalized as an adaptive trait by the leading evolutionary theories” (Fehr et al., 2002). However, **the empirical data are open to multiple explanations and do not support this claim.** First, in some cases, an equally valid explanation for the data is that humans are antisocial, rather than prosocial. In the ultimatum game, the expected strategy is for individuals to make minimal offers and for these to be accepted. If there is a chance that minimal offers will be rejected (punished) then individuals are expected to make larger offers (Gale et al., 1995). Consequently, the larger than minimal offers that are observed in experiments may just reflect the fact that individuals expect small offers to be punished. In this case, the unexpected behaviour is the rejection of small offers, and so, we might conclude that the data show that humans have a tendency to punish at a level greater than that expected from selfish interests. Note that our purpose here is not to argue that humans are particularly pro- or antisocial, just that it is easy to give multiple explanations for the data. Second, higher than expected levels of cooperation can be explained by individuals **making mistakes in laboratory settings.** The previous interpretation of economic games is based upon the implicit assumption that if individuals do not play perfectly, then this does not lead to a systematic bias in the level of cooperation (Kümmerli et al., 2010). This is a problem, because when the predicted behaviour is to not cooperate at all (e.g., in standard public goods games), then any deviations from perfection would automatically be perceived as greater than expected cooperation (Houser & Kurzban, 2002). Kümmerli et al. (2010) tested this possibility, by allowing individuals to play modified versions of public goods games, where 100% cooperation was the strategy that would maximise their personal financial game. They found that while this led to an increased level of cooperation, it did not lead to full cooperation (see also Houser & Kurzban, 2002; Laury & Holt, 2008; Saijo & Nakamura, 1995). If the logic from previous studies (e.g., Fehr & Schmidt, 1999) was applied to this result, then it would give a utility function that is negatively influenced by the success of others (an antisocial preference). Given that a simultaneous positive and negative regard to others is not possible, these data instead suggest that individuals have a tendency to avoid both full defection and full cooperation (Haselton & Nettle, 2006). Third, another possible explanation is that higher levels of cooperation are normally favoured and that this leads to a psychology that results in cooperation in one-shot experiments. The idea here is that, even if they are given perfect information, individuals find it hard to disassociate themselves from the real world, and so, cooperation occurs as a byproduct of the fact that is normally favoured (Bateson et al., 2006; Binmore, 2006; Burnham & Johnson, 2005; Hagen & Hammerstein, 2006; Haley & Fessler, 2005; Levitt & List, 2007; Nowak et al., 2000; Trivers, 2004; West et al., 2007b). Experimental support for this suggestion comes from a number of experiments which show that players taking part in one shot games, in which there are no future interactions, still adjust their level of cooperation in response to artificial cues, such as the presence of eyespot pictures on computer desktops (Bateson et al., 2006; Burnham & Johnson, 2005; Haley & Fessler, 2005) or interactions with individuals which do not influence the game (Houser & Kurzban, 2002; Kurzban et al., 2007). The idea here is that these cues trigger responses that have arisen in response to situations outside of the laboratory, where whether or not they are being observed will matter. Further support comes from cultural differences in experimental games (Gächter & Herrmann, 2006; Henrich et al., 2006; Henrich et al., 2005), which appear to reflect differences in how the game is perceived to relate to everyday events (Binmore, 2006). To put it another way, “Experimental play often reflects patterns of interaction found in everyday life” (Henrich et al., 2005, p. 798) and not just the game set up imposed by the experimenter. Furthermore, even in laboratory settings, behaviours such as punishment can provide a direct benefit if longer periods of interactions are allowed for (Gächter et al., 2008). The data discussed in the previous two paragraphs suggest that **humans have a psychology which can “misfire” in laboratory settings.** While it might be argued that the possibility that humans don't always behave perfectly is no surprise, **the more important point is that such imperfect behaviour can lead to a systematic bias towards higher than expected levels of cooperation.** Future work must address this issue, through the use of appropriate controls and by exercising greater caution when interpreting the absolute level of cooperation in particular treatments (Kümmerli et al., 2010). It would be useful to test whether there is a bias to accept evidence for humans being “extra cooperative,” without sufficient basis, due to a bias towards positive evidence or because this is a nice result to get. This “misfire” idea has been argued to be incorrect in several papers, where it is labelled the “big mistake” or maladaptation hypothesis (Boyd & Richerson, 2002; Fehr & Henrich, 2003; Gintis et al., 2003; Henrich, 2004). The implicit idea here is that humans should always behave perfectly. However, this hypothesis is clearly falsified by the numerous examples of how proximate mechanisms which have been previously favoured by natural selection lead to behaviours that do not maximise fitness under certain conditions. For example, the mismatch between real danger and our fear of snakes and spiders versus automobiles, various aspects of the porn industry, rises in obesity, or the decline in reproductive rate can be associated with better living conditions (Hagen & Hammerstein, 2006). It is even clearly falsified in the context of economic games, where, as discussed above, individuals show variation in behaviour in response to misleading “cues” of being observed, such as eye-spots on computers. It can be misleading to call such imperfect behaviour a maladaptation or a mistake, in the sense that it may be the optimal state, just that the benefits S.A. West et al. / Evolution and Human Behavior 32 (2011) 231–262 251 of improving a behaviour have to be balanced or traded off against the costs (Partridge & Sibley, 1991; Stearns, 1992). The general point here is that maximisation of fitness does not imply perfect behaviour in every possible situation, and that the selective regimen needs to be considered, as has been shown frequently in the animal behaviour literature (Davies, 1992; Herre, 1987; Herre et al., 2001; Krebs & McCleery, 1984; Pompilio et al., 2006; Wehner, 1987). Evolutionary theory does not predict that humans (or any other organism) should behave as perfect maximising agents in every situation in which they can be placed.

#### “Strong Reciporicty” theories confuse proximate and ultimate causes – humans may do something that appears altruistic for a PROXIMATE cause, but the deepest desire may still be self-interest

West, Mouden, and Gardner 11 <Stuart A, Professor of Evolutionary Biology, Oxford; Claire El, Post-Doctoral Prize Research Fellow, Fellow of Nuffield College, Oxford; and Andy, Professor of Biology, St. Andrews University, “Sixteen common misconceptions about the evolution of cooperation in humans,” Evolution and Human Behavior 32 (2011) 231–262, https://ac-els-cdn-com.proxy-um.researchport.umd.edu/S1090513810000905/1-s2.0-S1090513810000905-main.pdf?\_tid=b4c44800-3b4d-475d-a299-b1c4c3cacb49&acdnat=1532202873\_69651dceecc6439f51952635b2dffbf2>#SPS

It is useful to distinguish between ultimate and proximate explanations of traits or behaviours (Mayr, 1961; Tinbergen, 1963). Proximate explanations are concerned with the causal mechanisms underlying a behaviour (how questions). Ultimate explanations are concerned with the fitness consequences of a behaviour (why questions). Evolutionary biology attempts to explain features of an organism from an ultimate perspective—why are organisms the way they are? The key point is that these different methodologies are complementary and not competing alternatives. The Nobel Prize winner Niko Tinbergen (1963) famously clarified the distinction between ultimate and proximate explanations for animal behaviour, in the most influential paper of his career (Kruuk, 2003); less well known to many biologists is that Niko's brother Jan won the 1969 Nobel memorial prize in Economics. One of Tinbergen's classic studies to illustrate this distinction was on the removal of eggshells from their nests by black-headed gulls. The mechanistic (proximate) explanation for this is that individuals are more likely to remove objects from their nest if they are white- or egg-coloured, have frilly edges, and if they are feather-light. The evolutionary (ultimate) explanation for this is that it makes aerial predators such as herring gulls less likely to find their brood. These explanations are clearly not competing (each answer cannot provide a solution to the other problem), and a fuller understanding is gained by considering both. A clear example of the confusion that may be caused by conflating ultimate and proximate factors is provided by work on “**strong reciprocity**,” which is defined proximately but then given as a solution to an ultimate problem (Bowles & Gintis, 2004; Fehr & Gächter, 2002; Fehr & Rockenbach, 2003, 2004; Fehr & Fischbacher, 2003, 2004; Fehr et al., 2002; Gintis et al., 2003). A strong reciprocator has been defined as a combination of “a predisposition to reward others for cooperative, norm-abiding behaviours” and “a propensity to impose sanctions on others for norm violations” (Fehr & Fischbacher, 2003). This is a description of a proximate mechanism. However, it is then given as a solution to an ultimate problem—for example, “Strong reciprocity thus constitutes a powerful incentive for cooperation even in nonrepeated interactions when reputation gains are absent” (Fehr & Fischbacher, 2003), or “cooperation is maintained because many humans have a predisposition to punish those who violate group-beneficial norms” (Bowles & Gintis, 2004). This is illustrated even more clearly with a discussion of neurological work, where it is suggested that an explanation for the punishment of individuals who do not cooperate is that such punishment leads to “satisfaction” (Fehr & Rockenbach, 2004; Quervain et al., 2004). For example, in two adjoining sentences, Quervain et al. (2004, p. 1254) follow an ultimate question “Why do people punish violators of widely approved norms although they reap no offsetting material benefits themselves?” with a proximate answer “We hypothesize that individuals derive satisfaction from the punishment of norm violators.” **This does not solve the ultimate problem because it does not answer why evolution should have produced a psychology or nervous system that mechanistically encourages (rewards) such punishment.** This approach mixes up two different questions (how and why, or process and product). Claiming that cooperation is favoured because individuals have a predisposition to cooperate, and punish those that do not, is circular, as it does not explain why individuals should have a predisposition to cooperate and punish in the first place. The proximate question is how is cooperation maintained? The answer to this is a predisposition to cooperate and avoid punishment, i.e., what has been termed strong reciprocity. **The ultimate question is why is cooperation maintained, or more specifically, why are cooperation and punishment (strong reciprocity) maintained?** The possible answer to this is because it provides either a direct and/or an indirect fitness benefit (Gardner & West, 2004). We are not arguing that proximate questions are not interesting, and we appreciate that they are, with good reason, the focus of much human research. Instead, our point is that it is very misleading to mix and match by posing and justifying a problem from an ultimate perspective and then providing a proximate answer.

#### Reciprocal Altruism is insufficient to explain many parts of human behavior

Fehr and Fischbacher 03 <Ernst and Urs, University of Zürich, Institute for Empirical Research in Economics; “The nature of human altruism,” Nature volume 425, pages 785–791 (23 October 2003), http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043>#SPS

What are the ultimate origins behind the rich patterns of human altruism described above? It must be emphasized in the context of this question that a convincing explanation of the distinct features of human altruism should be based on capacities which are distinctly human—otherwise there is the risk of merely explaining animal, not human, altruism. Reciprocal altruism Reciprocal altruism [4](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref4),[5](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref5) in the form of tit-for-tat or similar cooperation-sustaining strategies in the repeated prisoners' dilemma is a powerful ultimate explanation for human altruism in small and stable groups. The experimental evidence unambiguously shows that subjects cooperate more in two-person interactions if future interactions are more likely [9](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref9),[10](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref10). There are, however, several aspects of human interactions that point towards the need to go beyond reciprocal altruism: first, with a few exceptions[26](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref26),[56](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref56), **the evolutionary analysis of repeated encounters has been largely restricted to two-person interactions but the human case clearly demands the analysis of larger groups.** Unfortunately, the evolutionary success of tit-for-tat-like strategies of conditional cooperation is extremely limited even in relatively small groups of 4–8 individuals. It turns out that in a repeated n-person prisoners' dilemma, the only conditionally cooperative, evolutionarily stable strategy prescribes cooperation only if all other group members cooperated in the previous period. The basin of attraction of this strategy is extremely small because a few selfish players suffice to undermine the cooperation of the conditional cooperators.[56](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref56) Second, the interacting individuals are forced to stay together for a random number of periods[6](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref6). This assumption is not only violated by many, if not most, animal interactions but it is also clearly violated in the case of humans. Throughout evolutionary history, humans almost always had the option to stop interacting with genetically unrelated individuals. Thus, the choice of cooperation partners has to be modelled explicitly, which brings issues of reputation formation to the forefront of the analysis. Recent experiments indicate that endogenous partner choice and the associated incentives for reputation formation have powerful effects on human cooperation[57](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref57). Third, reciprocal altruism relies on the idea that altruistic behaviour creates economic benefits for the altruist in the future. Therefore, it has difficulties explaining strongly reciprocal behaviour, which is characterized by the absence of future net benefits. Reciprocal altruism could only explain strong reciprocity if humans behave as if cooperation provides future net benefits, although there are none objectively. The ethnographic evidence suggests that—depending on the interaction partner—humans faced many different probabilities of repeated encounters so that situations often arose in which defection was the best strategy[58](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref58). Indeed, the very fact that humans seem to have excellent cheating detection abilities[59](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref59) suggests that, despite many repeated interactions, cheating has been a major problem throughout human evolution. Therefore, humans' behavioural rules are likely to be fine-tuned to the variations in cheating opportunities, casting doubt on the assumption that humans systematically overestimate the future benefits from current altruistic behaviours.

### Strong Reciprocity - Altruistic Punishment

#### Cultural evolution theories of altruism rely on the idea of “altruistic punishment” – that people will punish violations of social norms even if it hurts themselves to do so – experiments prove that doesn’t happen

Pedersen, Kurzban, and McCullough 13 <Eric J., Department of Psychology, University of Miami; Robert, Department of Psychology, University of Pennsylvania, Department of Economics, University of Alaska; Michael E., Department of Psychology, University of Miami; “Do humans really punish altruistically? A closer look,” Proceedings of the Royal Society B, Volume 280, issue 1758, 7 May 2013, http://rspb.royalsocietypublishing.org/content/280/1758/20122723.short>#SPS

Game structure. (a) Round 1: all players started with $5. Subject was either recipient or third party. The dictator (a fictitious player whose ‘decisions’ were determined by computer script) either took $0 (fair conditions; unbolded) or $4 (unfair conditions; bolded) from recipient. (b) Round 2: Players started with $5. Subject was dictator; previous ‘dictator’ was recipient; third party was excluded. Subject was allowed to give any portion of $5, do nothing or pay a 1 : 4 cost to deduct money from recipient. Money deducted from recipient in round 2 was ‘burned’—it was not gained by dictators as income. Prior to role assignment for the second round, subjects were informed that there would be no third party in round 2 (to avoid potential audience effects [22,34]); one player would be assigned to a different task and be unable to see the results of the interaction. We note that the presence of the experimenter can also induce audience effects [22]; we took great care to minimize this potential influence by (i) clearly informing participants during the consent process that their data would be stored completely anonymously and could not be connected to them in any way, and (ii) minimizing contact with the experimenter by presenting all instructions electronically. Although we cannot rule out experimenter audience effects completely, our results are as insulated from them as we believe was possible in the context of this experiment. All players were given another $5; because previous earnings had been ‘banked’, all players started round 2 with $5. The subject was assigned the role of dictator while the dictator from round 1 (who had treated either the subject or the other player fairly or unfairly) was assigned the role of the recipient (ostensibly by chance). Players were identified consistently throughout, so subjects were aware that the recipient in round 2 was the same player that had been the dictator in round 1. Subjects were instructed that they could give any amount of their $5 to the recipient, do nothing, or remove any amount of the recipient's $5 (the word ‘punishment’ was never used). Removing money cost one-quarter of the amount removed and, unlike in the first round, was not gained by the subject as income—it simply disappeared. Note that the cost of punishment used here, 1 : 4, was less expensive than the 1 : 3 cost typically used in the third-party punishment game; previous research has shown that punishment becomes more likely as the cost of punishment declines (see [10] for review). Following the completion of the round, the experiment ended and the experimenter debriefed the subject through an extensive, staged process to assess the believability of the experiment and to explain why deception was necessary [32]. (ii) Experiments 2a and 2b After providing consent, subjects were instructed to imagine themselves ‘in a particular situation in our laboratory. Please try to picture yourself in the situation we are describing. We will ask you to complete a series of questions regarding how you think you would think, feel and act in this situation.’ The layout and instructions of the game were presented as they were in experiment 1, and the rounds of the game and the self-report measures were the same. However, subjects did not complete a lexical decision task following the first round and were not debriefed following the completion of the experiment (because no deception was involved). (c) Psychometric information regarding the self-report measures (i) Self-rated emotions towards the other players Subjects were asked to describe their emotional responses towards both of the other players after the first round. They described their feelings towards both players to avoid demand effects that might have occurred by probing only about the dictator. (Emotional reactions to the other player were not of theoretical interest here and so, in the interest of brevity, we do not report them.) — Anger. Three-item composite of ratings on a scale from 0 (not at all) to 5 (extremely) of the extent to which the subject was ‘angry,’ ‘mad,’ and ‘outraged’ at the dictator (Cronbach's α = 0.94). — Envy. Two-item composite of ratings on a scale from 0 (not at all) to 5 (extremely) of the extent to which the subject was ‘envious’ and ‘jealous’ of the dictator (Cronbach's α = 0.84). (ii) Fairness/moral wrongness of the round 1 dictator's behaviour Subjects were asked to rate both how ‘fair’ and how ‘morally wrong’ the dictator's behaviour was towards the recipient in round 1 on a scale from 1 (not at all) to 9 (totally). Results (a) Experiment 1 **Third parties did not punish on behalf of strangers:** a one-sample Wilcoxon test (used because distributions were non-normal) revealed that the sample median of the distribution of dollars punished or rewarded in round 2 (in terms of the effect on the recipient, not the cost to the subject) by third-party witnesses of unfairness did not differ significantly from a hypothesized median of zero (z =−1.48, p = 0.140, n = 65; all p-values throughout manuscript are two-tailed). By contrast, victims of unfairness punished a non-zero amount (z =−3.52, p < 0.001, n = 61)—significantly more than mere witnesses of unfairness (p = 0.026, n = 126; two-sample median test; figure 2). If the function of punishment is to deter harmdoers from imposing costs on oneself (i.e. to bargain for better treatment for oneself [10,13,20]) or others in the future—or even if its function is to enforce adherence to social norms [6]—then the punishment must be strong enough to erase unfairly gained benefits [1,35]. Otherwise, the harmdoer retains a net profit from the transgression, and thus will retain an incentive to continue to behave unfairly towards others in the future. Because unfair dictators took $4 from recipients in round 1 of experiment 1, $4 was also the minimum amount of punishment that would be expected to deter unfair dictators from behaving unfairly in the future. Third-party punishment of this magnitude was extremely rare: only 2 of 65 (3%) witnesses imposed at least $4 worth of punishment on unfair dictators, a proportion no different from the proportion for witnesses of fairness (0 of 80; p = 0.199, Fisher's exact test). By contrast, 13 of 61 (21%) victims of unfairness punished at least $4, a proportion significantly greater than that for both recipients of fairness (0 of 64; p < 0.001) and witnesses of unfairness (p = 0.002). Indeed, most victims of unfairness who punished (13 of 21) imposed at least $4 worth of punishment. According to the self-report measures of emotion, third parties did not become angry at unfairness: when controlling for envy (which was highly correlated with anger; r = 0.637, p < 0.001; see the electronic supplementary material, S1.5), a 2 (target: self, other) × 2 (treatment: fair, unfair) ANCOVA revealed a significant target × treatment interaction for anger (F1,265 = 17.11, p < 0.001). Witnesses of unfairness (M = 0.533, s.e. = 0.095, n = 65) did not report more anger than did witnesses of fairness (M = 0.414, s.e. = 0.084, n = 80; p = 0.363, partial η2 = 0.00), but victims of unfairness (M = 1.28, s.e. = 0.098, n = 61) did report more anger than their fairly treated counterparts (M = 0.418, s.e. = 0.093, n = 64; p < 0.001, partial η2 = 0.13; figure 3; see the electronic supplementary material, S1.3 and figure S1 for a replication with an implicit measure based on reaction time data). Thus, people became angry when treated unfairly but not when they only witnessed the unfair treatment of a stranger (cf. [36]). Importantly, this difference in the anger of witnesses versus victims of unfairness was not due to different perceptions of the transgression's fairness or moral wrongness (see the electronic supplementary material, S1.4 and figure S2). Eleven of 65 witnesses of unfairness paid some cost to impose costs on the unfair dictator; with a much larger sample, one might argue, we therefore might have found statistical evidence for mild third-party punishment. However, because the witnesses of unfairness were not angry at the dictator (see above), we suspected that the predominant emotional response among witnesses of unfairness was envy, given that they had observed the unfair dictator obtain a higher payoff ($9) than they themselves had received ($5; see [37]). We found a significant target × treatment interaction for envy (with anger partialled out; F1,265 = 4.53, p = 0.034) witnesses of unfairness were more envious of the dictator than were the witnesses of fairness (p ≤ 0.001, partial η2 = 0.07). By contrast, victims of unfairness were no more envious than were their fairly treated counterparts (p = 0.306, partial η2 = 0.00). Thus, had we observed a significant amount of third-party punishment among witnesses of unfairness, it plausibly could have been motivated by envy towards the unfair dictator rather than by moralistic anger. This difference in the emotions of the witnesses and victims of the dictator's unfairness explains why third-party punishment was quite rare and mild: witnesses of unfairness were envious of the dictator's ill-gotten gains—but not angry—and so they were likely to be reluctant to spend their own money to punish the dictator. During experiment 1, we also ran a small fifth condition (n = 45; see the electronic supplementary material, S1.1) in which witnesses of unfairness started round 1 with $9, enabling us to test whether witnesses’ economic disadvantage relative to unfair dictators explained their surplus envy. Witnesses who started round 1 with $9 were significantly less envious of unfair dictators (controlling for anger) than were witnesses who started with $5 (F1,107 = 8.35, p = 0.005, partial η2 = 0.07). Self-reported anger (controlling for envy) did not differ between groups (F1,107 = 0.581, p = 0.448, partial η2 = 0.01). Witnesses of unfairness with $9 did not punish an amount significantly different from zero (z =−0.879, p = 0.379, n = 45), nor differently than did the witnesses of unfairness with $5 (p = 0.475, n = 110), even though doing so would have cost a smaller proportion of their stake. Thus, the emotional reactions of witnesses of unfairness were characterized by envy rather than moralistic anger. (b) Experiment 2a In experiments 2a (online sample) and 2b (undergraduate sample), we investigated how subjects’ affective and behavioural forecasts in a hypothetical scenario would compare with the results from experiment 1 (see the electronic supplementary material, tables S1 and S2 for descriptive statistics for both experiments). This experiment was conducted not because we thought that participants’ hypothetical responses would provide a reliable assay of how they would behave in a real-life situation, such as the one we explored in experiment 1, but because we wished to compare participants’ forecasts of how they might behave and feel with participants’ actual behaviour and emotional reactions in experiment 1. The rounds of the game were identical to experiment 1, except that subjects were instructed to report how they believed they would act and feel in response to a hypothetical vignette. In experiment 2a—and in contrast to experiment 1—witnesses of hypothetical unfairness forecast that they would administer a greater-than-zero amount of punishment (z =−2.38, p = 0.017, n = 137), as did victims of hypothetical unfairness (z =−2.66, p = 0.008, n = 148; witnesses and victims of hypothetical unfairness did not differ significantly, p = 0.391). Four additional results suggest that a different psychological process was at work in experiment 2a than in experiment 1. First, witnesses of hypothetical unfairness forecast a much higher likelihood of punishing at least $4 (the critical threshold for efficacious punishment) than did witnesses of hypothetical fairness (p = 0.002, Fisher's exact test). This proportion (22 of 137; 16%) was significantly larger than what we saw in the actual behaviour of experiment 1 subjects (2 of 65; 3%; p = 0.009, Fisher's exact test; figure 4). Second, the proportion of witnesses (16%) and victims (31 of 148; 21%) of hypothetical unfairness that forecast punishing at least $4 did not differ (p = 0.361), contrary to experiment 1. Third, witnesses of hypothetical unfairness forecast a significant amount of anger towards unfair dictators in experiment 2a, again in contrast to experiment 1: there was a significant target × treatment interaction (controlling for envy; F1,285 = 5.87, p = 0.016). Witnesses of hypothetical unfairness (M = 1.59, s.e. = 0.109, n = 133) forecast more anger than did witnesses of hypothetical fairness (M = 0.419, s.e. = 0.116, n = 147; p < 0.001, partial η2 = 0.16). Likewise, victims of hypothetical unfairness (M = 2.04, s.e. = 0.110, n = 139) forecast more anger than did recipients of hypothetical fairness (M = 0.345, s.e. = 0.108, n = 141; p < 0.001, partial η2 = 0.28). Fourth, both witnesses (F1,131 = 25.48, p < 0.001, partial η2 = 0.16) and victims (F138 = 10.69, p = 0.001, partial η2 = 0.07) of hypothetical unfairness forecast significantly more anger (controlling for envy) towards the unfair dictator than their counterparts reported in study 1 (figure 3). These results are therefore consistent with proposals that norm violations elicit ‘negative emotions’ [4,6], which in turn motivate altruistic punishment, but here they resulted from affective forecasting rather than from responding to real-time events. (Recall, in contrast, that in experiment 1, which involved real-time behaviour and emotional responses rather than forecasting, no such moral outrage was found.) (c) Experiment 2b The pattern of punishment results for experiment 2b was virtually identical to those of experiment 2a: the proportion of witnesses of hypothetical unfairness that forecast they would punish at least $4 (5 of 85; 6%) did not differ from that of victims of hypothetical unfairness (9 of 101; 9%; p = 0.580, Fisher's exact test)—a pattern that was similar to experiment 2a, but contrary to experiment 1. Interestingly, neither witnesses (z = 0.183, p = 0.855, n = 85) nor victims of hypothetical unfairness (z =−0.162, p = 0.872, n = 101) reported they would administer a greater-than-zero amount of punishment. Nevertheless, both witnesses of hypothetical unfairness (p = 0.031) and victims of hypothetical unfairness (p = 0.012) forecast they would punish significantly more than did their (hypothetically) fairly treated counterparts: this is because both witnesses (z = 2.33, p = 0.020, n = 97) and recipients of hypothetical fairness (z = 2.98, p = 0.003, n = 85) rewarded a greater-than-zero amount. Furthermore—and importantly—witnesses and victims of hypothetical unfairness did not forecast different amounts of punishment (p = 0.743). Thus, notwithstanding the fact that hypothetical punishment of unfairness appeared largely to have taken the form of withdrawing reward (rather than imposing costs) for subjects in experiment 2b, the punishment results largely replicated those obtained in experiment 2a (figure 4) Moreover, the pattern of emotion-related results of experiment 2b was identical to that of experiment 2a: witnesses of hypothetical unfairness (M = 1.43, s.e. = 0.100, n = 94) forecast more anger (controlling for envy) than did witnesses of hypothetical fairness (M = 0.392, s.e. = 0.104, n = 92; p < 0.001, partial η2 = 0.13). Likewise, victims of hypothetical unfairness (M = 1.44, s.e. = 0.096, n = 109) forecast more anger than did recipients of hypothetical fairness (M = 0.259, s.e. = 0.098, n = 99; p < 0.001, partial η2 = 0.16). Witnesses (F1,144 = 18.53, p < 0.001, partial η2 = 0.11) but not victims (F157 = 0.005, p = 0.945, partial η2 = 0.00) of hypothetical unfairness also forecast significantly more anger (controlling for envy) towards the unfair dictator than the subjects in experiment 1 actually experienced (figure 3). The overall pattern of forecast behaviour and emotion in experiment 2b suggests that the students who were the subjects in experiment 2b had a slight tendency to believe that they would reward fair distributions, which the non-student subjects in experiment 2a did not share, but in every other way the results are identical to those of experiment 2a: subjects forecast that both experiencing and witnessing unfairness would cause them to become angry and to punish dictators to a greater extent than did subjects who forecast their responses to either receiving or witnessing fair treatment. Furthermore, both experiencers and witnesses of unfairness forecast equivalent likelihoods of punishing at least $4. Experiment 1 indicates that, under the conditions we investigated, humans do not impose meaningful amounts of third-party punishment on behalf of absolute strangers. The nominal and statistically non-significant amount of punishment we did observe was apparently motivated by envy because of a comparatively unfavourable personal outcome rather than by moralistic anger on behalf of a mistreated stranger. Our finding that the emotional reaction to witnessing unfairness is characterized by envy rather than moralistic anger is particularly inconvenient for the altruistic punishment hypothesis: to categorize a behaviour as an adaptation for altruistic benefit delivery, one needs to provide evidence that the psychological mechanisms that produce the behaviour in question have been designed **for that specific function.** That is, one needs to demonstrate that the behaviour is not caused by mechanisms designed for a different function. The presence of envy, rather than moralistic anger, in response to witnessing unfairness suggests that the psychological mechanisms involved in third-party punishment are, at least in part, designed to process cues that another individual has obtained better outcomes than oneself [38]. By contrast, we found no evidence that they are designed to process cues that an anonymous stranger has been harmed. We do not mean to imply that humans do not impose any third-party punishment: under some circumstances, they do [22,35,39]. However, our results cast doubt on the proposal that the mechanisms that motivate third-party punishment are altruistic benefit-delivery systems that are motivated proximately by moralistic anger.

#### Experiments “proving” altruistic punishment are severely flawed – 5 reasons

Pedersen, Kurzban, and McCullough 13 <Eric J., Department of Psychology, University of Miami; Robert, Department of Psychology, University of Pennsylvania, Department of Economics, University of Alaska; Michael E., Department of Psychology, University of Miami; “Do humans really punish altruistically? A closer look,” Proceedings of the Royal Society B, Volume 280, issue 1758, 7 May 2013, http://rspb.royalsocietypublishing.org/content/280/1758/20122723.short>#SPS

In many animal species, including humans, individuals punish conspecifics that have harmed them [1–3]. Some researchers have recently argued that humans, unlike other animals, also altruistically punish individuals who have harmed others, even when the punisher has no chance of benefiting via reciprocity or benefits to kin [4–6]. Results from several economics experiments appear to support this claim [4,6,7], but some scholars have questioned both the adaptationist logic behind such theoretical claims [8–10] and the interpretation of the empirical results [8,10–14]. Here, we elide these theoretical debates and instead investigate a more basic empirical question: do people actually spontaneously punish individuals who have only harmed other individuals in anonymous settings in the laboratory? Put differently, do the empirical research findings often marshalled in support of the altruistic punishment hypothesis [4,6] provide a reliable guide to the presence or absence of a propensity for altruistic punishment in humans? In previous work, researchers claimed empirical support for the existence of altruistic punishment on the basis of results from public goods game experiments in which the individual being punished had harmed—or failed to help—the putative punisher as well as other victims [4], leaving open the possibility that the punishment was **vengeful**, rather than **altruistic** [10]. Results from similar experiments that exclude revenge as a possible motive suggest that investments in punishment in such contexts are conspicuously low [15]. Additional data frequently adduced in support of the altruistic punishment hypothesis come from third-party punishment games [6,7], in which a dictator chooses to give some portion of a sum of money (or none) to a passive recipient. A third player can punish the dictator (at a cost) in response to the dictator's transfer to the recipient. Many third parties in games with this structure pay a cost to punish stingy dictators, despite receiving no financial benefit from doing so [6,7]. However, five methodological limitations of the standard third-party punishment game might conspire to yield inflated estimates of humans’ propensity to punish strangers for having behaved unfairly towards other strangers. First, in the standard game, subjects are assigned to a third-party role that implies their task is to determine how much to punish the dictator; indeed, the only choices third parties can make are whether to punish the dictator [16] (and, if so, how much). Thus, any error will lead to an increase in the estimated quantity of punishment. Second, punishment in the third-party punishment game is typically administered with the presence (or inferred presence) of an audience: punishment of the dictator by the third party is witnessed by the initial victim because all players see the results of the game. The presence of an audience introduces reputational considerations that could motivate punishment as a means of pursuing indirect fitness benefits (e.g. by signalling one's quality as a cooperative partner [17,18], or one's formidability to prevent future exploitation of oneself [19,20] or one's friends and kin [21]). Indeed, it has been shown—though with a different paradigm—that observers of unfair treatment punish third parties significantly less when they are assured no one will see their decision [22] (however, see [23]). Third, the third-party punishment game is typically conducted with the ‘strategy method’ [24], which requires third parties to repeatedly respond to a series of hypothetical dictator choices—in advance of learning of the dictator's actual choice—that are progressively more (or progressively less) unfair [6]. Such methods can cause subjects to infer that the experimenters expect them to vary their responses according to some feature that varies across the set of repeated scenarios [25]. Consequently, owing to a well-known experimental artefact called demand, subjects might feel compelled to punish at least some of the time, calibrating those decisions to the only feature of the dictators’ repeated choices that varies: how unfair they are. This is especially problematic in the standard third-party punishment game because rewarding is not allowed; the only way subjects can vary their responses is to vary their amount of punishment. In a notable exception, Almenberg et al. [26] did add a rewarding option to the typical third-party punishment game (conducted with the strategy method), and a small amount of third-party punishment was observed, on average, when dictators transferred $0 (of $10) to the recipient. We note, however, that subjects in this experiment were informed, before making their decisions, that it was possible they would not be paired with another subject—in such a case, their decisions would not be enacted and they would retain all of their money (i.e. participants’ decisions were somewhat hypothetical; see below). Fourth, the strategy method also involves affective forecasting [27] in as much as it requires subjects to respond ex ante to dictator actions that have not yet occurred. Such behavioural commitments can differ from the actual behaviours people enact after experiencing social situations directly because people frequently weight the features of social situations differently during conscious deliberation than they do after experiencing those social situations in real time [28]. For example, as forecasters, people severely overestimate how upset they would feel by (and subsequently, how much they would attempt to avoid interacting with) someone who had made a racist comment; by contrast, subjects who have actually observed another individual express strongly racist attitudes (versus those in control conditions) respond with relative indifference to the racist individual [29]. Fifth, previous claims that anger is the predominant emotional response of third-party punishers have relied on self-reports of anger in response to hypothetical scenarios [4,6]. Self-reports of anger are typically highly correlated with self-reports of other, similar emotions—including envy [30]. To the extent that the covariation between self-reported anger and self-reported envy is not statistically controlled, estimates of third parties’ anger towards unfair strangers might actually reflect envy, which can also motivate costly punishment in pursuit of goals that are quite distinct from putatively altruistic goals such as enforcing norms or delivering deterrence benefits to strangers [31]. Specifically, if third parties’ punishment of individuals who have treated another individual unfairly is motivated by envy, but not by anger, then the mechanisms that motivate third-party punishment might process cues that another individual has obtained better outcomes than the self, rather than cues that an individual has violated a norm or harmed an anonymous third party in whom the punisher has no fitness interest [6,7].

### One-Ness

#### Seemingly altruistic behavior is actually driven by one-ness – confusing the other person’s needs for your own. This means it is actually self-interested.

Cialdini et al. 97 <Robert B. Cialdini, Stephanie L. Brown, Brian P. Lewis, Carol Luce, and Steven L. Neuberg, all of the Department of Psychology, Arizona State University, “Reinterpreting the Empathy- Altruism Relationship: When One Into One Equals Oneness,” Journal of Personality and Social Psychology, 1997, Vol. 73, No. 3. 481-494, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.473.5871&rep=rep1&type=pdf>#SPS

The data patterns of the three studies of this investigation are compelling in their consistency. In each, as relationship closeness increased, so did empathic concern for a needy other. In each, empathic concern significantly predicted willingness to help. And in each, it did so even after the influence of the egoistic factors of personal distress and sadness had been removed. More telling for the purposes of this research, however, was a fourth type of consistent internal replication: Invariably, when a nonaltruistic factor that covaries with empathic concern was introduced to the analyses, it reduced the impact of empathic concern to nonsignificance. That nonaltruistic factor, oneness, reflects a sense of interpersonal unity, wherein the conceptions of self and other are not distinct but are merged to some degree. The implications of these results for the empathy-altruism model are considerable. If the circumstances specified in the model as leading to truly altruistic acts (interpersonal closeness and perspective taking instructions) are the same circumstances that enhance the merging of self and other, as has been shown in the present studies as well as earlier studies (Aron et al., 1991; Aron et al., 1992; Davis et al., 1996), **then one can doubt whether those helpful acts reflect the selflessness required of true altruism.** As even the proponents of the model admit, **if self and other are not sharply distinct in a helper's mind, it is not possible to separate egoism from altruism in a helper's motive** (Batson, 1987; Batson, in press; Batson & Shaw, 1991). After all, as the self and other increasingly merge, helping the other increasingly helps the self. Moreover, one can doubt the empathy-altruism hypothesis even further when, as we have demonstrated, oneness both covaries with empathic concern and is the functional mediator of helping when the two factors are considered simultaneously. That is, although relationship closeness elevated the levels of both factors, only one (perceived oneness) predicted helping when the influence of the other factor was controlled. Overall, then, our findings suggest that empathic concern may have only appeared to mediate aid in much prior research because it is a concomitant of **perceived oneness**, a construct that offers a **nonaltruistic path to such aid**.

#### In-group preference in helping behavior proves that one-ness and self-interest actually drive altruistic behavior

Stürmer et al. 06 <Stefan, Alexandra Kropp, Birte Siem, Christian-Albrechts-Universität zu Kiel, and Mark Snyder, University of Minnesota, “Empathy-Motivated Helping: The Moderating Role of Group Membership,” PSPB, Vol. 32 No. 7, July 2006 943-956, DOI: 10.1177/0146167206287363>#SPS

The main objective of the two laboratory experiments presented here was to test a group-level perspective on empathy-motivated helping. Informed by an evolutionary perspective on human altruism (e.g., Burnstein et al., 1994; Cunningham, 1986; also Park & Schaller, 2005) and building on the idea of psychological essentialism (e.g., Medin & Ortony, 1989; Rothbart & Taylor, 1992), our group-level perspective suggests that similar group membership between the helper and the target strengthens the role of empathy in helping, whereas dissimilar group membership renders empathy motivated helping less likely. In line with this perspective and confirming our specific Empathy × Group Membership Moderation hypothesis, each experiment demonstrated that empathy had a stronger effect on helping when the helpee was an ingroup member than when the helpee was an outgroup member. Including Stürmer et al.’s (2005) studies, the Empathy × Group Membership Moderation hypothesis has thus been confirmed in four different studies employing different research methodologies (field research vs. controlled experimentation) and focusing on different intergroup contexts (natural vs. artificial groups) and different helping criteria (helping intentions vs. actual help). In addition, other researchers have observed similar ingroup/outgroup differences in empathy-motivated helping (e.g., Penner & Finkelstein, 1998). The convergence of empirical evidence thus speaks strongly and persuasively for the validity and generalizability of a group-level perspective on empathy. Some researchers have hypothesized that empathy may affect helping across ingroup/outgroup boundaries (e.g., Batson et al., 1997). Given the research reported earlier (including the present experiments), it seems more likely, however, that as ingroup/outgroup distinctions are salient, empathy-motivated helping is typically restricted to “us,” whereas empathy-motivated helping across group boundaries to “them” is less likely. From our group-level perspective, one could argue, however, that the likelihood of empathy-motivated outgroup helping increases as the outgroup is perceived as relatively similar to the ingroup. This should be so because perceptions of intergroup similarities should facilitate the attribution of a common essence shared by members of both the ingroup and the outgroup. In fact, a recent experimental study in the context of intercultural helping provides encouraging evidence for this reasoning (Siem & Stürmer, 2005). Some may wonder why our experiments—which demonstrated significant ingroup/outgroup differences in the role of empathy—did not also show significant ingroup/outgroup differences in the amount of help provided or in the strength of the intentions to do so. With regard to this issue, it should be taken into account that our experiments tested the effects of ingroup/outgroup categorizations in a very benign contact situation (in which students conversed with a fellow student in a research laboratory). As documented by previous research, overt outgroup discrimination in helping in such situations is rare (see Saucier et al., 2005). For the present experiments it seems quite possible for instance that the benign nature of contact in our experiments facilitated motivational processes that led participants to help outgroup members despite a lack of empathic motivation to do so (e.g., normative considerations and/or the desire to appear unprejudiced, Gaertner and Dovidio, 1977; Pryor et al., 2004). In intergroup contexts that are marked by conflict and animosity, such “compensatory” processes may be less likely to produce intergroup helping. Accordingly, in such contexts the consequences of the lack of empathic motivation should be far more severe, with outgroupers being unlikely to be helped or even actively discriminated against in helping. The design of our experiments on empathy does not allow us to precisely delineate the processes that led our participants to help an outgroup target, and we acknowledge this as a major limitation of the present work. It is telling, however, that in both experiments none of the “need-related” emotions (empathy, sadness, distress) that were in the focus of our research proved as a significant predictor of outgroup helping. At a more general level, this observation falls in line with previous research suggesting that when people contemplate offering help to an outgroup member they may be generally more hesitant to let themselves be guided by spontaneous experiences and base their decision on systematic and controlled information processing instead (e.g., Pryor et al., 2004). Before closing, we also wish to comment on important implications of our experiments for research on the relationships among empathy, interpersonal oneness, and helping. Cialdini and colleagues suggested that empathy serves merely as an emotional signal for interpersonal oneness, and that it is the perception of oneness and not empathy that ultimately promotes helping (Cialdini et al., 1997; Maner et al., 2002). In fact, in both Experiments 1 and 2, in the ingroup conditions, our measures of empathy and interpersonal oneness were positively correlated, rs ≥ .33, ps ≤ .064. Moreover, in both experiments, in the ingroup conditions, interpersonal oneness emerged as a unique predictor of helping intentions (or helping) even when empathy, sadness, and distress were considered as additional predictors. However, in contrast to Cialdini et al.’s (1997) perspective but in line with our reasoning, in these analyses empathy was also a significant and unique predictor of helping intentions (or helping). Based on the present findings (and other research demonstrating a unique role of empathy in helping, e.g., Batson et al., 1997; Stürmer et al., 2005, Study 2), instead of assuming that empathy serves merely as a signal for oneness, it could be argued that empathy and interpersonal oneness may in fact represent two related but distinct sources of people’s motivation to help. Cialdini and colleagues investigated the role of oneness in the context of cues indicating relationship closeness (Cialdini et al., 1997; Maner et al., 2002), whereas our own research focused on helping a stranger who happened to be an ingroup or an outgroup member. One might speculate then that the relevance of interpersonal oneness on one hand and feelings of empathy on the other hand in helping is contingent on the salience of different kinds of relationship cues, with oneness-based helping being more closely tied to cues indicating familiarity and close interpersonal relationships and empathy-based helping being more closely tied to perceptions of selfother similarity. CONCLUSION A starting point of the present research was the proposition that the motivations for helping “us” versus helping “them” are often of a fundamentally different nature (see Dovidio et al., 1997; Omoto & Snyder, 2002; Simon et al., 2000; Stürmer et al., 2005). Our two experiments clearly substantiate this proposition in that they show that even in benign intergroup encounters, **empathy is “deactivated” as a significant motivator of helping outgroup members.** Our group-level perspective suggests that the role of empathy in helping is contingent on perceived group-level (dis)similarities. Accordingly, intervention programs designed to emphasize commonalities rather than differences between groups could provide promising opportunities to increase empathymotivated outgroup helping.

#### One-ness is the best explanation for helping

Maner et al. 02 <Jon K. Maner, Carol L. Luce, Steven L. Neuberg, Robert B. Cialdini, Arizona State University, Stephanie Brown, University of Michigan, Brad J. Sagarin, Northern Illinois University, “The Effects of Perspective Taking on Motivations for Helping: Still No Evidence for Altruism,” PSPB, Vol. 28 No. 11, November 2002 1601-1610, DOI: 10.1177/014616702237586>#SPS

GENERAL DISCUSSION The present study met all of the methodological requirements proposed by Batson (1997) and Neuberg et al.(1997) to test the links between perspective taking, empathic concern, and helping. Results of the present study supported our predictions. First, the data were consistent with a model in which experimentally increasing oneness should increase helping, whereas instructing participants to take the target’s perspective should only increase helping when participants have no explicit information about oneness. Second, the source of altruistic motivation—empathic concern—did demonstrate a significant zero-order relationship with helping. Indeed, the zero-order relationship between empathic concern and helping was nearly as strong as the relationship between perceived oneness and helping. However, the empathy-helping relationship disappeared when statistically controlling for the set of nonaltruistic motivators. Finally, our data indicate that helping was functionally mediated by only nonaltruistic constructs (perceived oneness, nonempathic negative affect) and not by empathic concern. Note that the pattern of helping means across conditions closely mirrored the pattern of perceived oneness across conditions and did not mirror the pattern of empathic concern across conditions (see Tables 1 and 2).Accordingly, the same set of contrast weights significantly predicted both perceived oneness and helping, whereas the contrast weights that so nicely predicted the effects of perspective taking on empathic concern failed to predict helping. Thus, **these data pose a substantial challenge to the empathy-altruism hypothesis.** It is important to note that if we had only met Batson’s (1997) methodological criteria and had not also included a set of alternative motivator variables, the results would have appeared to support the empathy-altruism hypothesis. This is precisely why Cialdini et al. (1997) called into question data previously used to support the empathy-altruism hypothesis. Helping outcomes can erroneously appear to be motivated by empathic concern, and thus appear to be altruistic, if researchers do not also follow Neuberg et al.’s (1997) major methodological requirement: Examine the empathy-helping relationship while statistically controlling for a full set of plausible nonaltruistic mediators. In meeting both sets of methodological requirements, the current work takes a meaningful step forward in resolving the previous debate. The present study also takes a step forward in identifying issues relevant to the measurement of potential affective mediators of helping, in particular the construct of empathic concern. Our analysis suggests that researchers can meaningfully measure empathic concern but that doing so may not be as simple as measuring responses to items purported to indicate other-oriented emotion, as researchers on both sides of the empathy-altruism debate have done in the past (e.g., Batson et al., 1983, 1991, 1997; Batson, Turk, Shaw, & Klein, 1995; Cialdini et al., 1997). The present data suggest that such items tap into both empathic concern and more general negative affect and that a meaningful measure of other-oriented emotion can be achieved only once the more general affective component is removed. The general-specific measurement model we present is a step toward understanding how to meaningfully measure and differentiate such affective states.

## Biology

### Evolution

#### Evolutionary explanations of altruism confuse proximate and ultimate causes

Scott-Phillips, Dickins, and West 11 <Thomas C., School of Psychology, Philosophy and Language Sciences, University of Edinburgh; Thomas E., School of Psychology, University of East London, Centre for Philosophy of Natural and Social Science, London School of Economics; Stuart A., Department of Zoology, Oxford University; “Evolutionary Theory and the Ultimate–Proximate Distinction in the Human Behavioral Sciences,” Perspectives on Psychological Science, Vol 6, Issue 1, 2011, http://journals.sagepub.com/doi/10.1177/1745691610393528>#SPS

Proximate Phenomena Used to Address Ultimate Questions One area in which inclusive fitness theory has been very fruitfully applied is in the evolution of cooperation, in which it has been used to explain cooperation in an extremely wide range of organisms. **The central theoretical question for the evolution of cooperation is why an individual would perform a behavior that is beneficial to another** ([Hamilton, 1964](http://journals.sagepub.com/doi/10.1177/1745691610393528)). This is an ultimate level problem—it asks why such behaviors should exist at all. It thus demands an ultimate level solution. However, **much of the human literature has suggested various proximate mechanisms as solutions to this ultimate problem**. These include religion ([Johnson, 2005](http://journals.sagepub.com/doi/10.1177/1745691610393528); [Johnson & Krueger, 2004](http://journals.sagepub.com/doi/10.1177/1745691610393528)); a concern for praise and blame ([Ellingsen & Johannesson, 2007](http://journals.sagepub.com/doi/10.1177/1745691610393528)); costly ritual ([Power, 2009](http://journals.sagepub.com/doi/10.1177/1745691610393528)); and psychological predispositions to reward cooperators and punish noncooperators, otherwise known as strong reciprocity ([de Quervain et al., 2004](http://journals.sagepub.com/doi/10.1177/1745691610393528); [Fehr & Fischbacher, 2003](http://journals.sagepub.com/doi/10.1177/1745691610393528); [Fehr & Rockenbach, 2004](http://journals.sagepub.com/doi/10.1177/1745691610393528)). All of these describe proximate phenomena: They explain how cooperation might work. For example, one suggestion ([Johnson, 2005](http://journals.sagepub.com/doi/10.1177/1745691610393528); [Johnson & Krueger, 2004](http://journals.sagepub.com/doi/10.1177/1745691610393528)) is that the fear of supernatural punishment motivates individuals to behave cooperatively. This fear and the subsequent motivation are psychological phenomena—they explain (if the hypothesis is correct) how humans come to behave cooperatively. They do not, however, explain why such behavior is present in the population. It may, for example, be the case that individuals who behave cooperatively because of the fear of supernatural punishment have lower inclusive fitness than those who do not fear punishment and hence do not behave cooperatively—in which case such fear would be selected against. To explain why such fear exists in the first place, we must make reference to inclusive fitness effects. Further examples are numerous. Here is one: “Human altruism is considered an evolutionary puzzle… here, we investigate the hypothesis that altruism is caused by feelings of shame and pride and that these feelings are accentuated by others' verbal evaluation” ([Ellingsen & Johannesson, 2007](http://journals.sagepub.com/doi/10.1177/1745691610393528), p. 100). Similarly, “Why [from an evolutionary perspective] are unrelated members of human social groups who only infrequently interact so ready to cooperate with each other?… The only medium for securing such cooperation… is costly ritual” ([Power, 2009](http://journals.sagepub.com/doi/10.1177/1745691610393528), p. 257). In both of these cases, the question is posed from an evolutionary (i.e., ultimate) perspective, but a proximate mechanism is then offered as an answer. Another example is the literature on strong reciprocity. Strong reciprocity is defined as a combination of “a predisposition to reward others for cooperative, norm-abiding behaviors… [and] a propensity to impose sanctions on others for norm violations” ([Fehr & Fischbacher, 2003](http://journals.sagepub.com/doi/10.1177/1745691610393528), p. 785)—this defines it as a proximate mechanism. **Yet it is offered as a solution to the ultimate problem of why humans cooperate** (West, El Mouden, & Gardner, in press). The neurological work on strong reciprocity makes the same error (West et al., in press). It is not that the above-cited studies are worthless; on the contrary, they can help explain how cooperative behavior can lead to direct and indirect fitness benefits. Our point is only that they describe how cooperation might work and not why it exists. Similar issues occur in the literature on the evolution of language ([Scott-Phillips, 2007](http://journals.sagepub.com/doi/10.1177/1745691610393528)).

### Ontogenetics

#### Research into infants show that children only help in response to social cues – meaning altruism can’t be innate in human nature.

Barragan and Dweck 14 <Rodolfo Cortes Barragan and Carol S. Dweck, Department of Psychology, Stanford University, “Rethinking natural altruism: Simple reciprocal interactions trigger children’s benevolence,” PNAS December 2, 2014. 111 (48) 17071-17074, https://doi.org/10.1073/pnas.1419408111>#SPS

We began our investigation by examining whether a simple experience with reciprocity might serve as a critical cue for altruism. We found support for this contention. Reciprocal interactions triggered high levels of altruistic behavior on the part of 1- and 2-y-olds whereas parallel play did not. Moreover, **we found this with an age group in which altruism had been depicted as occurring naturally and without much need for social input** (8). We further found that an even shorter reciprocal interaction elicited substantially more altruism than a parallel interaction in preschoolers and that these interactions yielded not only the enactment of altruism but the expectation of it from others. Thus, consistent with anthropological (18, 19), economic (20, 21), evolutionary (22, 23), philosophical (24), psychological (25⇓–27), and sociological (28, 29) theories of human contractual processes, our young participants were, in a sense, capable of drawing broad inferences about the benevolent norm of the situation on the basis of reciprocal patterns of behavior. It is interesting to note that although prior research has shown that explicitly rewarding or encouraging helping seems to lower it (30, 31), in none of our studies did we directly model, teach, or reward altruistic actions for children. Rather, we exposed children to certain cues about the local “culture.” In the parallel play condition, children may have learned that, in this culture, people engage with each other by carrying out their own activities. By contrast, **in the reciprocal play condition, they may have learned that people engage with each other by being responsive to each other’s needs.** As such, our findings build on past research showing that infants’ helping may be triggered by actions that people carry out, such as intentionally providing toys to the infant (32). However, it is important to note that the experimenter in our studies provided toys in both the reciprocal and the parallel play conditions. What varied across the conditions was whether the experimenter and the child engaged reciprocally, not whether the experimenter showed an intention to provide. Our research also builds on past research showing that preschoolers understand the principle of economic reciprocity and use it to guide their decisions about what others should do (33). Our findings suggest that children will be more likely to act on the principle of economic reciprocity if they have themselves undergone even a simple reciprocal interaction. More generally, and consistent with research that shows that children are avid learners who eagerly look for clues about how the physical world operates (34, 35), our data build on theories of children’s social learning (36, 37) by documenting that reciprocal interactions trigger the enactment and expectation of altruism in young children. That is, after an experience with reciprocity, children seem to construct a community characterized by care and commitment. Thus, **the notion that socialization has little or no part to play in early occurring altruism** (8) **becomes less plausible.** Experimenters, parents, teachers, and others who regularly interact reciprocally with children may be implicitly communicating to children that in these contexts people help one another. Fortunately, children seem quite eager to adopt this “benevolent social contract.”

#### Children will sometimes help, but only in response to help – they believe in exchange, not altruism.

Dunfield and Kuhlmeier 10 <Kristen A. and Valerie A., Queens University, “Intention-Mediated Selective Helping in Infancy,” Psychological Science, 21(4), 523–527, 10.1177/0956797610364119>#SPS

Even in the earliest instances of helping behavior, infants are not indiscriminate, but instead they direct their help on the basis of previous interactions with individuals. More specifically, the underlying intentions of their interactants appear to strongly influence their subsequent prosocial behavior, more so in some cases than the actual outcomes of the interactions. Thus, the intention reading that has of late been documented in infancy (e.g., Behne et al., 2005, Meltzoff, 1995; Woodward, 1998) appears to influence helping behavior not just by determining someone’s overall goal (e.g., to obtain an out of reach object) but also by mediating decisions as to the target of helping behavior. This study also allows us to form some initial proposals regarding more specific aspects of the mechanism. Infants do not simply help all individuals except those who have previously displayed an overt unwillingness to provide. In Experiment 3, infants preferentially helped an individual who had previously willingly provided toys over an individual who had ultimately provided the toys but not indicated any intention to do so (or not to do so). Infants, thus, do not solely avoid helping previously unwilling individuals but will selectively help those who have shown a willingness to provide. Indeed, the extent to which infants avoid helping previously unwilling providers remains unclear, but note that the propensity to use negative information to inform subsequent action, and perhaps even a bias to do so in contexts such as social referencing, has been documented for infants (see Vaish, Grossman, & Woodward, 2008, for an extensive review). Further, the actresses the infants preferred to help were not simply “nice people” in relation to others, as all the actresses were friendly (e.g., all smiled and spoke in pleasant tones). Instead, it was their willingness to provide (even if unable to) that set them apart. It is possible that this served to define them as good partners, those with whom it might be beneficial to enter into a reciprocal helping relationship. Taken together, these findings suggest that specificity in the form of intentionbased reciprocity can be observed in early helping behavior, in turn indicating that some of the foundations for the complex nature of adult other-oriented behaviors are in place in infancy. Future research examining the situations under which aid is selectively provisioned on the basis of intentions versus outcomes will help to identify the unique cognitive mechanisms that allow for the ubiquitous and essential social cooperation of humans.

#### Warneken is wrong – altruism is only socially constructed, not fundamental

Dunfield and Kuhlmeier 10 <Kristen A. and Valerie A., Queens University, “Reply to Warneken: Social experience can illuminate early-emerging behaviors,” PNAS March 10, 2015. 112 (10) E1053; published ahead of print February 18, 2015. <https://doi.org/10.1073/pnas.1500252112>

Warneken acknowledges that early human helping is a “rich social interaction embedded in the context of social experiences more generally” (2). According to his past writings, Warneken simply does not believe that this rich social context plays a role in the emergence of early altruism: “Infants show altruistic tendencies at an age when socialization could not yet have had a major impact on their development” (3). Our research does not speak to possible biological contributions to altruism, but it does make a social contribution far more plausible (1). **Our studies revealed high levels of altruism in young children only after a reciprocal interaction with the experimenter.** When children had had highly similar and friendly—but nonreciprocal—play experiences with the experimenter, subsequent levels of altruism were alarmingly low, even though the experimenter’s bids for help were strong. Others, too, have found relatively low levels of altruism in young children (4). In our report (1) we go on to examine how and why reciprocal interactions may trigger altruism and to suggest how early reciprocal interactions may plausibly serve as a basis for altruism. Warneken questions our conclusions by contending that because chimpanzees sometimes help, “helping emerges in the absence of any relevant socialization experiences and adoption of human social norms” (2). This is a premature conclusion. It is quite possible that subtle reciprocal interactions in the developmental history or in-laboratory experiences of nonhuman primates foster their ability to help. Indeed, other theories of development suggest that reciprocal interactions in the early life of mammals could explain more sophisticated forms of interaction, such as helping (5). Warneken (2) also describes his other work, which showed that toddlers cooperated with a puppet even when that puppet failed to cooperate with them. As we suggest in our report (1), that study involved an extensive (likely reciprocal) warm-up experience with the puppet. It is possible that this lengthy initial interaction primed children’s altruism and overrode the uncooperative behavior of the puppet. In the end, Warneken (2) appears to concede that social interaction plays an important role by questioning the particular type of social interaction we have introduced. But if altruism is so inbred and automatic, why should subtle differences in the type of social interaction that precedes it be so critical? We seem to have made our point. Many researchers, like Warneken, readily interpret early-emerging behaviors as inherent and unlearned because there has been no direct teaching of the behaviors. By contrast, we suggest that there are numerous routes to early learning, aside from direct tuition. For example, 1-y-olds have rich mental representations of their attachment relationships with caregivers, despite no direct tuition (6). (Indeed the role of learning mechanisms in language acquisition was underappreciated until statistical learning was discovered.) The task before us is to understand the subtle forms of socialization embedded in social experiences and the capacity of infants to learn from those experiences.

### Neurology

#### Altruism does exist – but its explained entirely by neurological evidence and varies drastically from person to person – means it can’t be a primary driver

Cutler and Campbell-Meiklejohn 16 <Jo, PhD student in Psychology, University of Sussex, and Daniel, Senior Lecturer in Psychology, University of Sussex, “Are you a true altruist or driven by self-interest? Brain scan may give verdict,” The Conversation, March 3, 2016, https://theconversation.com/are-you-a-true-altruist-or-driven-by-self-interest-brain-scan-may-give-verdict-55545>#SPS

The reason why we help others at a cost to ourselves has long presented a puzzle for scientists. Why do some of us do it more than others? And are we doing it because we are truly moved by the suffering of others or simply because we feel we ought to return a favour or even get something in return? Looking at behaviour alone, it can be hard to tell. Both empathy and the principle of reciprocity – giving to return a favour or expecting others to do so – are proposed explanations for altruism which have been impossible to separate until now. Using functional magnetic resonance imaging (fMRI), which measures blood flow changes in the brain, a new study suggests that specific differences in connectivity between brain regions can predict whether someone is an empathy-driven altruist, a reciprocity-driven altruist – or just selfish. In the experiment, 34 female participants were divided into two groups. Those in the “empathy” group witnessed an actor receive painful electric shocks – and received shocks themselves (so they knew it hurt). In the “reciprocity” group, participants were paired up with actors who kindly paid money so the participant received fewer shocks (although both groups received the same number of shocks overall). Next, their brains were scanned. During the scanning, participants were asked to split a sum of money between themselves and another person. For the empathy group, the other person receiving the money was sometimes the partner they saw shocked. In the reciprocity group, the person was sometimes the partner who paid for the participant to receive fewer shocks. At other times, participants were simply asked to split the cash between themselves and a neutral person who neither received shocks nor did anything nice. The researchers could therefore divide the participants into those empathy-driven altruists and reciprocity-driven altruists based on the first part of the experiment. They could also use the way participants split the money in the second part to identify selfish individuals among these participants. Unsurprisingly, the initial analysis showed that participants gave, on average, larger sums of money to the empathy and reciprocity partners than to the neutral partner – and that both groups were equally generous. Those that most regularly chose splits involving more money for themselves than the other were classified as “selfish”. **But this was just the starting point.** The researchers used a complex and sophisticated follow-up to gain deeper insight. By looking at the timing of activity in the anterior cingulate cortex (known for a host of functions from pain and conflict to learning), the anterior insula cortex (associated with arousal and emotion) and the ventral striatum (associated with rewards and learning), the researchers created models of how information was passed between these areas. Then a computer algorithm tried to guess, based on these models, whether an individual’s altruistic decision had been motivated by empathy or by reciprocity. The high accuracy of these guesses at 77% shows the two groups of participants had brain activity patterns that differed enough to classify. In empathy-driven altruism, the anterior insula (emotion and arousal) and ventral striatum (rewards) showed a lower than average connectivity, while reciprocity-driven altruism showed increased connectivity between these regions. Connectivity in this sense can be imagined as how much one area is “talking to” another. Although the functions of these areas are broadly known, the meaning of changes in connectivity is still difficult to interpret. The results also showed differences in brains of those who had been classified as selfish or altruistic based on their decisions. Selfish individuals showed lower than average connectivity from the anterior cingulate cortex to the anterior insula whereas altruists had increased connectivity between these regions. Can we learn to be more altruistic? When it comes to implications, the differences between primarily selfish or primarily altruistic participants may be the most important finding. Inducing empathy, by seeing someone shocked, increased giving and associated neural connectivity for selfish individuals – they were more generous to the shocked partners than to the neutral person. The altruistic people, however, shared just as much with the neutral person as the shocked partner. The opposite was true for the reciprocity effect: increased giving to the partner who paid to prevent their shocks was seen in altruistic but not selfish participants. One could speculate that this implies that altruistic participants are already giving because of empathic motivation, so increasing empathy makes no difference – they are at their “empathy capacity”. Similarly, selfish participants may already be acting due to motivations more likely to benefit themselves too, such as reciprocity. Research on altruism regularly concludes that people have an empathetic motivation but this paper suggests potential for future studies to check whether this is the case for each individual participant. The authors also open doors to more specific measures and targets for further research on reciprocity and empathy. The paper shows the importance of analysing subtle differences in brain communication rather than overall activity. Looking at different brain regions working together, rather than in isolation, can identify previously elusive psychological concepts, such as underlying motivations. Future research is needed on whether these increases in altruism and neural connectivity could last, perhaps with ongoing training. For example, if the techniques used to induce empathy in the study could be employed in some sort of treatment for antisocial behaviour. However, charities can already make the most of the current findings. They suggest empathy-inducing appeals may be most effective for new supporters, who are not yet “altruistic enough” to donate. Existing supporters, who are already altruistic, may respond more to receiving a token gift they feel they can reciprocate by increasing their donations. The effectiveness of these techniques, already used by many charities, may be explained by the findings. But with limited resources, new insight into cognitive processes that might be harnessed by appeals could help society be a bit more generous.

## Political Science

### Attitudes

#### Attitudes in politics are clearly motivated by self-interest

Weeden and Kurzban 17 <Jason Weeden, Pennsylvania Laboratory for Experimental Evolutionary Psychology, Robert Kurzban, University of Pennsylvania, “Self-Interest Is Often a Major Determinant of Issue Attitudes,” Advances in Political Psychology, Vol. 38, Suppl. 1, 2017, https://onlinelibrary.wiley.com/doi/abs/10.1111/pops.12392>#SPS

If we were to stress one thing from our work on political attitudes, it would be this: Across lots of issues, there are substantial domain-specific connections between public opinion and the relevant concrete details of people’s own lives. We’ve covered several examples in this review. When it comes to short-term economic issues, short-term economic factors are often predictive (as when the unemployed support more robust measures to benefit the unemployed, or when those experiencing economic decline move generally to the left on economic issues, or when those particularly affected by tax policies oppose efforts to increase the taxes of people like themselves). When it comes to more general views on income redistribution and government support for the poor, income itself is a major predictor, along with open to them. Most particularly, motives are typically very hard to study, and there are various ways to try to claim that apparently self-interested opinion patterns don’t actually arise from self-interested motives. These matters are tremendously complex, and we don’t claim to have settled them once and for all. But we do claim to have provided substantial evidence that undercuts a recurring objection to self-interest claims, namely, the objection that there really aren’t substantial, plausibly self-interested patterns present in public opinion that need explaining in the first place. In other words, for example, one should feel free to explore alternative motivational accounts for why poorer people would often prefer more robust governmental income redistribution, but it would be misleading to state that poorer people don’t, in fact, have such preferences on average. (And, as we discussed, showing that such a relationship can be mediated by a third variable doesn’t necessarily show that the relationship never existed.) In exploring motivational accounts, we think it’s important to pay attention to psychological research suggesting that people are often consciously ignorant—and indeed often strategically self-deceptive—regarding their own motives (Kurzban, 2010; Weeden & Kurzban, 2014). This situation creates additional hurdles in studying self-interested political views. Asking questions about selfinterest too directly can trigger self-presentational maneuvers, including self-deceptive ones, in which individuals routinely seek to encourage others to view the individual as reasonable and generous rather than self-serving. Similarly, asking questions about whether respondents think that certain policies are beneficial to society as a whole or whether they view the beneficiaries of a given policy to be deserving are likely to access self-presentational discourse strategies rather that straightforward internal motives. Thus, we think that it’s generally appropriate in studying self-interested political views to piece together a more objective view of respondents’ personal interests (through demographic and related measures) rather than asking them to directly characterize their motives or their views on who benefits from particular policies. To be clear, we’re not claiming that self-interest is always and everywhere the only or plainly dominant determinant of issue attitudes. Elsewhere, we have been explicit about some issue areas that don’t obviously fit in an interest-based perspective. One example was our examination of Kinder’s (1998) claim about working women: “Women employed outside the home do not differ from homemakers in their support for policies intended to benefit women at work” (p. 801). This was the one area from Kinder’s list that we tested that seemed to be plainly accurate (Weeden & Kurzban, 2014, chap. 2). We have also discussed policies involving the environment, the military, and right-to-die as examples of areas where it seems to us that a self-interest-based perspective can advance the ball only in limited ways (Weeden & Kurzban, 2014, chap. 10). We similarly don’t deny that there are relevant factors beyond those involving concrete aspects of everyday life. While we do have substantial doubts about the value of values (given our concerns that they are likely not independent causes of closely related policy preferences, even if they can have really large correlations), we view party identification and ideology as partial causes (in addition to being partial effects) of issue opinions (Weeden & Kurzban, 2014, 2016). But we have also noted that the alignment of lots of issue positions into left-or-right bundles is both a recent phenomenon and mostly limited to college-educated Whites (Weeden & Kurzban, 2016). To be sure, this has led to demographic patterns among college-educated Whites that don’t seem obviously interest-based—the key examples involve the recent rise of substantial contrasts across a very wide range of issues between Christians and non-Christians and between those with only four-year degrees and those with graduate degrees (Pew Research Center, 2016; Weeden & Kurzban, 2016). These kinds of patterns supplement, but do not supplant, the usual domain-specific interest-based demographic patterns. While the authors of The American Voter Revisited stated that the “current scholarly consensus holds that self-interest is not a major determinant of issue attitudes” (Lewis-Beck et al., 2008, p. 197), our view is that self-interest (or something closely related to it) is often a major determinant. Yes, it’s complicated. Yes, there are exceptions. Yes, there are other major determinants. Self-Interest and Issue Attitudes 87 In fact, there remains much work to be done in unravelling when and how self-interest (or closely related factors) impact individuals’ policy views. But in order to move forward with this work, researchers need to acknowledge that self-interest-minimizing claims are an inappropriate point of departure. **Researchers need to recognize consistently that there are many domain-specific patterns in public opinion that are strongly suggestive of something closely related to self-interest and that many of these patterns are grounded in predictive variables that have relatively clear claims to being early in the causal flow.**

#### Self-interest is the clearest explanation of political attitudes

Weeden and Kurzban 17 <Jason Weeden, Pennsylvania Laboratory for Experimental Evolutionary Psychology, Robert Kurzban, University of Pennsylvania, “Self-Interest Is Often a Major Determinant of Issue Attitudes,” Advances in Political Psychology, Vol. 38, Suppl. 1, 2017, https://onlinelibrary.wiley.com/doi/abs/10.1111/pops.12392>#SPS

The prior section reviewed some ways in which self-interest minimizers have selectively applied evidentiary standards and narrowed definitions to reduce the appearance of self-interest effects. This section reviews evidence that, nonetheless, these self-interest effects often persist. Even in its narrow form, there are widely acknowledged empirical exceptions to the general selfinterest-minimizing claim. These exceptions pertain to policy issues in which “the material benefits or harms ... are substantial, imminent, and well publicized” (Kinder, 1998, p. 802), where “the material outcomes ... are very clear, very large, and very imminent” (Taber, 2003, p. 448), that “have large, clear, and certain effects on an individual’s interests” (Huddy, 2013, p. 741), “that offer unambiguous benefits or impose tangible costs” where “the policy being considered was clearly going to help or hurt some elements of the population more than others” (Chong, 2013, pp. 102–103). Two commonly cited examples involve taxes. Sears and Citrin (1985) found that homeowners supported a local proposal to cut property taxes more than nonhomeowners. Green and Gerken (1989) found that smokers opposed cigarette taxes more than nonsmokers. Similar examples appear regularly. Doherty, Gerber, and Green (2006) found that “lottery winnings have a profound effect on views toward estate taxes, which touch directly on respondents’ material interests” (p. 456). In “a striking example of the power of self-interest to disrupt and transform political views,” Erikson and Stoker (2011) examined the political effects of randomly assigned draft lottery numbers in the Vietnam era: “Those who were ... handed an adverse draft number tended to turn against the war and against the new draft policy’s champion, President Richard Nixon .... In almost every comparison, lottery status outstrips preadult party identification in accounting for the political views draft-eligible men came to hold by their mid-twenties” (pp. 235–236). Looking at changes in opinions about the Affordable Care Act between 2008 and 2010, Henderson and Hillygus (2011) found that those “expressing the greatest concern about medical expenses are far less likely to join the growing opposition to universal coverage” (p. 954). Hacker, Rehm, and Schlesinger (2013) found that “[d]irect economic experiences are strongly correlated with support for risk-buffering social policies—at times rivaling partisanship and ideology as correlates” (pp. 32–33). Margalit (2013) found “compelling evidence of the strong impact of personal economic circumstances, particularly the loss of employment, on individuals’ preferences on welfare spending” (p. 98). Nteta (2013) found “in line with the expectations of self-interest theory that class membership alongside a number of objective and subjective measures of self-interest are among the most consistent determinants of African American, and to a lesser extent White, opinion on restrictive immigration policies” (pp. 160–161). Owens and Pedulla (2014) found that “preferences for government redistribution do respond to changes in employment and income, supporting theories of political preference formation that emphasize the material underpinnings of political preferences” (p. 1104). The narrow definition of “self-interest” equates it with short-term economic outcomes, yet there are numerous exceptions to the general self-interest-minimizing claim. When we make both the 72 Weeden and Kurzban narrow definition and the exceptions explicit, we end up with something like this: Short-term economic self-interest doesn’t often matter in political issue opinions unless the issue in question is a shortterm economic one where people recognize the short-term economic implications. This is very close to a case in which the exception swallows the rule. In short, we agree that short-term economic self-interest tends only to apply to short-term economic issues where people recognize the relevant implications. We also think that self-interest based on having a certain religious identity is particularly relevant when it comes to issues based on religious identity in which people recognize the implications for people with their own religious identity. We further think that self-interest based on having a certain sexual and reproductive lifestyle doesn’t often matter in political issue positions, unless the issue is a sexual/reproductive one where people recognize the implications for their own sexual/reproductive lifestyle. The point is that, taken together, these kinds of statements actually imply a substantial role for self-interest in political issue opinions; it’s just that the particular connections between issues and demographics are often domain-specific. Nonetheless, even on the narrow self-interest-minimizing claim, perhaps the exceptions are greatly outnumbered by nonexceptions. Indeed, extended discussions of the limited role of self-interest typically are supported by several specific examples of the lack of self-interest effects. But we have an empirical worry here as well: Many such claims that animate this discussion are misleading. For example, Kinder’s (1998) review of the determinants of the public’s political opinions in The Handbook of Social Psychology has been influential among psychologists. In it, he produced a list of specific claims that carry much of the weight of the overall argument that self-interest rarely matters much in politics: For the self-interested citizen, then, the question is always and relentlessly, What’s in it for me and my family—what’s in it for me and mine now? Defined in this way, selfinterest is surprisingly unimportant when it comes to predicting American public opinion. ... Consider these examples. When faced with affirmative action, white and black Americans come to their views without calculating personal harms or benefits. The unemployed do not line up behind policies designed to alleviate economic distress. The medically indigent are no more likely to favor government health insurance than are the fully insured. Parents of children enrolled in public schools are generally no more supportive of government aid to education than are other citizens. Americans who are subject to the draft are not especially opposed to military intervention or to the escalation of conflicts already under way. Women employed outside the home do not differ from homemakers in their support for policies intended to benefit women at work. On such diverse matters as racial busing for the purpose of school desegregation, anti-drinking ordinances, mandatory college examinations, housing policy, bilingual education, compliance with laws, satisfaction with the resolution of legal disputes, gun control, and more, self-interest turns out to be quite unimportant. ... American society is marked by huge differences in income, education, and wealth, but such differences generally do not give rise to corresponding differences in opinion. (pp. 801–802, citations omitted) To be sure, Kinder is basing his summary on numerous published studies by many researchers. The studies are complex—and many involve the sorts of self-interest-minimizing maneuvers we discussed above—and summaries like Kinder’s boil these down to rather simple claims that, if true, would indeed help make the case that self-interest is surprisingly unimportant. But are these simple claims actually true? We used General Social Survey (GSS) data to run basic tests of many of these straightforward statements (Weeden & Kurzban, 2014, chap. 2). In most cases, our interpretation of the results contrasts sharply with Kinder’s summary. We found, for example, that 74% of the unemployed thought that it should be the government’s responsibility to provide a decent standard of living for the unemployed, and 57% thought that government spending on unemployment benefits should be increased; these opinions were in marked contrast to people working full time, among whom only 46% and 27% agreed, respectively. These results echo longitudinal studies finding substantial effects of unemployment on economic policy views (Margalit, 2013; Owens & Pedulla, 2014). We therefore resist the conclusion that the “unemployed do not line up behind policies designed to alleviate economic distress.” Similarly, we found large racial differences in views on race-based workplace affirmative action. Some might say that this is a case of group interest and not self-interest, but we went further into individual circumstances. In particular, African Americans who feared losing jobs to Whites supported affirmative action more than African Americans who did not fear such losses; at the same time, Whites who feared losing jobs to African Americans opposed affirmative action more than Whites who did not fear such losses. One might call this a self-interest-laden case of group interest, or, if one prefers, a group-interest-laden case of self-interest. These analyses suggest that affirmative action views are related to “personal harms and benefits.” On the claim about the “medically indigent,” similar to other studies (e.g., Henderson & Hillygus, 2011), we found that poorer people and those lacking medical coverage supported government help with healthcare more than richer people and those who had health coverage. Also in line with other studies (e.g., Wolpert & Gimpel, 1998), we found gun owners to be substantially more opposed to gun ownership restrictions than nonowners. In a final example, while Kinder claimed that differences in income generally do not give rise to differences in opinion, we found that when we looked at opinions relating squarely to whether the government should reduce income differences and provide for the poor, there were in fact substantial correlations with income. These findings align with a large number of studies noting important differences in policy preferences between the rich and the poor (e.g., Gilens, 2012; Hacker & Pierson, 2010). In sum, **while it is often said that self-interest is of minimal importance to issue attitudes, the case is weak.** Such claims rely on a narrow definition of self-interest and on viewing what are surely closely related phenomena (like demographic effects and group interest) as irrelevant or even as evidence against self-interest. In addition, the list of exceptions is substantial, growing, and seems to cut to the heart of the narrowed definition of self-interest. And, further, when we look at self-interestminimizing examples on their face, accepting for purposes of the exercise the narrow definition of self-interest, many of the specific supporting claims are arguably misleading. We therefore conclude that the change that occurred from the original version of The American Voter to its revisited version was not due to basic changes in how Americans choose their positions. Despite efforts to make it disappear, self-interest was visible then and remains visible today. To be clear, we’re not saying that self-interest is the only determinant that matters, or that it always matters, or related extreme positions. But when it comes to issues impacting short-term material positions, short-term material self-interest is often one of the major determinants of individuals’ opinions.

### Motivated Reasoning

#### Your brain is lying to you – even when you attempt to do altruistic things, your brain is secretly doing them because they are in your interest

Weeden and Kurzban 16 <Jason, PhD in psychology from the University of Pennsylvania, Professor at the University of Pennsylvania and recently served as the Editor-in-Chief of Evolution and Human Behavior, “The Hidden Agenda of the Political Mind: How Self-Interest Shapes Our Opinions and Why We Won't Admit It,” Princeton University Press, 2014>#SPS

Our take is neither partisan nor polite, and might make many uncomfortable. Our explanation for political disagreements begins with something obvious but often overlooked: The policies people fight over have real-life consequences that help some people and harm others. In our view, all sides typically seek to advance their interests and are hypocritical in the way they present their views. No side is particularly motivated by being fair or reasonable or public-spirited. Indeed, when it comes to policy disputes, we think that one’s perceptions of what’s “fair” or “reasonable” are themselves typically driven by one’s interests. People are generally neither boobs nor saboteurs, but social animals competing over advantages for themselves, their families, and their social networks. It doesn’t take one very far to divide the country (much less the whole of humanity) into two or three ideological boxes. If one wants to understand the variety of public opinion, one needs to think about specifics. The key, we have argued, is to look at people’s lives and interests, focusing on demographic features that provide clues to the particular outcomes that will help or harm them. On sexual and reproductive issues, differences in Freewheeler and Ring-Bearer lifestyles help determine whether people gain or lose when higher costs are placed on Freewheeler lifestyles—when casual sex carries moral costs, when partying carries legal costs, and when family planning is restricted. These lifestyles influence people’s decisions to affiliate with or avoid religious groups. People’s religious and lifestyle patterns strongly predict their views on issues related to premarital sex, pornography, abortion, birth control, and marijuana legalization. About group-based issues, we proposed that the two key factors in determining people’s competing interests are, first, group identities (race, religion, etc.) and, second, accumulated human capital (education and related cognitive abilities). Analogous to talented African American baseball players in our allegory, people with lots of human capital who are also members of traditionally subordinate groups do better when the rules abolish group-based barriers and give advantages to those with lots of human capital. Analogous to less talented white baseball players in our allegory, people with less human capital do better when advantages are given to their own groups and other groups are held back. People’s views on issues involving sexual orientation, religion, immigration, and race are well predicted by their group identities and levels of human capital. Finally, on economic issues, people differ not only in how much they stand to benefit (or lose) when wealth is redistributed, but also in, first, how much they might need hard-times programs in the future and, second, how much they might rely on their own social groups and private charities when hard times hit. So, while income predicts people’s economic views to a degree, race, age, gender, religion, sexual orientation, and human capital are also important for understanding and predicting preferences for public hard-times programs. **Because people generally adopt issue opinions that advance their multifaceted inclusive interests, they wind up frequently adopting, buffet-like, sets of particular views that fall outside of a simple left-right framework.** When someone’s interests point to “liberal” policy preferences on one set of issues and to “conservative” policy preferences on a different set of issues, that’s usually how things turn out. Focusing on interests points the way to finding people who are typically liberal, typically conservative, typically libertarian, and typically whatever-we-shouldcall-the-opposite-of-libertarian, along with other nameless position profiles that are completely absent from the usual discussions of the political map. We view it as a good sign that our efforts line up with certain aspects of political targeting by campaign professionals, the people who get paid to get such things right. We have tried to add to these perspectives by providing a psychological framework that can reveal interests in play in a wide range of issues (beyond the usual suspects involving economic redistribution). In particular, instead of viewing “social” or “cultural” or “religious” issues as symbolic and disconnected from the concrete concerns of real life, we’ve made the case that battles over sexual lifestyles and social status regimes have real-life effects as concrete as the results of fights over money. Without necessarily knowing the real reasons, across a range of policy areas, people are motivated to seek outcomes that advance the everyday goals of themselves, their families, their friends, and their wider circles of social allies. On that point, we’ve also argued that human minds are designed for spin, to hide their strategic foundations behind socially attractive veneers. The Public Relations Departments of people’s minds craft stories about the benevolent wisdom of their own views and the malevolent idiocy of their opponents’ views, with Spokespersons almost wholly ignorant of the nature of the game. Public political discourse is frequently a battle between prickly Spokespersons fighting over made-up stories that have little to do with the underlying motives of people’s mental Boards of Directors. Admitting that one’s political opponents would often be worse off under one’s own policy preferences interferes with the goal of advancing one’s own agenda. People’s desires to advance favorable policy outcomes typically trump any desire to express coherent views of themselves and others. Observers can predict, with error, to be sure, other people’s political positions and priorities by taking into account the other person’s inclusive interests, considering their religion, lifestyle, sexual orientation, race, immigration status, education, intelligence, income, and so forth, despite the fact that most people are themselves unaware that these interest-relevant features are important in shaping their own views. In fact, most people, most of the time, will strongly deny, for example, that their opposition to abortion has anything to do with suppressing others’ sexual promiscuity. Virtually no one says they favor meritocracy because it helps smart people like themselves beat less-smart people in social competitions. People’s Public Relations Departments don’t let their Spokespersons know such things, let alone say them out loud; they are the kinds of accounts, indeed, that people find insulting, regardless of how well the accounts explain the facts. We think we’ve provided the basics to understand these kinds of political opinions, but we acknowledge the limits of the approach. We don’t want to give the impression that we think our view explains the totality of the expanse of American political opinion. People are, in a word, complicated. We think we’ve given a foundation that is really useful, but it’s obvious there’s more to the story.

# Neg

## General

### Self-fullfilling prophecy

#### The belief in the self-interest assumption causes people to act as if they were self-interested – it is a self-fullfilling prophecy, not a fundamental human drive

Miller 99 <Dale T., Department of Psychology Princeton University, “The Norm of Self-Interest,” The American Psychologist, December 1999 Vol. 54, No. 12, 1053-1060, http://faculty-gsb.stanford.edu/millerd/docs/1999amerpsyc.html>#SPS

With the publication of Leviathan, Thomas Hobbes (1651/1950 ) enthroned self-interest as the cardinal human motive, a that the power accorded self-interest, especially by neoclassical economics, promotes a misleading and impoverished view of the human agent that is as dangerous as it is demeaning. A more adequate view of human agent, dissenters contend, acknowledges the power of many additional, sometimes dominating, motives, such as public spiritedness ( Etzioni, 1988 ; Mansbridge, 1994 ), empathy ( Batson, 1991 ; Kohn, 1990 ), commitment ( Sen, 1977 ), and justice ( Lerner, 1980 ; Tyler, Boeckmann, Smith, & Huo, 1997 ). That the explanatory power of the self-interest assumption has been the major focus of the self-interest debate is understandable, but it has also proven limiting. In particular, it has forestalled consideration of the theory's causal power–a serious omission, as the assumption of self-interest is not simply an abstract theoretical concept but a collectively shared cultural ideology ( Kohn, 1990 ; Lerner, 1982 ; Miller & Ratner, 1996 ; Sampson, 1983 ; Wallach & Wallach, 1983 ) . Demonstrating that the theory of self-interest has causal power–specifically, that it plays a role in its own confirmation–is the main objective of this article. Scientific theories, by generating self-knowledge and self-images, are always at risk of becoming self- fulfilling ( Berger & Luckmann, 1966 ; Gergen, 1973 , 1982 ; Schwartz, 1997 ; Wallach & Wallach, 1983 ) . **Nowhere would this risk seem greater than in the case of the self-interest assumption**, for the image of humans as being predominantly self-interested is central not only to neoclassical economics but to many other influential theories of human behavior, including evolutionary biology, behaviorism, and psychoanalytic theory ( Etzioni, 1988 ; Kohn, 1990 ; Schwartz, 1986 ; Wallach & Wallach, 1983 ) . Indeed, the self-fulfilling properties of the self-interest assumption have frequently been commented on ( Kagan, 1989 ; Kohn, 1990 ; Lerner, 1982 ; Schwartz, 1997 ). How exactly does the assumption that humans are self-interested become a self-fulfilling prophecy? Previous analyses of the self-fulfilling impact of the self-interest assumption have focused on the role that the theory plays in the structuring and configuring of **social institutions** ( Etzioni, 1988 ; Kohn, 1990 ; Lerner, 1982 ; Schwartz, 1997 ). According to these arguments, the image of humans as being self-interested leads to the creation of the kinds of social institutions (e.g., workplaces, schools, governments) that transform the image into reality. The present analysis accepts the premise that a positive feedback loop exists between theory and social structure and that this loop contributes importantly to theory confirmation ( Schwartz, 1997 ). I seek to go further, however, and show that scientific theory also impinges on individual consciousness and interpersonal behavior. More specifically, I attempt to show that the theory of self- interest has spawned a norm of self-interest, the consequence of which is that **people often act and speak in accordance with their perceived self-interest solely because they believe to do otherwise is to violate a powerful descriptive and prescriptive expectation.** Defining the Norm of Self-Interest The norm of self-interest prescribes that people pursue their self-interest, but, more than this, it prescribes that they pursue their self-interest narrowly defined. To satisfy the strictures of the self-interest norm, people's actions must conform to, at least crudely, the strictures of neoclassical economic theory. Acting so as to maximize positive emotions (e.g., pride) or minimize negative emotions (e.g., guilt) while meeting many definitions of self-interested behaviors ( see Mansbridge, 1990b ) does not meet the standards of most rational choice models nor, as we shall see, those of the self-interest norm. The latter requires that the interests motivating people be material ones ( e.g., economic profit). Consider the act of donating to charity as an example. It is often claimed that acts of this type, even anonymous ones, are self-interested because they serve various psychological interests of the self, such as self-esteem enhancement or guilt reduction ( Frank, 1988 ; Mansbridge, 1990b ). Such motivational constructions, however, do not satisfy the logic of most rational choice models (e.g., Mueller, 1979 ; Olson, 1965 ), because the fact that actors "feel good" after contributing to a public good (e.g., a new museum) is of no material consequence to them. Indeed, according to many rational choice models, the only truly rational action taken with respect to public goods is to partake of them without contributing to them– after all, public goods exist whether one contributes to them or not. 1 The norm of self-interest does not go so far as to mandate that people eschew charitable donation in favor of free riding, but it has, as we shall see, led the layperson to feel increasingly uncomfortable donating to charity in the absence of a tangible quid pro quo. The Layperson as Self-Interest Theorist If a scientific theory is to exert power over people's behavior, its validity must be widely acknowledged. This precondition certainly seems to be met in the case of the self- interest assumption, for the belief in the power of self- interest, far from being confined to theoreticians closeted in ivory towers, is held, in some form or another, by people in all segments of society: politicians, policy analysts, educators, captains of industry, athletic coaches, and, most importantly, the layperson. That the average person is a self-interest theorist, and one in the neoclassical mode at that, is well-documented in general surveys ( Kohn, 1990 ; Wuthnow, 1991 ) as well as in more systematic studies ( Miller & Ratner, 1996 , 1998 ). For example, the layperson believes that people's attitudes and behaviors are highly influenced by monetary incentives as well as by other personal stakes ( Miller & Ratner, 1996 , 1998 ). Thus, people who would benefit materially from the implementation of a social policy are expected to have more favorable attitudes toward that policy than are people who would not. Similarly, those offered financial compensation for undertaking a socially beneficial act (e. g., giving blood) are expected to be more willing to undertake it than are those who are not. The most impressive evidence that individuals are self-interest theorists is the finding that the predictive power they accord self-interest is largely unaffected by the explanatory power it has for their own behavior. Even when people's own attitudes toward a social policy are incongruent with their acknowledged level of vested interest in it (either too positive or too negative), they still think that the attitudes of others will be congruent with their self- interest ( Miller & Ratner, 1998 ) . The finding that laypersons believe fervently in the power of self-interest to shape attitudes and behaviors will strike few as surprising, certainly not those social scientists who adhere to the central assumptions of rational choice theory (see Green & Shapiro, 1994 ) . However, it should give pause to anyone familiar with the large body of research showing consistently weak links between self-interest and social attitudes (see Sears & Funk, 1990 , 1991 ). The small actual effects of self-interest stand in sharp relief to the substantial assumed effects of self-interest and as such pose a puzzle: How is it that people come to embrace the theory of self-interest when **everyday life provides so little evidence of it?** To answer this question, a closer examination of self-interest rhetoric is in order. The Origins of the Layperson's Belief in the Self- Interest Motive Knowledge pertaining to the relative power of different motives is acquired through the same means as are other forms of social knowledge: instruction and experience. Consider first how instruction might foster the belief in the potency of self- interest. Every culture, through its scientific theories, provides its members with answers to fundamental questions about the nature of the human condition ( Fiske, Kitayama, Markus, & Nisbett, 1998 ; Shweder, 1991 ). As Schwartz (1997 ) described this process, "science creates concepts, ways of understanding the world and our place in it, that have enormous effect on how we think and act" (p. 21). Few questions about the nature of the human condition are as important as that of the source of human action, and most cultures forge collective representations that specify both what does motivate people ( descriptive theories) and what should motivate people ( prescriptive theories). The dissemination of these collective representations takes both direct and indirect routes. Direct instruction. Repeated instruction (whether at the parent's knee or in the classroom) on the power of self-interest is unlikely to be lost on its audience. Unfortunately, there has been very little research on the extent or effect of socialization of this type. What evidence there is, however, suggests that self-interest motivation can indeed be taught. For example, Frank et al. (1993 ) demonstrated that efforts to evangelize on behalf of the self- interest motive can produce converts even among those who could be assumed to already have strong theories of human motivation (i. e., college students). The specific question addressed by Frank et al. (1993 ) was this: Does exposing college students to the precepts and findings of rational choice theory influence the power they perceive self-interest has, or at least should have, over their own lives and that of the average other? The researchers examined this question by assessing students' responses at both the beginning and end of the semester to two ethical dilemmas ("Would you return a lost envelope with $100 in it?" and "Would you report a billing error that benefited you?"). Students were members of one of two different microeconomics classes or of a class unrelated to economics (astronomy). Of the economics classes, one was taught by an instructor who specialized in game theory (a field in which self-interest is axiomatic), the other, by an instructor who specialized in economic development in Maoist China. The results supported the hypothesis that studying economics can foster self-interest. Over the course of the semester, the responses of students in the game theorist's class increased in self-interest more than did those of students in the other economist's class; these latter students' responses, in turn, increased more in self-interest than did those of students in the control (astronomy) professor's class. Similar changes emerged on measures assessing students' expectations of the actions of the average person. The most significant finding of Frank et al. (1993 ) for the present analysis is that the experience of taking a course in microeconomics actually altered students' conceptions of the appropriateness of acting in a self-interested manner, not merely their definition of self-interest. Instruction in economics, it would appear, does not make cynics out of students by persuading them that the motivation behind people's actions, whatever it appears to be, is inevitably self-interest. Frank et al.'s (1993 ) participants did not emerge from Economics 101 believing that it actually is in one's self-interest to report a favorable billing error because, for example, it preempts guilt or fosters a reputation for honesty. Rather, they emerged apparently believing that not reporting a favorable billing error, in addition to being self-interested, is also the rational and appropriate action to take, however guilty one feels doing so. Indirect instruction. One need not take a course in microeconomics to learn about the descriptive and prescriptive power of self-interest, of course. The putative power of self-interest can be communicated much more subtly. For example, myriad daily appeals to the self-interest motive reinforce the belief that this motive should be powerful. More than merely pointing out the instrumental value of a particular activity (e.g. , "ethics pay"), these appeals communicate an important message about the virtue of instrumental value: It is good to pursue what pays. Thus, appeals to self- interest, although explicitly pointing to the means by which the goal of self-interest can be achieved, implicitly point to the wisdom and appropriateness of pursuing self-interest ( Frey, 1997 ; Kohn, 1990 ). This last point suggests a novel perspective on the finding that people's sensitivity to their self-interest increases after it has been made salient to them ( Green & Cowden, 1992 ; Sears & Lau, 1983 ). Pointing out to people where their material self-interest lies may make it salient to them, but it also makes the norm of self- interest salient to them, and it may be due as much to the latter as to the former that this intervention leads them to act more in accordance with self-interest. As an illustration, consider then-candidate Ronald Reagan's famous refrain in the 1980 U.S. presidential campaign: "Ask yourself, are you better off today than you were four years ago?" This refrain reflected a message as well as a question. In addition to asking voters to reflect on the link between their self-interest and their electoral choice, it instructed them on what should motivate their choice of political candidate–self- interest. In effect, by telling voters that a vote for him would be good for their pocketbook, Reagan was telling them that it was normative to vote on the basis of their pocketbook. It is interesting to note that the correlation between self-interest (defined by personal financial situation) and candidate preference in the 1980 U.S. presidential election was only .08 ( Sears & Funk, 1991 ). However, it rose to a more substantial .36 by 1984 ( Lau, Sears, & Jessor, 1990 ), tempting one to speculate that the tenor of the self-interest-celebrating rhetoric of Reagan's first term contributed to the social construction of voting behavior (and perhaps to many other aspects of social life) as being something that self-interest does and should dominate. People, it would appear, can be taught to act "naturally." Why Does Personal Experience Not Invalidate Self-Interest Indoctrination? Self-interest rhetoric may be ubiquitous in contemporary society, but social science research suggests that this will not be the message of everyday social experience. For example, research suggests that everyday experience should reveal self- interest to be related only weakly, if at all, to people' s attitudes toward social policies ( Sears & Funk, 1990 , 1991 ) and to their satisfaction with the outcomes provided by social institutions ( Tyler, 1990 ). However, this assumes that social actors are as comfortable revealing divergences between their self-interest and their attitudes as they are revealing convergences, which, as we shall see, is not the case. The fact is that people's everyday actions and words exaggerate the actual power self- interest has over them, thereby providing the layperson with spurious "evidence" of the power of self- interest. In this way, everyday experience not only fails to contradict societal rhetoric but actually reinforces it. The Norm of Self-Interest and Behavior Social norms can be defined as shared perceptions of appropriate behavior that possess the power to induce people to act publicly in ways that deviate from their private inclinations ( Miller & Prentice, 1996 ) . The norm of self-interest is a case in point: It **induces people to act publicly in ways that maximize their material interests**, **whether or not they are so inclined privately**. This norm, like most other norms, reflects both a descriptive belief (people are self-interested) and a prescriptive belief (people ought to be self-interested). **Either of these beliefs is sufficient to induce the layperson to act more in line with self-interest than he or she might personally be inclined to do.** The power of the norm of self-interest does not end with its influence on people's behavior. By influencing behavior and rhetoric it also influences lay theories as well as more formal theories of human behavior. The reason for this is simply that the layperson (along with the professional observer) frequently interprets evidence of the strength of the norm of self-interest to be evidence of the strength of the motive of self-interest. The consequence of this misinterpretation is a strengthening of the norm itself. How precisely the norm of self-interest contributes to a false impression of the inevitability and naturalness of the self-interest motive is the topic to which we now turn. Descriptive Power The expectation that others will act in a particular way frequently leads people to act similarly (see Darley & Fazio, 1980 ; Miller & Turnbull, 1986 ; Snyder & Stukas, 1998 ) . There are at least two reasons why the expectation of self-interested behavior from others could lead people to act (contrary to their private preferences) in a self- interested manner themselves and thereby unwittingly provide further evidence of the power of self-interest. One reason people pursue self-interest when they anticipate self-interested behavior from others is the fear that to do otherwise would lead to their exploitation. This dynamic is nicely captured in experimental games where the more individuals expect those with whom they are interacting to pursue a self-interested strategy, the more inclined they are to pursue a self-interested strategy themselves ( Bouas & Komorita, 1996 ; Kelley & Stahelski, 1970 ; Messé & Sivacek, 1979 ). The experience of experimental games or any similarly structured social relationship thus reinforces people's belief that individuals are "out for themselves" and leaves them even more convinced that pursuing a competitive orientation is the rational and appropriate thing to do ( Kelley & Stahelski, 1970 ; Miller & Holmes, 1975 ) . A second reason people pursue self-interest when they anticipate self-interested behavior from others is the fear that to do otherwise would be a waste of time and effort. This assumption is common in situations where actors must decide whether to undertake action on behalf of a cause they support but in which they, like most potential actors, have no personal stake. Individuals in this circumstance tend to assume that their supportive position is not shared by others and hence that any action of theirs is likely to be solitary and ineffectual ( Miller & Ratner, 1998 ) . As Coontz (1992 ) described this dynamic, "The major barrier to social involvement is not people's commitment to a purely individualistic way of life but their feeling of helplessness, the fear that they are the only people who feel this way" (p. 284). Prescriptive Power The descriptive component of the norm of self-interest primarily affects behavior by influencing an actor's calculation of the material consequences of deviating from a self-interested course of action. In contrast, the prescriptive component of the norm of self-interest primarily affects behavior by influencing the actor's calculation of the social consequences of deviating from a self-interested course of action. The fear that deviating from one's material self-interest will provoke dismay, suspicion, or derogation can be as powerful a deterrent as the fear that it will prove futile or render one vulnerable to economic exploitation. For an example of how the prescriptive component of a self- interest norm can affect behavior, consider once again the experimental game context. A reliable finding in this context is that decisions made by groups tend to be more self-interested than those reached by individuals ( Schopler & Insko, 1992 ) . The traditional explanation for this finding points to the feelings of anonymity and diffusion of responsibility experienced by individuals in groups. Ratner (1999 ), however, has speculated that enhanced self-interest in groups may actually reflect the prescriptive component of the self- interest norm, reasoning that even if people themselves are comfortable pursuing a non-self-interested course in such a situation, their belief that their teammates are not leads them to embrace a competitive team strategy. Supporting her speculation, Ratner (1999 ) found participants in an experimental game believed that their teammates were less comfortable with a cooperative strategy than they themselves were, and this belief led the individuals to play more competitively when their outcomes and those of their teammates were interdependent. People's wish to avoid the disapprobation of peers and observers is not the only reason they conform to the norm of self- interest. Even when deviation would neither affect nor be known by anyone else, actors might still override their private preferences and conform to the dictates of narrowly defined self-interest simply because they have internalized the belief that it is appropriate and rational to do so ( Tyler, Huo, & Lind, 1999 ). Thus, although research consistently demonstrates that even positive material outcomes (e.g. , high grades) can leave people upset if they are produced by procedures that people consider unfair, most people are reluctant to express their dissatisfaction in these circumstances, because they have internalized the belief that one ought not to be upset when the bottom line favors the self ( Greenberg, 1987 ). Of course, by concealing from public view their dissatisfaction with unfairly produced positive outcomes, people only provide further evidence of the (social) fact that people are self- interested. The Norm of Self-Interest and Accounts for Behavior The norm of self-interest influences more than the actions people take and the attitudes they express. It also influences the accounts people provide for their actions and opinions ( Mills, 1963 ). The easiest way to avoid the negative reactions behavioral and attitudinal deviations from material self-interest provoke is simply to refrain from taking such actions or expressing such attitudes. However, it is also possible to normalize seemingly deviant actions or opinions by providing accounts for them that emphasize their basis in self-interest ( Miller & Ratner, 1996 ) . The prominence of self-interest-speak in behavioral accounts can be seen in the different reasons people provide for their voting preferences at different temporal points. Sears and his colleagues have found that the relation between self-interest and voting behavior is much higher in exit polls than in either preelection or postelection surveys ( Sears & Lau, 1983 ). Whether people distort their votes in the direction of their self- interest or vice versa, it appears clear that people who have recently cast a vote are motivated to tell a story that closely links their vote with their self-interest (see also Stein, 1990 ). People also provide self-interested accounts when asked to explain their attitudes. An interesting example of this is provided by a study in which researchers examined the impact vested interest had on Whites' attitudes toward school busing as a means of achieving racial integration ( Green & Cowden, 1992 ; Sears & Funk, 1990 ). When asked to explain their position, most opponents of busing focused on narrowly defined self-interest arguments: concern for their children's safety, diminished property values, inconvenience to kids, and so forth ( Sears & Funk, 1990 ). However, the data also indicated that whether or not respondents were parents or property owners had virtually no affect on the likelihood of their expressing an antibusing position. It would appear, then, that although people were comfortable acknowledging to an interviewer that they were opposed to busing, they were not comfortable offering an account for that opposition that did not focus on material self-interest. People's motivation to make their actions appear self- interested is not limited to the political sphere. They seem motivated to cast even highly pro-social acts in terms of self-interest ( Lerner, 1982 ). For example, on the basis of extensive interviews and surveys, Wuthnow (1991 ) claimed that although people actually engage in many acts of genuine compassion, they are loathe to acknowledge that these acts may have been motivated by genuine compassion or kindness. Instead, people offer pragmatic or instrumental reasons for them, saying things such as "It gave me something to do" or "It got me out of the house." Indeed, the people Wuthnow interviewed seemed to go out of their way to stress that they were not a "bleeding heart," "goody two-shoes," or " do-gooder." On the basis of Wuthnow's findings, it would appear that accounts for actions that emphasize self-interest, rather than diminishing the actions, actually may normalize or license them. Self-interest provides a sufficient account for what might otherwise seem inexplicable. Interestingly, the claim that people–at least Americans–often conceal their more noble sentiments under the guise of self-interest is not new. Over 150 years ago, the French social philosopher Alexis de Tocqueville observed that "Americans ... enjoy explaining almost every act of their lives on the principle of self-interest ... I think that in this they often do themselves less than justice, for sometimes in the United States, as elsewhere, one sees people carried away by the ... spontaneous impulses natural to man. But the Americans are hardly prepared to admit that they do give way to emotions of this sort" ( 1835/1969 , p. 546) .

#### Studies prove you’re biased to believe in self-interest, even when it isn’t there.

Critcher and Dunning 11 <Clayton, Haas School of Business, University of California, Berkeley, and David, Department of Psychology, Cornell University, “No good deed goes unquestioned: Cynical reconstruals maintain belief in the power of self-interest,” Journal of Experimental Social Psychology, Volume 47, Issue 6, November 2011, Pages 1207-1213, https://doi.org/10.1016/j.jesp.2011.05.001>#SPS

In this manuscript, we examined how people may retain their beliefs about the power of self-interest even in the face of what should be disconfirming evidence. We suggested that after being engaged in thought about why people act in an apparently selfish manner, people reconstrue that behavior more as a product of selfishness than they had before. No such shift would occur when people considered selfish behaviors. Our studies provide evidence for this attributional cynicism. Upon learning that the prevalence of seemingly selfless behaviors is especially high, people do not revise their beliefs about the overall prevalence of selfishness in the world. **Instead, they decide that seemingly selfless behaviors must be selfish after all** (Study 1). Studies 2 and 3 moved beyond this initial finding to demonstrate that such revision is not simply a necessary consequence of receiving new empirical information. Instead, mere attributional thought prompted people to see more self-interest in a seemingly selfless act. Their attributions, in contrast, for a seemingly self-interested act do not systematically deviate from their prior beliefs in a more selfless direction (Studies 2a and 2b). Finally, Study 3 found that such cynicism is particularly robust. As participants thought more about leading philanthropists—people selected by media outlets for their unusually kind, generous acts—their thoughts became increasingly cynical. Notably, participants more frequently appealed to selfish than selfless motives when explaining the philanthropists' actions. We have discussed bias by highlighting the way in which one's ultimate attributions did not follow from one's prior beliefs. But one could describe this effect in the opposite way, placing “bias” in one's prior beliefs—an unjustifiably definitive view of what self-interest predicts that is “corrected” with further thought. Regardless of where bias is placed (i.e., in the prior or posterior beliefs), what is clear is that the difference between prior and posterior beliefs indicates a counternormative dynamic process. For those cases in which participants received no new information, no Bayesian would endorse the systematic revisions in theory that participants made. **This dynamic process leads people to maintain a belief in the predictive value of the self-interest construct, but then to fudge their interpretation of data in evaluating the theory's accuracy.** Before considering a behavior, our participants believed that self-interest was a useful concept predicting which behaviors were more likely than others (e.g., dishonesty over honesty). But, after consideration, they **shifted toward believing that self-interest could predict anything** (both honesty and dishonesty). This left participants with “confirmed” views of the power of self-interest, **regardless of the data** (real or hypothetical). However, with faith in the power of self-interest affirmed, people can approach a new situation all-too-eager to once again rely on their confidence that seemingly selfish behavior (rather than seemingly selfless ones) will emerge (e.g., Miller & Ratner, 1998).4

### Tautology

#### Claiming everything that appears altruistic is actually selfish is tautological and wrong

Bowles and Gintis 02 <Samuel, Professor of Economics at the University of Siena, Emeritus Professors of Economics, University of Massachusetts and Director of the Economics Program at the Santa Fe Institute and Herbert, is a member of the External Faculty of the Santa Fe Institute, Emeritus Professors of Economics, University of Massachusetts, “The Origins of Human Cooperation,” July 24th, 2002, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.590.6817&rep=rep1&type=pdf>#SPS

Two themes run through our account of the origin of cooperation among humans: one is the importance of groups in human evolution and the power of multi-level selection, and the other is the underlying dynamic of gene-culture co-evolution. We close with comments on what we consider two mistaken approaches. These are, first, the tautological extension of self-interest to the status of the fundamental law of evolution, and second, the representation of culture as an epiphenomenal expression of the interaction of genes and environments. Like de Tocqueville’s “Americans,” a distinguished tradition in biology and the social sciences has sought to explain cooperative behavior “by the principle of self interest, rightly understood.” From J. B. S. Haldane’s quip that he would risk his life to save eight drowning cousins to the Folk Theorem of modern game theory, this tradition has clarified the ways that relatedness, repeated play, and other aspects of social interactions among members of a group might confer fitness advantages on those engaging in **seemingly unselfish behaviors.** The point is sometimes extended considerably by noting that if the differential replication of traits by selection operating on either culturally or genetically transmitted traits is monotonic in payoffs, only traits that on average have higher payoffs will be evolutionarily successful. If selfish behaviors are then defined as those that on average have higher payoffs, the principal of self-interest becomes the fundamental law of evolution. Some prominent researchers in evolutionary biology have taken precisely this tack. Richard Dawkins (1989[1976]), for instance, states, in the course of the first four pages of The Selfish Gene, “a predominant quality to be expected in a successful gene is ruthless selfishness. This gene selfishness will usually give rise to selfishness in individual behavior.…Let us try to teach generosity and altruism, because we are born selfish.”3 Similarly, drawing out the philosophical implications of the evolutionary analysis of human behavior, Richard Alexander (1987) says, “ethics, morality, human conduct, and the human psyche are to be understood only if societies are seen as collections of individuals seeking their own self-interest…That people are in general following what they perceive to be their own interests is, I believe, the most general principle of human behavior.” (pp. 3,35) Like de Tocqueville, **we object to this tautological extension of the principle of self interest**. Our concern is not with the fitness-based or other payoff monotonic dynamic process assumed in this approach. It goes without saying that traits that experience lower fitness in a population will be handicapped in any plausible evolutionary dynamic: even cultural evolution may be strongly biased towards the proliferation of behaviors leading to individual material success. Rather, our concern is with the **distortion of the term “self-interest.”** Those who, in Darwin’s words, were “ready to warn each other of danger, to aid and defend each other” would tautologically be deemed “selfish” if, as Darwin (1973) suggested, tribes in which these behaviors were common would “spread and be victorious over other tribes.” We have eschewed the terms “selfish” or “self-interested” to avoid these confusions, and have instead defined cooperative behaviors in terms of their costs to the individual and their beneficial consequences for group members. Our models and simulations show that these behaviors may proliferate under plausible conditions due to the group structure of human populations and the success of groups in which cooperators are common.

## Psychology

### Strong Reciprocity - General

#### **Strong reciprocity is both evolutionarily useful and empirically verifiable**

Fehr and Rockenbach 04 <Ernst, Institute for Empirical Research in Economics, University of Zurich, and Bettina, Chair in Microeconomics, University of Erfurt, Nordhaeuser, “Human altruism: economic, neural, and evolutionary perspectives,” Current Opinion in Neurobiology, Volume 14, Issue 6, December 2004, Pages 784-790, https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?\_rdoc=1&\_fmt=high&\_origin=gateway&\_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y#!>#SPS

Why is cooperation observed at all and what are the mechanisms that enable and sustain human cooperation in social dilemma situations, even in an environment with (a considerable number of) selfish subjects? Recent research indicates that **strong reciprocity** is crucial for the establishment of cooperation in groups with a share of selfish individuals. A person who is willing to reward fair behavior and to punish unfair behavior, **even though this is often quite costly and provides no material benefit for the person, is called a ‘strong reciprocator’** [13.](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib13), [14.](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib14), [15.•](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib15). Because strong reciprocity is costly for the individual reciprocator, the question arises as to how such behavior could evolve evolutionarily. It has been shown, however, that a positive share of strong reciprocators in the population can be part of an evolutionarily stable situation [16.••](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib16), [17.](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib17), [18.](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib18). Strong reciprocity has been observed in sequential social dilemma experiments, even in interactions with completely anonymous strangers [14.](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y#bib14), [19.](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib19), [20.](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib20), across many different cultures [[21]](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib21), and under stake sizes of up to three months income [[22]](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib22). Strong reciprocity contributes to moderate levels of cooperation in sequential dilemma settings. If, however, effective punishment opportunities are available, high levels of cooperation are achieved because the cooperative group members can discipline selfish subjects [23.](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib23), [24.](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib24). In these experiments, subjects are given the possibility of reducing the other subjects’ income at their own cost after having seen the others’ contribution to the public good. **These punishment possibilities are heavily used**, and the lower an individual's contribution relative to the group average, the more the individual is punished. As a result, a large increase in cooperation is observed (see [Figure 1](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "fig1)). Punishment in this experiment could, in principle, be attributed to selfish incentives because of repeated interactions between the subjects. The absence of any material gain from punishment is ensured in the study by Fehr and Gächter [[25]](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib25), because the punished and the punishing subjects never interact again. Nevertheless, punishment is frequently observed, and punished subjects typically increase their cooperation in future interactions with other subjects, so the future interaction partners of the punished subjects benefit from the punishment. Recent evolutionary models show that altruistic punishment even survives evolutionary pressures in relatively large groups [16.••](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y#bib16), [18.](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y#bib18).

What are the proximate mechanisms behind strong reciprocity? Recent neuroeconomic studies that scan subjects’ brains while they are making decisions in interactive economic experiments provide interesting results on the neural foundations of strong reciprocity [33.](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib33), [34.](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib34), [35.](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib35), [36.••](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib36), [37.••](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib37). They support the hypothesis that neural representations of emotional states guide human decision-making and they suggest that **subjects derive specific rewards from mutual cooperation and the punishment of norm violators.**

#### Neurobiology supports strong reciprocity

Fehr and Rockenbach 04 <Ernst, Institute for Empirical Research in Economics, University of Zurich, and Bettina, Chair in Microeconomics, University of Erfurt, Nordhaeuser, “Human altruism: economic, neural, and evolutionary perspectives,” Current Opinion in Neurobiology, Volume 14, Issue 6, December 2004, Pages 784-790, https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?\_rdoc=1&\_fmt=high&\_origin=gateway&\_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y#!>#SPS

A recent study [[36••]](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y#bib36) demonstrated the importance of the interplay of emotions and cognition in economic decision-making. Nineteen participants who responded to fair and unfair offers in a bargaining game were scanned using functional magnetic resonance imaging (fMRI). Less fair offers activated the bilateral insula, which has been implicated in negative emotional states such as disgust, pain, hunger, and thirst. Subjects with stronger insula activation to unfair offers were also more likely to reject these offers. Unfair offers from a human partner also caused stronger insula activation than unfair offers from a computer partner, which suggests the importance of the social context for the insula activation. Unfair offers also activated the dorsolateral prefrontal cortex (DLPFC) and the anterior cingulate cortex (ACC). These activations are interesting because the DLPFC is a region that is often associated with goal maintenance and executive control and the ACC has been implicated in detection of cognitive conflict. In fact, if the insula activation to unfair offers was stronger than the DLPFC activation subjects tended to reject the offer, whereas subjects tended to accept an unfair offer if the DLPFC activation was stronger. **fMRI analysis of subjects playing a PD indicates that mutual cooperation with a human partner yields stronger activation of the brain's reward circuit** (components of the mesolimbic dopamine system including the striatum and the orbitofrontal cortex) than mutual cooperation with a computer partner that yields the same monetary payoff does [[34]](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y#bib34). Moreover, there is also evidence implying a negative response of the dopamine system if a subject cooperates but the opponent defects. **These findings indicate that there is a neural basis for strong reciprocity.** This interpretation receives further support from an imaging study that scanned subjects while they were making gender judgments of faces that were previously attached to opponent players in a sequentially played PD [[38••]](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib38). Some faces were associated with cooperative decisions, some with defections, and some were neutral. The study shows that the presentation of faces of intentional cooperators caused increased activity in left amygdala, bilateral insula, fusiform gyrus, superior temporal sulcus, and **reward-related areas.** Moreover, a particularly noteworthy result is that **merely seeing cooperators’ faces** during the gender judgment task activated reward-related areas. One of the major puzzles posed by the existence of strong reciprocity is the fact that many cooperative subjects punish defectors in one-shot PD games although punishing is also costly for punisher. A new study that combines a sequential PD experiment with positron emission tomography (PET) provides a solution to this puzzle. A punishment opportunity augmented the PD in this study because the cooperating player could punish a defecting player. In the effective punishment condition the cooperator could reduce the defector's economic payoff by punishing him, whereas the cooperator could only punish the defector symbolically in a control condition, that is, the assignment of punishment points to the defector did not reduce the defector's payoff in this condition. The contrast between the effective and the symbolic punishment condition activated the dorsal striatum, which is well known for its reward processing properties. The study also shows that those subjects with a higher activation in the dorsal striatum impose a greater punishment on defectors. Moreover, additional analyses suggest that the activation in the dorsal striatum reflects the anticipated satisfaction associated with the punishment. The previous results indicate a neural basis for certain forms of strong reciprocity. However, we do not know at present the neural basis of third-party punishment [[32•]](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib32), which plausibly requires empathizing with the victims of norm violations. A study in which the brain activity of humans experiencing pain was compared to the brain activity of humans observing a loved one experiencing a similar pain stimulus [[39]](https://www-sciencedirect-com.proxy-um.researchport.umd.edu/science/article/pii/S0959438804001606?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y" \l "bib39)reveals that empathy with the pain of others does not activate the whole pain matrix, but is based on the activation of areas that represent solely the affective dimension of pain. This observation yields the neural basis of empathy (between loved ones). Hence, an interesting question is whether the same brain areas are activated in third party punishment, that is, when people empathize with strangers who became the victim of a norm violation. **Economic experiments show that strong reciprocity is a key force in human cooperation,** and evolutionary models indicate that it can be a stable and adaptive trait. In addition, neuroeconomic studies examined the neural basis of strong reciprocity. The anterior insula seems to play a crucial part in the willingness to reject unfair outcomes, and reward-related circuits involving the ventral and dorsal striatum seem to be important for human cooperation and the punishment of norm violations. These exciting results suggest that the combination of interactive economic experiments with brain imaging techniques constitutes a fertile area for future research that promises a better understanding of complex social behaviors that form the basis of human societies.

#### Altruism is an important driver of human decision making over and above self-interest

Fehr and Fischbacher 03 <Ernst and Urs, University of Zürich, Institute for Empirical Research in Economics; “The nature of human altruism,” Nature volume 425, pages 785–791 (23 October 2003), http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043>#SPS

Repeated interactions and reputation formation A reputation for behaving altruistically is another powerful mechanism for the enforcement of cooperation in public goods situations. If people are engaged in bilateral encounters as well as in n-person public goods interactions, a defection in the public goods situation, if known by others, may decrease others' willingness to help in bilateral interactions43. Suppose that after each round of interaction in a public goods experiment, subjects play an indirect reciprocity game44. In this game, subjects are matched in pairs and one subject is randomly placed in the role of a donor and the other in the role of a recipient. The donor can help the recipient, and the donor's costs of helping are lower than the benefits for the recipient. The recipient's reputation is established by his decision in the previous public goods round and his history of helping decisions in the indirect reciprocity game. It turns out that the recipients' reputations in the public goods game are an important determinant for the donors' decisions. Donors punish the recipients by being significantly less likely to help when the recipients defected in the previous public goods game. This, in turn, has powerful cooperation-enhancing effects in the future rounds of the public goods game. Helping behaviour in indirect reciprocity experiments has also been documented in the absence of interactions in public goods games45,46. A crucial element in these experiments is that direct reciprocity is ruled out because no recipient will ever be put in a position where he can give to one of his previous donors. **Helping rates between 50 and 90% have been achieved and recipients with a history of generous helping decisions are significantly more likely to receive help themselves.** This suggests that the donors' behaviour may be driven by the desire to acquire a good reputation. However, it is also possible that altruistic rewarding drives helping behaviour. A recent study examines this question by allowing only half of the subjects in the experiment to acquire a reputation47. This means that one can compare the behaviour of donors who cannot gain a reputation with the behaviour of those who can. **The data show that both altruistic rewarding and reputation-seeking are powerful determinants of donors' behaviour.** Donors who cannot acquire a reputation help in 37% of the cases whereas those who can gain a reputation help in 74% of the cases. These results indicate that humans are very attentive to possibilities of individual reputation formation in the domain of rewarding behaviours. **They exhibit a sizeable baseline level of altruistic rewarding,** and when given the opportunity to gain a reputation for being generous, helping rates increase strongly. Humans are similarly attentive to the possibility of repeated interactions with the same individual (reciprocal altruism). The cooperation rate is much higher in social dilemmas if subjects know that there is a possibility of meeting the same partners again in future periods10.

### Strong Reciprocity – Altruistic Punishment

#### Altruistic punishment is empirically verified

Haurwitz et al 16 < Kristin M. Brethel-Haurwitz, Sarah A. Stoycos, Elise M. Cardinale, Bryce Huebner & Abigail A. Marsh, Department of Psychology, Georgetown Univeristy, Scientific Reports volume 6, Article number: 18974 (2016), http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974>#SPS

Costly rejection of unfair offers in the UG is often characterised as altruistic[5](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref5), despite little direct evidence linking costly punishment to other forms of altruism [7](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref7),[8](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref8). Because inter-individual differences in altruism are moderately heritable and stable[7](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref7),[29](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref29),[30](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref30),[31](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref31), and altruistic kidney donors exhibit altruism across multiple settings[24](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref24), we hypothesised that if costly punishment in the UG reflects simple altruistic motivations, this population would be more likely to punish unfair offers. **Our findings did not support this hypothesis,** as kidney donors and controls rejected unfair offers at equal rates. Our alternate hypothesis was that if **costly punishment instead reflects a reaction to inequity driven by norm-enforcing strong reciprocity** [18](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref18),[19](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref19),[20](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref20),[21](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref21),[40](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref40), then costly punishment would more closely correspond to self-reported frequency of normative altruism. Our findings supported this hypothesis, showing a significant relationship between self-reported normative altruism scores and punishment of inequity. These findings are consistent with the idea that punishment of inequity is motivated at least in part by sensitivity to social norms. Both norms of prosociality and norms of punishing defection may result from patterns of behaviour consistent with strategies that are most adaptive or common [20](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref20),[21](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref21), though adopting such patterns of behaviour need not be explicit or conscious. To the extent that norm conformity may drive self-reported, public altruism, this motive may explain both SRA responses and costly rejections of unfair offers in the UG, which are known to follow local cooperative norms [22](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref22),[23](http://www.nature.com.proxy-um.researchport.umd.edu/articles/srep18974#ref23).

#### A wide variety of experiments prove altruistic punishment is common

Fehr and Fischbacher 03 <Ernst and Urs, University of Zürich, Institute for Empirical Research in Economics; “The nature of human altruism,” Nature volume 425, pages 785–791 (23 October 2003), http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043>#SPS

Altruistic punishment The ultimatum game 21 nicely illustrates that a sizeable number of people from a wide variety of cultures 22,23 even when facing high monetary stakes 16,17, are willing to punish others at a cost to themselves to prevent unfair outcomes or to sanction unfair behaviour. In this game, two subjects have to agree on the division of a fixed sum of money. Person A, the proposer, can make exactly one proposal of how to divide the money. Then person B, the responder, can accept or reject the proposed division. In the case of rejection, both receive nothing, whereas in the case of acceptance, the proposal is implemented. A robust result in this experiment is that proposals giving the responder shares below 25% of the available money are rejected with a very high probability. This shows that **responders do not behave to maximize self-interest, because a selfish responder would accept any positive share.** In general, the motive indicated for the rejection of positive, yet ‘low’, offers is that responders view them as unfair. Most proposers seem to understand that low offers will be rejected. Therefore, the equal split is often the modal offer in the ultimatum game. The decisive role of rejections is indicated by the dictator game, in which the proposer unilaterally dictates the division of the money because the responder cannot reject the offer. The average amount given to the responders in the dictator game is much lower than that in the ultimatum game20,24. Rejections in the ultimatum game can be viewed as altruistic acts because most people view the equal split as the fair outcome. Thus, a rejection of a low offer is costly for the responder and it punishes the proposer for the violation of a social norm. As a consequence, the proposer is likely to obey the norm in the future by making less greedy offers. For the purpose of this review, we ran an experiment with ten proposers who met a different responder in ten successive rounds. We observed that proposers who experienced a rejection in the previous round increased their offers in the current round by 7% of the available sum of money. In the ultimatum game, the proposer's action directly affects the responder. However, a key element of the enforcement of many social norms, such as food-sharing norms in hunter-gatherer societies 2,3, is that people punish norm violators not for what they did to the punisher but for what they did to others 25,26. Norm enforcement involves the punishment of norm violations even by those who are not economically affected by the violation. To study this question experimentally, we conducted a third-party punishment game involving three subjects—an allocator, a recipient, and a third party.27 The allocator is endowed with 100 monetary units (MUs), the recipient has no endowment, and the third party is endowed with 50 MUs. The allocator is free to give whatever he wants to the ‘poor’ recipient. After the third party has been informed about the allocator's transfer to the recipient, he can spend money to punish the allocator. Every MU spent on punishment reduces the allocator's income by three MUs. Because it is costly to punish, no selfish third party will ever punish. But if a fairness norm applies to the situation, altruistic punishers are expected to punish unfair transfers. In fact, 55% of the third parties punish the allocator for transfers below 50, **and the lower the transfer, the higher the punishment (**Fig. 1). Moreover, between 70 and 80% of the recipients expect that allocators will be punished for unfairly low transfers. Similar results have been observed when third parties are given the chance to punish subjects in a ‘prisoners' dilemma’27. In this case, they frequently punish a defector if his opponent cooperated. If it is anticipated, the punishment by third parties thus deters non-cooperation. Altruistic rewarding Sequentially played social dilemmas are a powerful tool for the study of altruistic rewarding. They come in various forms—as gift exchange games28, trust games29 or sequentially played prisoners' dilemmas30—but the basic structure is captured by the following example. There is a truster and a trustee, both of whom are endowed with 10 MUs. First, the truster decides how many, if any, MUs to transfer to the trustee. Then the trustee decides how much of his endowment to send to the truster. The experimenter doubles any amount sent to the other subject so that, collectively, the two subjects are best off if both transfer their whole endowment: if both keep what they have, each one earns 10; if both transfer their whole endowment, each earns 20. However, a selfish trustee will transfer nothing regardless of how much he received and, therefore, a selfish truster who anticipates this behaviour will never transfer anything in the first place. **This experiment mimics the essence of a vast number of real-life situations.** A similar structure characterizes any sequential exchange that takes place in the absence of contracts that are enforced by the courts. In these situations, both players are better off exchanging their goods and favours but there is also a strong temptation to cheat. Despite the incentive to cheat, however, more than 50% of the trustees transfer money and their transfers are the higher the more the truster transferred initially 28,29,30. Like altruistic punishment, the presence of altruistic rewarding has also been documented in many different countries31, in populations with varying demographic characteristics32, and under stake levels approaching 2–3 months' income 18. Strong reciprocity and multilateral cooperation A decisive feature of hunter-gatherer societies is that cooperation is not restricted to bilateral interactions. Food-sharing, cooperative hunting, and warfare involve large groups of dozens or hundreds of individuals1. To what extent does strong reciprocity contribute to cooperation in public goods situations involving larger groups of individuals? By definition, a public good can be consumed by every group member regardless of the member's contribution to the good. Therefore, each member has an incentive to free-ride on the contributions of others. Altruistic rewarding in this situation implies that an individual's contributions increase if the expected contributions from the other group members increase. Individuals reward others if the latter are expected to raise their cooperation. In public goods experiments that are played only once, subjects typically contribute between 40 and 60% of their endowment, although selfish individuals are predicted to contribute nothing33. There is also strong evidence that higher expectations about others' contributions induce individual subjects to contribute more33,34,35. Cooperation is, however, rarely stable and deteriorates to rather low levels if the game is played repeatedly (and anonymously) for ten rounds36,37. The most plausible interpretation of the decay of cooperation is based on the fact that a large percentage of the subjects are strong reciprocators but that there are also many total free-riders who never contribute anything35. Owing to the existence of strong reciprocators, the ‘average’ subject increases his contribution levels in response to expected increases in the average contribution of other group members. Yet, owing to the existence of selfish subjects, the intercept and the steepness of this relationship is insufficient to establish an equilibrium with high cooperation (Fig. 2). In round one, subjects typically have optimistic expectations about others' cooperation but, given the aggregate pattern of behaviours, this expectation will necessarily be disappointed, leading to a breakdown of cooperation over time. This breakdown of cooperation provides an important lesson. Despite the fact that there are a large number of strong reciprocators, they cannot prevent the decay of cooperation under these circumstances. In fact, it can be shown theoretically that in a population with a clear majority of strong reciprocators, a small minority of selfish individuals suffices to render zero cooperation the unique equilibrium 38. This implies that **it is not possible to infer the absence of altruistic individuals from a situation in which we observe little cooperation**. If strong reciprocators believe that no one else will cooperate, they will also not cooperate. To maintain cooperation in n-person interactions, the upholding of the belief that all or most members of the group will cooperate is thus decisive. Any mechanism that generates such a belief has to provide cooperation incentives for the selfish individuals. The punishment of non-cooperators in repeated interactions 39,40,41 or altruistic punishment 27,42 provide two such possibilities. If cooperators have the opportunity to target their punishment directly towards those who defect they impose strong sanctions on the defectors. Thus, in the presence of targeted punishment opportunities, strong reciprocators are capable of enforcing widespread cooperation by deterring potential non-cooperators 39,40,42. In fact, it can be shown theoretically that even a minority of strong reciprocators suffices to discipline a majority of selfish individuals when direct punishment is possible38.

### One-Ness

#### Self-other merging, or “One-ness” is a deeply flawed view of altruism

May 11 <Joshua, Assistant Professor of Philosophy at the University of Alabama at Birmingham, “Egoism, Empathy, and Self-Other Merging,” Southern Journal of Philosophy Vol. 49 (2011), pp. 25-39, http://www.joshdmay.com/wp-content/media/may-egoism-merging.pdf>#SPS

5. Proposals and Problems First, it is unclear that removing “oneness” from empathy is conceptually possible. Cialdini et al. want to pry these two apart, but one natural conception of “oneness” seems rather essential to the perspective-taking that is crucial to empathy. If so, then this would undermine their claim that it is oneness and not empathy that increases helping. Nevertheless, what is more instructive is to explore the details of this empathy-merging hypothesis to see how exactly it is meant to explain the empathy-helping relationship (or the oneness-helping relationship).12 We must clarify what such metaphorical talk as “blurring the distinction between self and other” means. This requires specifying the content of the ultimate desire to determine whether it is egoistic or altruistic, and then provide the proposed beliefs this connects with to provide an explanation of the empathy-helping relationship. We can explore the prospects of several versions of the self-other merging proposal, all of which face serious problems. 5.1 Proposal 1: Peculiar Beliefs The simplest version of the empathy-merging hypothesis would just attribute the belief to empathically aroused subjects that they exist in two separate bodies. Their ultimate desire would be to benefit themselves, but they believe they exist in the other body. **So they instrumentally desire to help themselves in their other incarnation,** so to speak. This appropriately explains why they help and in a clearly egoistic way. But this, I submit, is a wildly implausible belief to attribute to normal people, even if the belief is subconscious. **We might even count it as delusional.** Of course, on certain metaphysical views we are all one, in which case the belief that one exists in two separate bodies would not be an error, and so perhaps not a delusion. But even if such a view about the world is correct, we need an argument to show that people would believe it. After all, typical Western subjects (as these are) do not seem to hold such a metaphysical view. Moreover, it is even more implausible to attribute the relevant belief given that it must be empathy-specific. That is, to explain the empathy-helping relationship, one must explain the helping behavior by appealing to some feature of empathy or something that empathy causes, since only the empathically aroused subjects help in such higher numbers. So not only must we attribute what appears to be a delusional belief to subjects on the self-other merging account, we must also hold that empathy in particular induces such a belief. It is important here to keep in mind that we needn’t rule out every egoistic explanation, however implausible it may be. We don’t, for example, seriously consider any hypotheses involving aliens from outer space playing some role. In a similar context to ours, Dale Jamieson makes something like this general, theoretical point: “The tortured, ad hoc explanations offered by [psychological egoism] are more like those given by the Ptolemaic view than those provided by relativity theory or quantum mechanics. All things considered, it is more plausible to suppose that [psychological altruism] is true than [psychological egoism]. The experiments of Batson and others are part of what supports this judgment” (2002, p. 706). The general point applies in this context as well. In any event, Cialdini et al. (1997) do themselves attempt to avoid attributing to subjects such peculiar beliefs: “What is merged is conceptual, not physical. We are not suggesting that individuals with overlapping identities confuse their physical beings or situations with those of the other” (p. 482). 5.2 Proposal 2: Indeterminate Identities Cialdini and his collaborators only maintain that the representations of oneself versus another are to some extent “blurred.” One way to read this is that the identities of the people represented in the content of the relevant mental states are indeterminate.13 One problem with this proposal, though, is that it requires a substantial claim about the psychological ability to represent indeterminate identities of persons. Cialdini et al. do mention previous research they claim suggests the existence of such an ability—the “contemporary theory of the self” as dynamic and malleable. They also mention the idea of “inclusive fitness” in evolutionary theory, which they take to suggest that we can see ourselves in others by evidence of shared genes (pp. 482-3). But this is all quite compatible with explanations employing less drastic conceptual resources. The psychological research merely reveals phenomena such as “more similar self-other word associations” (p. 482); and evolutionary theory only shows that we are attune to the presence of similar genes and so forth. Such data do not preclude the idea that we are still only representing determinate identities of self and other and connecting them in various ways.14 In any event, even granting the ability to represent indeterminate personal identities, there are two additional problems with using this as a non-altruistic explanation of the empathy-helping relationship. First, it won’t adequately explain the behavior of such empathically aroused individuals. The more fundamental purpose of attributing these ultimate motives to people is to explain and predict their behavior, or at least to develop an account that can. Other egoistic models purport to explain the helping behavior by holding that the motivation to help the other individual is simply instrumental to the ultimate motivation to help or benefit oneself. For example, one egoistic explanation (“aversive-arousal reduction”) holds that helping is merely a means to eliminating empathic arousal, which is hypothesized to be a feeling to which subjects are quite averse. This explains the behavior by positing a desire, here an instrumental one, to help the other, where this is represented wholly as an individual distinct from oneself. This is then coupled with the ultimate desire to benefit oneself and a belief (among others) that one can do this by helping the other when no one else apparently can. These materials are sufficient to explain, in an egoistic way, why the empathically aroused individual performs the action of helping a person physically distinct from oneself.15 What, on the empathy-merging model, is going to likewise explain why empathically aroused subjects tended to help someone distinct from themselves?16 Recall that Cialdini et al. maintain that what is merged is “conceptual, not physical.” However, the two cannot be pried apart so easily. A conceptual representation is meant to guide behavior to the physical counterpart. My desire for some coffee, for example, must link at least with a belief that there is coffee in front of me in order to explain my picking it up and drinking it. What is crucial in this toy example is that the explanation of my action of interacting with an object at a distance from myself must be grounded in mental representations of myself as distinct from the object that is at a distance from me. Similarly, a conceptual representation of oneself as not at a distance from oneself will not—barring delusions, etc.—appropriately explain one’s behavior of interacting with objects that are distinct from oneself.17 It seems, then, that we need a representation of another as distinct from oneself to avoid delusional beliefs and to provide an explanation of the helping of a distinct individual. But where exactly does this enter into the mental states of the individual? This representation of distinct identities could occur in the belief alone. But then what is the ultimate desire? If empathically aroused subjects believe that the other person is distinct from themselves, the ultimate desire cannot be to help another— that would be altruistic. The only egoistic ultimate desire will be to connect helping the other person with benefit to oneself. But this is the model for egoistic explanations past, not the newfangled empathy-merging explanation, which appeals to the merging of personal identities to secure a non-altruistic explanation of the empathy-helping relationship. I have argued that this cannot be done in a way that appropriately explains the behavior of helping a distinct other. But let us set aside this problem and raise another. Even if one could explain the behavior by appealing to indeterminate identities, the second problem is that this does not necessarily amount to a nonaltruistic motive. If one acts on a motive that represents another’s benefit at all, even partially, then we have the materials for at least a partially altruistic motive. This relies to some extent on the conceptual framework I have previously developed, which I haven’t argued for here (May forthcoming). But we can provide some general, additional grounds for such a conclusion. In short, admitting partial representations of another in the content of one’s ultimate desire is a significant concession on the part of psychological egoists. After all, their view is not meant to be ultimately confined to empathy-induced helping. They are making a claim about the nature of all ultimate desires of humans—namely, that they are egoistic. If we allow the existence of partially altruistic motives, then the project of a parsimonious account of human motivation is undermined. Once the gates are open to a pluralism of pro-social motives, there is little reason to stop at admitting only partially altruistic ones. The theoretical case for psychological egoism employing principles of parsimony could not be maintained if the competing views both posit two distinct kinds of motives. At best we would be left with no theoretical grounds for preferring the empathy-merging to the empathy-altruism hypothesis. In fact, at such a point we may have grounds for preferring the latter given its likely support from introspection: doesn’t it seem, for example, that your motivation to promote the well-being of your children, say, isn’t instrumental to any other desire to benefit yourself? While introspection’s value may be suspect in gathering the primary data, we might reasonably use it as a sort of tie-breaker. It seems, then, that the self-other merging explanation is subject to a formidable dilemma. In order to properly explain the behavior of helping another, it must posit a representation of another. But if a partial representation of another in the content of the ultimate desire can explain such behavior, then it is arguably altruistic. The alternative is to attribute to empathically aroused subjects the belief that they exist in two separate bodies. **But this is exceedingly implausible**; we have no reason to believe that empathy induces such peculiar beliefs. So far, then, at best the empathy-merging explanation is either wildly implausible or can’t explain the helping behavior and is potentially altruistic. I see only one final way of interpreting the merging explanation. 5.3 Proposal 3: Properties At this point, the challenger might attempt to appeal to merging representations of aspects or properties of self and other. Cialdini and colleagues do at times seem to conceive of their proposal along such lines when they suggest that empathically aroused people “locate more of themselves in the others to whom they are closely attached” (p. 483). Here the explanation of the act of helping another is supposed to be explained, it seems, by appealing to an ultimate desire to benefit oneself combined with a belief that aspects of oneself are, as it were, over there in that other body. This provides the tools for explaining the behavior of helping another distinct from themselves since two individuals are represented—it is only the properties that are identified. But this has two problems, which aren’t mutually exclusive. First, it seems again rather implausible to attribute such beliefs to subjects, especially when the attribution must be empathy-specific. Furthermore, even if empathically aroused subjects do believe that aspects of themselves exist in other bodies, we are still left unable to appeal to non-altruistic desires while still explaining the behavior. Ultimately, these “aspects” must be properties of individuals, and believing that one shares certain properties with another doesn’t entail that the ultimate desire of the individual is not to help another. Even believing that one’s own properties are instantiated in another does not clearly support an egoistic motive. Whether it is egoistic depends on the property. Perhaps, for example, the motive would be egoistic if the self-other property identity is specifically the property of being happy (or some such benefit). For example, one might hold that the ultimate desire of the agent is to benefit herself and that she believes she can do this by promoting the property of being happy (e.g.) in this other person, since she also believes this is the exact same property of herself. After all, empathy is often induced by perceiving the other to be similar to oneself.18 However, this explanation would then be too narrow to account for the empathy-helping relationship in its entirety. Research on empathy indicates that there are various similarities (e.g. same gender, age, background, etc.) that can induce higher levels of empathy for another perceived to be in need. People need not believe they are similar to the needy other in respect of harm or benefit to demonstrate the empathy-helping relationship (see the discussion of the “similarity manipulation” of empathy by Batson, 1991, p. 114). While it is plausible that normal subjects will always believe that the other will benefit from help, there is no reason to think that empathy always induces the belief that they are similar in respect of that benefit or harm, let alone identical. 6. Conclusion Once we take psychological egoism seriously as a live thesis about our ultimate motives, there is every reason to address empirical work that bears on it. While Batson and his collaborators seem to have provided a great deal of support for the empathy-altruism hypothesis, it is quite natural to suspect that the role of perspective-taking and similarity in manipulating empathy is more amenable to an egoistic explanation of the empathy-helping relationship. However, as we’ve seen, the details of such a self-other merging explanation lead to deleterious results. On its most plausible construal, the empathy-merging hypothesis cannot appropriately explain the increased helping behavior of empathically aroused individuals, which is the name of the game in this approach to the egoism-altruism debate. While the considerations raised in this paper certainly don’t rule out all egoistic explanations of the empathy-helping relationship, they fend off one that is especially plausible on its face.19

### Context-Dependence

#### Selfishness is context dependent – meaning it is not PRIMARY

Benkler 11 <Yochai, Berkman Professor of Entrepreneurial Legal Studies at Harvard Law School, faculty co-director of the Berkman Klein Center for Internet & Society at Harvard University, “The Unselfish Gene,” Harvard Business Review, July-August 2011, https://hbr.org/2011/07/the-unselfish-gene>#SPS

What would the world be like if some people consistently operated as self-interested rational actors while others did not? Take the experiments that Lee Ross and his colleagues conducted with American college students and Israeli fighter pilots. As we know, in prisoner’s dilemma games, the two players will both be better off if they cooperate, but neither can trust the other to do so. Game theory predicts that both players will choose not to cooperate instead of taking the risk of losing out by cooperating. Extensive experimental work, however, has shown that people actually cooperate more than the theory predicts. Ross and his collaborators told half the players in their experiments that they were playing the Community Game and the other half that they were playing the Wall Street Game. The two groups were identical in all other respects. Yet, in the Community Game group, 70% started out playing cooperatively and continued to do so throughout the experiment. In the Wall Street Game group, the proportions were reversed: 70% of the players didn’t cooperate with one another. Thirty percent started out playing cooperatively but stopped when the others didn’t respond. This experiment illustrates a couple of points. One, we are not all the same. About 30% of players cooperated even in the Wall Street Game while another 30% acted with self-interested rationality even when told they were in the Community Game. Two, many of us are influenced by context. According to Ross, **the framing of the games influenced 40% of the sample.** The players who thought they were acting in a context that rewarded self-interest behaved in a manner consistent with that expectation; participants who felt they were in a situation that demanded a prosocial attitude conformed to that scenario. When Ross and his colleagues asked the subjects’ teachers or commanders to predict who would and wouldn’t cooperate, it turned out that the **game’s framing forecast behavior better than the teachers and commanders could.** It seemed that participants who were seen as self-interested could be induced to cooperate if the games they were playing were reframed.

## Biology

### Evolution (Cultural)

#### Group selection in evolution leads to a physiological benefit of altruism

Fehr and Fischbacher 03 <Ernst and Urs, University of Zürich, Institute for Empirical Research in Economics; “The nature of human altruism,” Nature volume 425, pages 785–791 (23 October 2003), http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043>#SPS

Yet, what is the evidence for cultural group selection? There is quite strong evidence that group conflict and warfare were widespread in foraging societies[78](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref78),[79](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref79). There are also examples [70](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref70),[80](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref80) suggesting that group conflict contributes to the cultural extinction of groups because the winning groups force their cultural norms and institutions on the losing groups. However, although these examples are suggestive, they are not conclusive, so further evidence is needed. If cultural group selection was a significant force in evolution, then the human propensity to reward and punish altruistically should be systematically affected by inter-group conflicts. In particular, altruistic cooperation should be more prevalent if cooperative acts contribute to success in a group conflict. Likewise, people should be more willing to punish defectors if defection occurs in the context of a group conflict. There is evidence from inter-group conflict games indicating that altruistic cooperation in prisoners' dilemmas indeed increases if the game is embedded in an inter-group conflict[81](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref81). However, there is no evidence so far showing that inter-group conflicts increase altruistic punishment.

#### Purely biological understandings of evolution ignore complex human choice – adding cultural evolution to the mix explains fundamental altruism

Fehr and Fischbacher 03 <Ernst and Urs, University of Zürich, Institute for Empirical Research in Economics; “The nature of human altruism,” Nature volume 425, pages 785–791 (23 October 2003), http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043>#SPS

Gene–culture coevolution The birth of modern sociobiology is associated with scepticism against genetic group selection[67](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref67); although it is possible in theory, and in spite of a few plausible cases[25](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref25), genetic group selection has generally been deemed unlikely to occur empirically. The main argument has been that it can at best be relevant in small isolated groups because migration in combination with within-group selection against altruists is a much stronger force than selection between groups. The migration of defectors to groups with a comparatively large number of altruists plus the within-group fitness advantage of defectors quickly removes the genetic differences between groups so that group selection has little effect on the overall selection of altruistic traits[68](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref68). Consistent with this argument, genetic differences between groups in populations of mobile vertebrates such as humans are roughly what one would expect if groups were randomly mixed[69](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref69). Thus, purely genetic group selection is, like the gene-based approaches of reciprocal altruism and indirect reciprocity, unlikely to provide a satisfactory explanation for strong reciprocity and large-scale cooperation among humans. However, the arguments against genetic group selection are far less persuasive when applied to the selection of culturally transmitted traits. Cultural transmission occurs through imitation and teaching, that is, through social learning. There are apparent large differences in cultural practices of different groups around the world and ethnographic evidence indicates that even neighbouring groups are often characterized by very different cultures and institutions[70](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref70). In addition, a culture-based approach makes use of the human capacity to establish and transmit behavioural norms through social learning—a capacity that is quantitatively, and probably even qualitatively, distinctly human [1](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref1),[71](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref71). Recent theoretical models of cultural group selection[72](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref72),[73](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref73) or of gene–culture coevolution[71](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref71),[74](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref74) could provide a solution to the puzzle of strong reciprocity and large-scale human cooperation. They are based on the idea that norms and institutions—such as food-sharing norms or monogamy—are sustained by punishment and decisively weaken the within-group selection against the altruistic trait. If altruistic punishment is ruled out, cultural group selection is not capable of generating cooperation in large groups ([Fig. 4](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#f4)). Yet, when punishment of non-cooperators and non-punishers is possible, punishment evolves and cooperation in much larger groups can be maintained [73](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref73). This is due to the fact that the altruistic punishment of non-cooperators in combination with the imitation of economically successful behaviours prevents the erosion of group differences with regard to the relative frequency of cooperating members. If there are a sufficient number of altruistic punishers, the cooperators do better than the defectors because the latter are punished. Therefore, cooperative behaviour is more likely to be imitated. Moreover, when cooperation in a group is widespread, altruistic punishers have only a small or no within-group disadvantage relative to pure cooperators who do not punish. At the limit, when everybody cooperates, punishers incur no punishment costs at all and thus have no disadvantage. Thus, small cultural group selection effects suffice to overcome the small cost disadvantage of altruistic punishers that arises from the necessity of punishing mutant defectors. To what extent is there evidence for the role of culture and group selection in human altruism? There is strong evidence from intergenerational ultimatum and trust games that advice from players who previously participated in the experiment increases altruistic punishment and altruistic rewarding [75](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref75). Recent intergenerational public good games where advice is given indicate that later generations achieve significantly higher cooperation levels even in the absence of punishment opportunities[76](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref76). Ultimatum and dictator games with children of different ages show that older children are more generous and more willing to punish altruistically [77](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref77). Although these changes in children's behaviour could be a result of genetic developmental processes, it seems at least as plausible to assume that they are also a product of socialization by parents and peers. Why, after all, do parents invest so much time and energy into the proper socialization of their children if this effort is futile? Perhaps the strongest evidence for the role of cultural norms comes from a series of experiments in 15 small-scale societies[23](http://www.nature.com.proxy-um.researchport.umd.edu/articles/nature02043#ref23), showing decisive differences across societies in the behaviour of proposers and responders in the ultimatum game. Some tribes like the Hazda from Tanzania exhibit a considerable amount of altruistic punishment whereas the Machiguenga from Peru show little concern about fair sharing. Thus, taken together, there is fairly convincing evidence that cultural forces exert a significant impact on human altruism.

### Ontogenetics

#### Experiments with Children prove they are altruistic – it is our most fundamental drive

Warneken and Tomasello 07 <Felix and Michael, Max Planck Institute for Evolutionary Anthropology Leipzig, Germany, “Helping and Cooperation at 14 Months of Age,” INFANCY, 11(3), 271-294>#SPS

Prosocial behaviors such as helping and cooperation are interesting both cognitively and motivationally: To help someone with a problem, the helper must understand the other’s unachieved goal and possess the altruistic motivation to act on behalf of the other. Whereas in the case of helping, understanding another’s individual goal of action might be sufficient, cooperative activities are based on the formation of a shared goal. That is, two or more persons have to perform interdependent roles directed at a shared goal and possess the motivation to mutually support each other’s action to reach that goal. **These kinds of prosocial behaviors are at the core of the human condition.** Indeed, humans might act altruistically and cooperate in ways not found in other primates (e.g., Alexander, 1987; Richerson & Boyd, 2005), giving rise to social-cognitive skills such as complex mind reading and communication With regard to helping, children as young as 12 months show concern for others in distress and sometimes intervene by comforting them (for an overview see Eisenberg & Fabes, 1998). In addition, children occasionally point to objects another person is looking for as a form of helping through informing others (Liszkowski, Carpenter, Striano, & Tomasello, 2006; see also Dunn & Munn, 1986). Most recently, it has been shown that 18-month-old children perform unrewarded acts of instrumental helping spontaneously and flexibly in diverse situations (Warneken & Tomasello, 2006; see also Rheingold, 1982); for example, children helped an experimenter retrieve an out-of-reach object like a marker he had accidentally dropped on the floor, or opened cabinet doors when the experimenter was unable to open them himself. **Importantly, children did not perform these actions in control conditions where no help was needed.** In these acts of instrumental helping, the children understood the other’s unachieved goal and were motivated to help him achieve it.

#### Infants are naturally altruistic – thus, so are humans

Warneken and Tomasello 07 <Felix and Michael, Max Planck Institute for Evolutionary Anthropology Leipzig, Germany, “Helping and Cooperation at 14 Months of Age,” INFANCY, 11(3), 271-294>#SPS

Our claim is thus that **the altruistic tendencies seen in early human ontogeny reflect a natural predisposition.** Socialization can build upon this predisposition, but it is not its primary source. Human cultures cultivate rather than implant altruism in the human psyche. And even if we are wrong about this ontogenetic proposal, and human adults do in fact train altruism in developing young, it is worth asking where this tendency of adults came from? We do not see the adults of other species attempting to implant altruistic tendencies in their offspring. If the data we have presented here are valid, infants are genuinely altruistic early in ontogeny. The starting state of altruism in ontogeny is characterized by children’s tendency to help others spontaneously (i.e. in novel situations, without being encouraged to help, and without the expectation of rewards). It even appears that infants help rather indiscriminately, without taking into account if the beneficiary is a relative or a stranger, whether the other will reciprocate, or how their behaviour will affect their reputation. However, it is implausible from an evolutionary perspective that such a naive altruism in which people help without regard of any of these factors could persist. As Dennis Krebs points out: ‘Evolutionary theory leads to the expectation that dispositions to engage in indiscriminate altruism should not evolve.’ (Krebs, 2006, p. 48). For altruism to be sustained as an evolutionarily stable strategy, it must be complemented by safety measures to avoid being exploited by others and bias altruism towards certain individuals under certain circumstances. Thus, mechanisms that make altruism function selectively must be operative as well. However, this does not necessarily imply that all these mechanisms are co-present with the altruistic tendencies in early ontogeny. For instance, the ability to detect cheaters who profit from altruistic acts but do not repay the costs in the future is potentially of less relevance early in ontogeny when children are mainly surrounded by family members, who – even if not always trustworthy – at least share genes with the altruist so that inclusive fitness benefits are likely. The ability to tell apart other altruists from cheaters probably becomes important only later in life as the interaction with strangers increases. Our proposal is thus that children start out as rather indiscriminate altruists who become more selective as they grow older. Children’s emerging social-cognitive understanding and new experiences will enable them to act altruistically more frequently and across a variety of situations, but this should not just blindly lead to more helping, but more selective helping. This general notion of a differentiation process of prosocial behaviours across development has first been introduced by Dale Hay (1994), who quotes Machiavelli’s motto that ‘A prince must learn how not to be good.’ and hypothesizes that factors such as individual differences, gender roles and other norms should lead to a differentiation later in childhood. Our model corresponds to that by Hay in the general statement that a rather undifferentiated prosocial predisposition is differentiated out during later childhood (see also Caplan, 1993; Peterson, 1982), but in our model we focus on other factors as the cause of this differentiation. Namely, we derive our model from evolutionary theory, leading to the proposal to investigate how the proximate mechanism entailed in different evolutionary models begin to play a role during children’s development. Namely, these are kin selection, direct reciprocity, indirect reciprocity, and the transmission of norms.

#### Babies first instinct is to help others achieve their goals – even at cost to themselves

Warneken and Tomasello 09 <Felix and Michael, Department of Developmental and Comparative Psychology, Max Planck Institute, for Evolutionary Anthropology, Leipzig, Germany, “The roots of human altruism” British Journal of Psychology (2009), 100, 455–471, 2009>#SPS

Instrumental helping has both a cognitive and a motivational component. Cognitively, to help an actor achieve her goal, one must first recognize what that goal is – the change of state in the environment the actor wishes to bring about. Motivationally, to help an actor achieve her goal, one must be motivated by the sight of her achieving her goal, or perhaps the sight of her pleasure upon achievement. 2Note that we draw a distinction between agents performing acts of altruistic helping towards another person’s individual goal and collaborative activities in which two or more agents form shared intentions to jointly act towards a shared goal, usually leading to mutualistic outcomes. See Tomasello et al. (2005) for a conceptual framework of collaboration with shared intentions, as well as Warneken (in press) for the argument that instrumental helping and collaboration might be rooted in different psychological processes. With regard to the cognitive component, it is well known that infants from 12 to 18 months of age (if not earlier) understand other person’s behaviours in terms of the underlying goals and intentions (for an overview see Tomasello, Carpenter, Call, Behne, & Moll, 2005). For instance, infants in this age range can differentiate purposeful from accidental actions (Carpenter, Akhtar, & Tomasello, 1998) and even infer what another person was trying to achieve without actually witnessing the intended outcome (Meltzoff, 1995). With regard to the motivational component, a number of studies demonstrate that infants as young as 12 months of age begin to comfort victims of distress, based upon responses to and alterations of the emotional need of another person (Bischof-Ko¨hler, 1988, 1991; Johnson, 1982; Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992). This kind of prosocial behaviour can also be construed as ‘emotional helping’ as in contrast to instrumental helping, it critically depends on the response to the other’s emotional state (rather than an unachieved goal or conative state), which the helper tries to alter (see Warneken & Tomasello, 2006 for more details). Despite the early emergence of these two components of instrumental helping in separate spheres of activity, it remains an open question whether young infants put these together to perform acts of instrumental helping in which they assist another person in achieving an unfulfilled goal. Do young children combine their cognitive understanding of others’ goals and their altruistic motivation to help others instrumentally? When we first addressed this question, we were surprised that there were virtually no experimental studies on helping in young children. The only experimental study was that by Rheingold (1982) with children at 19 to 32 months of age who participated inhousehold chores simulated in the laboratory. Any act which was appropriate for the task and contributed to its completion qualified as helping and under these criteria many of the young children helped adults by holding a dustpan or putting groceries away. Therefore, this study showed that children can perform sophisticated prosocial behaviours and that the children knew the goals of the rituals. However, it was not assessed whether children would be attentive to the other’s unachieved goal, that is, would be able to infer the intended goal when witnessing a failed attempt. Moreover, no control or baseline condition assessed whether children actually performed the behaviours because the other needed help or whether they were motivated to engage in them for its own sake, independently of the other actually needing help. Therefore, we conducted a series of studies to explore the cognitive and motivational components of helping in infancy. In a first study, we presented 18-month-old infants with ten different situations in which an adult experimenter was having trouble achieving his goal (Warneken & Tomasello, 2006). The variety of situations probed the children’s ability to discern a variety of goals and intervene in a variety of ways. For instance, the experimenter used clothespins to hang towels on a line, when he accidentally dropped a clothespin on the floor and unsuccessfully reached for it. In this case, helping consisted in picking up the clothespin and handing it to the experimenter. In another situation, the ‘cabinet task’, the experimenter was trying to put a stack of magazines into a cabinet, but he could not open the doors because his hands were full. Thus, the child could help by opening the doors for him. For each of the ten tasks there were control conditions to rule out the possibility that children would perform the target behaviour (offering the clothespin; opening the door) irrespective of the other’s need, e.g. because they like to hand things to adults or like to open cabinet doors when their attention is drawn to it. In these control conditions, the same basic physical situation was established, but with no indication that the experimenter needed help. The finding of this study was that children display spontaneous, unrewarded helping behaviours when another person is unable to achieve his goal (but performed these behaviours significantly less often in control conditions where no help is necessary, see Figure 1). Infants did so spontaneously; that is, they intervened without being explicitly asked for help and never being rewarded or praised for their effort. Moreover, helping was observed in diverse situations. Infants handed out-of-reach objects; they completed an action after his failed attempt of stacking books; they opened the door of a cabinet for the other and they brought about the other’s goal by different means such as accessing a box by lifting a flap rather than using the wrong means which the adult was using by unsuccessfully reaching through a tiny hole. This initial experiment showed that 18-monthold infants spontaneously provide instrumental help and do so in a wide range of situations. In a follow-up study we tested even younger infants on several of these tasks and found that 14-month-old infants also helped under some circumstances (Warneken & Tomasello, 2007). Namely, they reliably helped with out-of-reach objects such as the ‘clothespin task’, but did not help in the other types of tasks such as the ‘cabinet task’. One possible interpretation for this finding is that they are willing to help, but can do so only in cognitively less demanding situations with more obvious goals such as a person reaching for an object, but fail to do so in situations with presumably more complex goals and more complex types of intervention. **Thus, even 14-month-old infants help spontaneously in situations in which they are able to determine the other person’s goal.** These acts of instrumental helping are some of the earliest manifestations of altruism in human ontogeny: **children acting on behalf of others without a benefit for themselves.** To further examine the interpretation that these acts of instrumental helping are the result of an altruistic motivation, we tested children’s helpfulness by on the one hand manipulating the costs for helping, and on the other hand varying whether the helper would benefit from this helpful act. **Young children even helped when the costs for helping are slightly raised.** In one study, we once again tested 18-month-old infants in a situation in which an object was on the floor and the experimenter was unsuccessfully reaching for it (Warneken, Hare, Melis, Hanus, & Tomasello, 2007; Experiment 2). But this time the children had to surmount an array of obstacles to pick up the object for the other. This can be quite effortful for toddlers who have just started to walk. But even these obstacles did not hinder them from helping the other person over a test session of ten trials. A similar result was obtained when we made the helping act costly in a different way. In particular, we gave 20-month-old children the opportunity to play with attractive toys in one corner of the testing room (Warneken & Tomasello, 2008). The experimenter was located in the opposite corner, so that when she encountered a problem and needed help, the child had to stop playing and leave the toys behind in order to provide help. **Again, children continued to help in the majority of cases and did so over repeated trials** – even when they had the alternative to play an attractive game. Children are thus willing to put some effort in helping – but do they maybe expect to be rewarded in return? In one experiment we directly contrasted whether 18-month-old children are motivated by the other person’s goal or an immediate benefit for themselves (Warneken et al., 2007; Experiment 1). We experimentally varied whether the helpee would or would not offer a reward in return for their helping effort. The results could not be clearer: Children once again helped by picking up objects the experimenter was unsuccessfully reaching for – **and did so irrespective of being rewarded for it**. Rewarding was neither necessary nor did it increase the rate of helping. **Thus, what determined children’s helping was the other’s unfulfilled goal, not an immediate benefit for themselves.** This suggests the possibility that **young children have an intrinsic motivation to act altruistically.** To further test this hypothesis of an intrinsic motivation for helping, we took advantage of a curious feature of intrinsic motivation: It is a well-established phenomenon that intrinsic rewards can be undermined by salient extrinsic rewards – what has also been called the ‘overjustification effect’ (Deci, 1971; Lepper, 1981). Social-psychological theories suggest that an induced extrinsic motivation to perform an enjoyable activity in order to receive a reward supplants the previously intrinsic motivation, so that when the extrinsic reward is no longer forthcoming, the motivation for the activity decreases. Would we find such an undermining effect of extrinsic rewards also in the case of altruistic helping? When we used the method established in overjustification research in an experiment with 20-month-olds, we found that children who had received a material reward for helping at an earlier time were subsequently less likely to engage in further helping than children who had not received such a reward (Warneken & Tomasello, 2008; see also Fabes, Fultz, Eisenberg, May-Plumlee, & Christopher, 1989, for school-aged children). This rather surprising finding provides even further evidence for the hypothesis that children’s helping is driven by an **intrinsic rather than an extrinsic motivation.** Rewards are often not only superfluous, but can have even detrimental effects as they can undermine children’s intrinsic altruistic motivation. In sum, these series of studies demonstrate that the ontogenetic roots of altruism are apparent early in childhood. Infants from 14 to 18 months of age display spontaneous, unrewarded helping when another person is unable to achieve his goal. They are willing to help multiple times and continue helping even when the costs are raised. Additional experiments confirm that they are actually motivated by the other person’s need rather than an immediate benefit for themselves, as rewarding is neither necessary nor does it increase their rate of helping. On the contrary, children appear to have an intrinsic motivation to help, and extrinsic rewards seem to undermine it. Taken together, **these results indicate that the ontogenetic roots of human altruism are present from very early in human ontogeny.**

## Political Science

### Attitudes

#### Self-interest fails to explain a huge amount of political beliefs

Funk 2000 <Carolyn L, VCU, “The Dual Influence of Self-Interest and Societal Interest in Public Opinion,” Political Research Quarterly, Vol. 53, No. 1 (March 2000): pp. 37-62, http://journals.sagepub.com.proxy-um.researchport.umd.edu/doi/pdf/10.1177/106591290005300102>#SPS

Political theorists have long held that virtuous citizens would willingly set aside self-interest in favor of political judgments based on societal interest (Burtt 1993; Mansbridge 1980, 1990; Scalia 1991; Sinopoli 1992). For many, this model of civic virtue appears impossible due to a belief in the primacy of selfinterest to the exclusion of other motives. These findings argue that normative ideals of civic virtue are not outside the realm of possibility. In contrast to the notion that any self-sacrifice is impossible, some citizens set aside personal considerations and adopt political positions that more closely fit their views of what is good for society. Self-sacrifice appears more likely under certain conditions, however; self-sacrifice of personal costs appears more likely to occur than selfsacrifice of personal benefits. **One's perspective as to what constitutes the public interest remains subjective, but that perspective has a powerful influence on evaluations of public policy** The present study demonstrates that a value commitment to societal interest explains a variety of policy positions; this relationship cannot be explained away as a simple reflection of self-interest motives. Several lines of research bolster the hypothesis that societal interest motivates policy attitudes. The absence or weakness of self-interest to empirically predict political positions across a range of cases where they might be expected to occur argues for the need to consider other bases for political attitudes (Citrin and Green 1990; Sears and Funk 1990, 1991). The long-standing observation that citizens seem to separate private from public concerns suggests that a more 54 public-level perspective might underlie political judgments. Evidence that **national economic conditions** **outperform personal considerations** as explanations for presidential evaluations suggests that societal-level concerns form the primary basis for political judgments. The tendency for those with a stronger sense of civic duty to participate in elections despite personal costs and unclear benefits is consistent with the notion that motives other than self-interest also influence political behavior. Experimental studies of small groups suggest that collective benefits can outweigh personal benefits when group identity is made salient. Individual differences in value orientations similarly predict a willingness to sacrifice direct self-interest in ongoing close relationships. And, the willingness of some people to engage in voluntary and charitable behaviors even at substantial risk to their own well-being or without clear personal benefit further underscores the need to consider motivations for other or collective benefit. While it may be possible to increase reliance on societal interest considerations by manipulating the question or political context, this study suggests that consideration of societal interest occurs for at least some citizens even without special efforts at manipulation. Cognitive processes that contribute to the "morselization" of private and public experience may help explain the tendency to rely on public-level considerations in public policy judgments. Civic norms that encourage citizens to evaluate political issues in terms of the public interest may also play a role. Clearly, not all individuals follow these norms on all issues or to the same degree. The present study suggests, however, that some citizens can and do approximate a model of the self-sacrificing citizen who restrains selfinterested concerns when making political judgments. But, citizens appear more likely to restrain self-interest when it involves endorsing benefits for others even at some personal cost than when it involves giving up benefits for themselves. Both self-interest and societal interest, then, appear to influence public opinion on public policy issues.

#### Self-interest only explains about 1% of the decisions people make on the largest scale - politics

Lau and Heldman 09 <Richard R., Rutgers University, and Caroline Heldman, Occidental College, “Self-Interest, Symbolic Attitudes, and Support for Public Policy: A Multilevel Analysis,” Political Psychology, Vol. 30, No. 4, 2009, http://fas-polisci.rutgers.edu/lau/articles/Lau-Heldman\_SelfInterestSymbolicAttitudesPublicPolicySupport.pdf>#SPS

One of the enduring questions in public opinion research is the extent to which citizens’ political beliefs and behavior are guided by their own narrowly defined tangible self-interest. Virtually every Western moral philosophical and psychological theory gives a prominent role to such self-interest (Sears & Funk, 1991), yet when David Sears started putting this question to empirical test he found surprisingly little support for it. For example, white Los Angeles residents who felt personally threatened by possible neighborhood desegregation, or by economic competition from blacks, or who feared black violence in their neighborhood, were no more likely to support the conservative white former police chief, Sam Yorty, over liberal Black city councilman Tom Bradley in the Los Angeles mayoralty elections of 1969 and 1973 (Kinder & Sears, 1981). Likewise, Los Angeles residents personally affected by the energy crisis of 1974 were no more or less likely to support President Nixon, or “the political system” more generally, than were Angelinos who were not directly affected by the crisis (Sears, Tyler, Citrin, & Kinder, 1978). Similarly, white adults across the United States living in areas where busing for school integration was occurring or threatened, or who had children in public school, were no more opposed to busing than adults without those markers of self-interest (Sears, Hensler, & Speer, 1979). Lastly, adults with close friends or family members serving in the military in Vietnam were no more opposed to the Vietnam War in 1968 than citizens without any such personal involvement with the war (Lau, Brown, & Sears, 1978). In every one of these instances, however, what Sears called symbolic beliefs—early learned affective responses to familiar political symbols such as Democrats and Republicans, conservatives and liberals, blacks and whites—had clear and strong influence over the political attitude or behavior in question. Perhaps the broadest exploration of the self-interest versus symbolic attitudes debate was provided by Sears, Lau, Tyler, and Allen (1980), who took advantage of an unusually large number of viable self-interest items in the 1976 American National Election Study (ANES) to explore[d] the question across four distinct public policy issues: guaranteed jobs and incomes, busing of school children for integration, the rights of people accused of crimes, and government health insurance. The authors found that the personal effects of a recent economic recession were weakly (though statistically significantly) associated with support for guaranteed jobs and incomes, but neither being unemployed or currently laid off from work, nor feeling “worse off” financially than a year ago, had any effect whatsoever on support for this issue in 1976. Likewise, adults with a child in a neighborhood public school, or living in an area where busing was occurring or rumored to occur shortly, were no more opposed to busing for school integration than other respondents. When it came to the rights of those accused of crimes, respondents who had personally seen or been a victim of a crime in the past year were weakly (though again significantly) associated with more conservative views toward law and order, but two other indicators of self-interest—not feeling safe walking alone in your neighborhood at night and staying away from certain parts of town because of fear of crime—had no relationship to law and order beliefs. Sears et al. found the largest effect of self-interest with the issue of national health insurance, where both those without any form of health insurance, and those whose insurance was inadequate to protect against a major illness were more supportive of government health insurance. In all four policy domains studied, some combination of symbolic attitudes— liberal-conservative ideology, party identification, and racial prejudice—proved to be much stronger predictors of attitudes toward the policy question at hand. 514 Lau and Heldman On average across the four issue domains, the various self-interest indicators explained only 1% of the variance over that already explained by symbolic beliefs, while the symbolic beliefs collectively explained 10 times more of the explainable variance in policy attitudes above that already accounted for by self-interest. Sears et al. went on the explore five special circumstances that might be expected to maximize the potential effect of self-interest: private-regarding values, believing the government is generally responsive to the public, feeling politically efficacious, believing the issue is a very important national concern, and political sophistication. None of these five potential moderators had any effect on the relative importance of self-interest and symbolic attitudes in 1976, however. Nor did self-interest in any of these four domains have anything to do with the vote for president in 1976. The conclusion seemed inescapable: narrowly defined **tangible self-interest rarely has much to do with citizens’ political beliefs and behavior.** Subsequent reviews of the literature (Citrin & Green, 1990; Lau, 1990; Sears & Funk, 1991) have revealed a few exceptions to the general rule—e.g., when the stakes are unusually large (Sears & Citrin, 1985)—but not enough to alter the general conclusion offered above.

### Observation

#### Cooperation is more common than selfishness – interdisciplinary evidence proves

Benkler 11 <Yochai, Berkman Professor of Entrepreneurial Legal Studies at Harvard Law School, faculty co-director of the Berkman Klein Center for Internet & Society at Harvard University, “The Unselfish Gene,” Harvard Business Review, July-August 2011, https://hbr.org/2011/07/the-unselfish-gene>#SPS

Like biology, however, the discipline of economics has changed over the years. In 2009, Elinor Ostrom was awarded the Nobel Prize in economics for showing how commons can—**and do**—sustain themselves for centuries as well-functioning systems. The most striking example is in Spain, where thousands of farmers have been managing their access to water through self-regulated irrigation districts **for more than five centuries.** To take another example, 75% of U.S. cities with populations of more than 50,000 have successfully adopted some version of community policing, which reduces crime not by imposing harsher penalties but by humanizing the interactions of the police with local communities. Overcoming our assumptions about self-interest is critical to diagnose the risks that new business rivals pose. In 1999, two experts showed how Microsoft’s entry into the encyclopedia market with Encarta symbolized the transformation made possible by networked information economics. Here was a major player leveraging a powerful position, gained by early-mover advantages and network effects, to bundle a product and distribute it widely at a low cost. Britannica’s lumbering 32-volume, multi-thousand-dollar offering didn’t stand a chance. Ten years later, Britannica had been pushed to a different model—but not by Encarta. Microsoft stopped producing Encarta in 2009 because of competition from a business model that is inconceivable according to the belief in self-interested rationality: Wikipedia. If you feel that Wikipedia—the seventh or eighth most trafficked website, with more than 300 million visitors a month—is unique, ask Zagat’s how the user-generated Yelp has affected its market or Fodor what it thinks about TripAdvisor. The rise of open source software is an example of the same dynamic. For more than 15 years, companies have used open source Apache software for mission-critical web applications, with Microsoft’s server software trailing a distant second. Companies such as Google, Facebook, and Craigslist have also found ways to become profitable by engaging people. **Our old models of human behavior did not—could not—predict that.** The way these organizations work flies in the face of the assumption that human beings are selfish creatures. For decades, economists, politicians, legislators, executives, and engineers have built systems and organizations around incentives, rewards, and punishments to get people to achieve public, corporate, and community goals. If you want employees to work harder, incorporate pay for performance and monitor their results more closely. If you want executives to do what’s right for shareholders, pay them in stock. If you want doctors to look after patients better, threaten them with malpractice suits. Yet, all around us, we see people cooperating and working in collaboration, doing the right thing, behaving fairly, acting generously, caring about their group or team, and trying to behave like decent people who reciprocate kindness with kindness. The adoption of cooperative systems in many fields has been paralleled by a renewed interest in the mechanics of cooperation among researchers in the social and behavioral sciences. Through the work of many scientists, we have begun to see evidence across several disciplines that people are in fact more cooperative and selfless—or behave far less selfishly—than we have assumed. Perhaps humankind is not so inherently selfish after all. Dozens of field studies have identified cooperative systems, many of which are more stable and effective than incentive-based ones. Evolutionary biologists and psychologists have found neural and possibly genetic evidence of a human predisposition to cooperate, which I shall describe below. After years of arguments to the contrary, there is growing evidence that evolution may favor people who cooperate and societies that include such individuals. In fact, a distinct pattern has emerged. In experiments about cooperative behavior, a large minority of people—about 30%—behave as though they are selfish, as we commonly assume. However, 50% systematically and predictably behave cooperatively. Some of them cooperate conditionally; they treat kindness with kindness and meanness with meanness. Others cooperate unconditionally, even when it comes at a personal cost. (The remaining 20% are unpredictable, sometimes choosing to cooperate and other times refusing to do so.) In no society examined under controlled conditions have the majority of people consistently behaved selfishly.