

# *Explaining human altruism*

**Michael Vlerick**

**Synthese**

An International Journal for  
Epistemology, Methodology and  
Philosophy of Science

ISSN 0039-7857

Synthese

DOI 10.1007/s11229-020-02890-y



**Your article is published under the Creative Commons Attribution license which allows users to read, copy, distribute and make derivative works, as long as the author of the original work is cited. You may self-archive this article on your own website, an institutional repository or funder's repository and make it publicly available immediately.**



# Explaining human altruism

Michael Vlerick<sup>1,2,3</sup>

Received: 24 March 2020 / Accepted: 19 September 2020  
© The Author(s) 2020

## Abstract

Humans often behave altruistically towards strangers with no chance of reciprocation. From an evolutionary perspective, this is puzzling. The evolution of altruistic cooperative behavior—in which an organism’s action reduces its fitness and increases the fitness of another organism (e.g. by sharing food)—only makes sense when it is directed at genetically related organisms (kin selection) or when one can expect the favor to be returned (reciprocal altruism). Therefore, evolutionary theorists such as Sober and Wilson have argued that we should revise Neo-Darwinian evolutionary theory. They argue that human altruism evolved through group selection in which groups of altruists were naturally selected because they had a comparative advantage over other groups. Wilson and Sober’s hypothesis attracted followers but is rejected by most of their peers. The heated debate between advocates and critics of group selection often suffers from a lack of conceptual clarity. In response, I set out to clearly distinguish ‘genetic’ from ‘cultural’ group selection (developed by Boyd, Richerson & Henrich) and argue that the latter does not face the potentially debilitating problems plaguing the former. I defend the claim that human altruistic dispositions evolved through cultural group selection and gene-culture coevolution and offer empirical evidence in support. I also argue that actual altruistic behavior often goes beyond the kind of behavior humans have evolved to display. Conscious and voluntary reasoning processes, I show, have an important role in altruistic behavior. This is often overlooked in the scientific literature on human altruism.

**Keywords** Human altruism · Behavioral game-theory · Group selection · Genetic group selection · Cultural group selection · Moral reasoning

---

✉ Michael Vlerick  
michaelvlerick@gmail.com; m.m.p.vlerick@tilburguniversity.edu

<sup>1</sup> Department of Philosophy, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

<sup>2</sup> University of Johannesburg, Johannesburg, South Africa

<sup>3</sup> Afsnee, Belgium

## 1 Introduction

Humans often behave altruistically towards strangers with no chance of reciprocation. Many people donate blood and funds for the benefit of people they will never meet and often do so anonymously. In experimental settings, people often cooperate with strangers in one-shot prisoner's dilemma's (in which 'defecting' always yields a higher individual payoff) and offer something rather than nothing in dictator games to strangers (when they could have kept everything for themselves) (Camerer and Thaler 1995; Camerer 2003; Henrich et al. 2001; Fehr and Rockenbach 2004; Gächter and Herrmann 2009). Many people are also willing to incur costs to punish those who have harmed the group or others. This too is altruistic behavior. (Fehr and Gächter 2002a). While there is variation between cultures, altruistic behavior is a human universal (Gächter and Herrmann 2009; Vakoch 2013).

The question I pose in this paper is the following: why do humans often exhibit altruistic behavior towards non-kin with no chance of reciprocation? From an evolutionary perspective, this is puzzling. The evolution of altruistic cooperative behavior—in which an organism's action reduces its fitness and increases the fitness of another organism (e.g. by sharing food) only makes sense when it is directed at genetically related organisms or when one can expect the favor to be returned. The first kind of altruism is referred to as 'kin altruism' and was elucidated by Fisher (1930), Haldane (1932) and Hamilton (1964) who understood that the altruistic organism was in fact increasing its evolutionary success since it was helping genetically related organisms. The second kind of altruism is known as 'reciprocal altruism' and was elucidated by Trivers (1971) who understood that the altruistic organism was in fact behaving in an 'enlightened' self-interested way since it could expect the favor to be returned in the future (Ruse 1979, p. 49).

Human altruism directed at non-kin with no chance of reciprocation cannot be satisfactorily explained in terms of 'kin selection' or 'reciprocal altruism'. Behavioral and evolutionary scientists and philosophers of science have consequently looked for alternative explanations of human altruism. These explanations often invoke 'group selection'. Influential scholars such as David Wilson (1975, 2005) and Elliot Sober and Wilson (1998) have developed group selection accounts of human altruism and many have followed their lead. Group selection theories, however, remain very controversial and are strongly rejected by an important numbers of scientists and philosophers of science (e.g. Dawkins 1994; Dennett 1994; Maynard Smith 1998; Pinker 2012).

Amid the controversy, a promising account of the evolution of human altruism: cultural group selection and gene-culture coevolution, is often brushed aside for no good reason. Cultural group selection, however, does not face the problems associated with traditional group selection (henceforth: genetic group selection). Human altruistic dispositions, I will argue, evolved through a combination of cultural group selection leading to a highly cooperative niche characterized by prosocial norms and punishments and standard (individualistic) natural selection of (altruistic) psychological traits in this altered social environment (that rewards altruism and punishes free-riding and other anti-social behavior). Such an interaction between cultural and genetic evolutionary processes is referred to as 'gene-culture coevolution' by Boyd and Richerson (1985) and Richerson and Boyd (2005).

While this hypothesis provides us with a plausible and evidence-based explanation of the evolution of altruistic psychological dispositions, it cannot explain many instances of human altruism and moral behavior in general that evidently go beyond the kind of behavior for which these dispositions evolved. The evolutionary story only provides us with half of the story of why humans often behave altruistically towards non-kin with no chance of reciprocation. We must also consider the important role of conscious and voluntary reasoning processes in moral decision-making. This is often overlooked in the scientific literature on human altruism.

In this paper, I have three objectives. Firstly, I want to clearly distinguish between ‘genetic’ and ‘cultural’ group selection and argue that the latter does not face the potentially debilitating problems plaguing the former. This is important since many group selection accounts combine (and do not clearly distinguish between) genetic and cultural group selection. Secondly, I aim to provide a plausible account of the evolution of human altruistic dispositions in particular and human moral psychology in general *and* support my hypothesis with evidence. Finally, I aim to complete extant naturalistic explanations of human altruism that focus on its evolutionary underpinning, by showing and describing the important role of reasoning processes in altruistic behavior.

In Sect. 2, I elucidate the notion of altruism by distinguishing between biological and psychological altruism and discuss the proximate explanations of human biological altruism. In Sect. 3, I take on the ultimate explanation of human (biological) altruism: group selection. I distinguish between ‘genetic’ and ‘cultural’ group selection and argue that the latter—in conjunction with gene-culture coevolution—offers a theoretically satisfactory and empirically supported explanation for the evolution of human altruistic dispositions. In Sect. 4, I discuss the evidence for the existence of a highly cooperative cultural niche in which recent human evolution took place. In Sect. 5, I argue that evolutionary dynamics only provide us with a partial answer to the question why some humans behave altruistically towards strangers with no chance of reciprocation. When explaining altruistic human behavior (and norms) we must also take into account conscious and voluntary reasoning processes. In Sect. 6, I conclude.

## 2 Proximate explanations

### 2.1 Psychological versus biological altruism

What do I mean by altruism? In its vernacular sense, altruism refers to other-regarding and selfless acts and dispositions. Altruists (are predisposed to) engage in costly behavior aimed to benefit others without an ulterior selfish motive (such as enhancing one’s reputation or expecting the beneficiary to return the favor). In the scientific literature on altruism, however, altruism does not take on this vernacular sense but refers to either psychological or biological altruism. Psychological altruism is solely concerned with motives. It refers to the desire to benefit another. Biological (or evolutionary) altruism, on the other hand, is solely concerned with acts. It refers to acts that increase the fitness (the chances of survival and reproduction) of the recipient and decrease the fitness of the actor. The *desire* to share a candy bar is a matter of psychological altruism, while

the *act* of sharing food qualifies as biological altruism (see Sober 1988; Sober and Wilson 1998; Ananth 2005).

Biological altruism is a common occurrence in the natural world. At all levels of complexity, organisms act in ways that reduce their own chances of survival and reproduction and increases the chances of survival and reproduction of other organisms. Most often, the beneficiaries of altruistic acts are offspring or genetically related organisms. This is referred to as ‘kin altruism’ and it makes good evolutionary sense. Enhancing the fitness of genetically related organisms enhances one’s own evolutionary success (i.e. the success an organism has in spreading its genetic material) since it helps organisms carrying similar genetic material to spread their genetic material. From a gene-centric perspective on natural selection (famously popularized by Dawkins 1976) kin altruism is readily understood: genes coding for altruistic behavior towards kin are great replicators (and can therefore be expected to spread), since they ‘help’ copies of themselves in other organisms (i.e. in those genetically related organisms).

Fisher (1930) and Haldane (1932) were the first to formalize this process of ‘kin selection’. It explains most instances of altruistic animal behavior, including its most extreme manifestations such as the sacrificing behavior of eusocial insects like bees and ants for the hives and colonies (of genetically related organisms) they belong to. Later, Hamilton (1964) developed and formalized the concept of ‘inclusive fitness’. According to Hamilton, genes that underlie behavior that benefits a genetically related organism contribute to the inclusive fitness of that organism if the benefit is larger than the cost given the degree of relatedness. So, given that I share 50% of my genes with my brother (on average), my inclusive fitness goes up if my actions boost his fitness by a factor of 10 and reduce my fitness (the cost of my altruistic act) by less than 5.

Biological altruism towards non-kin is less prevalent, but it does occur. Some birds give warning calls when they spot a predator (thereby potentially attracting the attention of the predator), vampire bats share food with conspecifics that didn’t have a successful hunt, and meerkats routinely go on the lookout for danger (and also emit warning cries) while the others are foraging and feeding. In all of these cases, the immediate fitness (chances of survival and reproduction) of the actors decreases and the fitness of the recipient (and often non-related) group members goes up. Despite the fact that it decreases the (inclusive) fitness of the altruistic organism in the short term, such altruistic behavior evolved because it is reciprocal. The altruists are repaid the favor (and free-riders are denied future favors), so the altruist benefits in the long term (its fitness increases). Robert Trivers (1971) elucidated the concept of reciprocal altruism and showed that it is ‘enlightened self-interest’ (Ruse 1979, p. 49).

Humans, however, often engage in altruistic acts directed at strangers (non-kin) and with no chance of reciprocation. Evidence for this unique form of altruism (it has not been observed in any other species) can be found both in the field and in the lab. Many people donate blood and money, they offer their seat to unrelated pregnant ladies and help old people cross the street. All of these altruistic acts come with no expectation of reciprocation. In the lab, behavioral game-theoretic experiments—such as one-shot prisoner’s dilemma, public good games or dictator games played for real money—reveal that a sizable percentage of test subjects will indeed act altruistically towards total strangers. They forego a larger payoff to benefit the other player(s), even

when they know that the recipients cannot repay the favor. Many participants will also give up part of their allocated sum in public good games to punish free-riders who do not contribute to the public good. (For an overview of these experiments and the results they yield: see Camerer and Thaler 1995; Camerer 2003; Fehr and Gächter 2002a, b; Fehr and Rockenbach 2004; Gächter and Herrmann 2009).

## 2.2 Proximate explanations

What causes this peculiar behavior? Proximate explanations of human biological altruism—explanations in terms of the direct causes—are not hard to come by. Many people are endowed with psychological altruism or other-regarding preferences: they often desire to help another even if that comes at a personal cost. Furthermore, they possess a sense of fairness and a desire or a feeling of obligation to act fairly. Finally, they want to follow social norms that require them to act fairly and engage in altruistic acts or feel obligated to so. Evidence for these psychological preferences and their universality comes from diverse strands of research such as neurology, anthropology and developmental psychology.

Neuroimaging studies reveal that altruistic behavior activates brain regions that are associated with cognitive and emotional empathy and reward processing (Filkowsky et al. 2016; Sonne and Gash 2018). Engaging in altruistic behavior stimulates the feel good hormones of the brain: dopamine, oxytocin and serotonin (Brueining 2016). In fact, in a clever experiment in which a large sample of people were randomly assigned to spend money on themselves or to spend it on others, Dunn and colleagues (2008) found that the group that was asked to spend it on others reported (significantly) greater happiness than the group that was asked to keep it for themselves. This evidence suggests that engaging in altruistic acts often follows from a genuine concern for others (emotional empathy<sup>1</sup>) and is inherently rewarding. This explains why many people behave altruistically (without expecting anything in return).

There is also good evidence that humans are endowed with an innate sense of fairness and a desire to act accordingly. Tomasello and colleagues found that young children possess a set of (innate) intuitions about distributive fairness. They tend to share spoils equally after having collaborated equally to obtain them—even if they could keep them for themselves (Warneken and Tomasello 2009; Warneken et al. 2011)—they understand and defend the entitlement of others (Schmidt et al. 2013), and give less to free-riders than to collaborators (Melis et al. 2013). According to Binmore (2005), a universal deep structure of fairness underlies human fairness considerations, analogically to Chomsky's (1955) deep structure underlying natural language acquisition (the so-called 'universal grammar'). In support of his claim, Binmore (2005) points at strong cross-cultural similarities in human fairness norms. This innate sense of fairness (present in very young children and in all cultures), explains why people would behave altruistically in certain contexts (such as proposing equal divisions in dictator games and cooperating in one-shot prisoner's dilemma).

Finally, people tend to follow social norms. According to Bicchieri (2005, p. 42), they can be moved to do so for a number of reasons. Their compliance may be born

<sup>1</sup> See Bateson (2011) for an extensive account of how empathic concern produces altruistic motivation.

out of fear of incurring reputation damage or of being punished (informally) by their peers for breaking the rules. People may also follow norms because they accord value to these norms or because they want to fulfill the legitimate expectations of others. Whatever the underlying reason, it stands beyond doubt that most humans have the inclination to follow social norms. As Bicchieri (2005, p. 55) rightly points out: if this were not the case, social norms could not exist. Interestingly, cross-cultural research with behavioral game-theoretic experiments gauging altruism and fairness in different societies, shows that the actions of participants in these games tend to mirror the patterns of interaction in their society (Gintis 2006, p. 26). In other words, participants often follow the social norms that govern the social interaction in their societies. The reason for many altruistic acts therefore may be that people follow social norms requiring them to act altruistically.

While these proximate explanations of human altruistic behavior—the underlying psychological features (and the neurological underpinnings)—are well documented and widely accepted, the same cannot be said for the ultimate explanation. Why did humans evolve such altruistic dispositions in the first place?

### 3 Group selection

Ultimate explanations of the evolution of altruistic dispositions, leading to behavior that benefits others in the group at the expense of the altruistic individual, often invoke group selection. The reasoning goes as follows: groups of altruists have a higher fitness than (and often outcompeted) groups of non-altruists. Therefore, altruistic individuals making up these successful groups, generously contributed to the genepool. Groups of non-altruists—on the other hand—eventually perished, so their members left no descendants. Darwin himself pointed this out:

When two tribes of primeval man, living in the same country, came into competition, if (other things being equal) the one tribe included a great number of courageous, sympathetic and faithful members, who were always ready to warn each other of danger, to aid and defend each other, this tribe would succeed better and conquer the other. (Darwin 1871, p. 166).

Ancestral human groups had much to gain with altruistic cooperation. Altruistic cooperation makes hunting more successful since it allowed ancestral humans to take down big game. It also reduces the risk of famine through food sharing. It provides a huge advantage when it comes to warfare: imagine a group of individuals willing to risk life and limb for the group facing a group of individuals not willing to do so (Bowles and Gintis 2011, pp. 3–4, Wilson 2005, p. 12). Finally, it allows for cooperative child rearing, in which ‘allo-parents’ share some of the long and arduous work to raise children. This raises the reproductive success of members in the group (Hrady 2009). Given the large benefits produced by altruistic cooperation, it safe to assume that throughout human evolutionary history groups of altruistic cooperators would have thrived at the expense of groups of non-altruists.

However, any explanation that invokes the (natural) selection of traits that benefit the group at the expense of the individual faces an obvious challenge. The consensus



among evolutionary biologists is that natural selection will retain traits in organisms that provide those organisms with an advantage in terms of survival and reproduction over conspecifics that do not possess these traits or possess them to a lesser extent. So how could altruistic traits have been selected? Free-riders would readily drive altruists to extinction within the group. They would profit from the altruism of others without bearing any of the costs of altruism and pass down their egotistical genes in greater numbers than the altruists would. Darwin (1871, p. 88) understood this too: “He who was ready to sacrifice his life, as many a savage has been, rather than betray his comrades, would often leave no offspring to inherit his noble nature.” So, how could such a trait evolve?

One possible explanation is that free-riding within groups is prevented and that natural selection driven by between group dynamics (selecting for group beneficial traits in individuals) offsets natural selection driven by within group dynamics (selecting for individually beneficial traits). This is Sober and Wilson’s (1998) view. They propose a so-called ‘multi-level selection’ account, claiming that not only genes and/or organisms are units of natural selection but also groups. While theoretically possible, this conjecture faces important and potentially debilitating problems (that I will mention below).

There is however another explanation for the evolution of altruistic dispositions in humans, which does not face these problems and is supported by extensive evidence. Between-group competition did not select directly for ‘altruistic genes’ in humans but selected (culturally not biologically) for strong prosocial norms in groups. These cultural features, in turn, have shaped a radically altered social environment in which altruistic traits are naturally selected because they boost the fitness of individuals. Such an explanation invokes cultural group selection (and gene-culture coevolution) rather than genetic group selection.

This distinction between ‘genetic’ and ‘cultural’ group selection is not always clearly made in the literature. Many accounts invoke both kinds of group selection. Sober and Wilson (1998, p. 147), for instance, argue that groups needed rules and regulations to become adaptive units (of natural selection). In other words, cultural group selection yielded rules and regulations, which then brought about a process of genetic group selection. Similarly mixed accounts have been proposed by others (e.g. Boehm 1997; Wilson and Kniffin 1999; Wilson 2005; Fehr and Gächter 2002b; Gintis et al. 2003; Bowles and Gintis 2011). This may have led sceptics (such as Pinker 2012) to dismiss any form of group selection, including cultural group selection, at the outset (while only offering arguments targeted at genetic group selection).

### 3.1 Genetic group selection

Sober and Wilson (1998, see also Wilson 2005) argue that individual natural selection cannot select for altruistic behavior because such behavior decreases the relative fitness of individuals within the group and would be selected against. Therefore, they conclude, it must have been naturally selected at the level of groups. The position they defend is often referred to as ‘multi-level selection’ (Sober and Wilson 1998; Okasha 2005): natural selection does not only act on the level of individuals, but also on the

level of groups. Altruistic behavioral dispositions, by this rationale, evolved because natural selective pressure at the level of the group outweighed selective pressure at the level of the individual.

Simply put, advocates of explanations of altruism in terms of genetic group selection claim that altruistic dispositions evolved because altruistic individuals making up altruistic groups had greater reproductive success than less altruistic individuals making up less altruistic groups. Group selection in this explanation is acting directly on the genome. Such a position is not only championed by Sober and Wilson (1998), others followed in their wake (e.g. Okasha 2005; Fletcher and Doebeli 2009; Bravetti and Padilla 2018).

The obvious challenge to genetic group selection accounts of the evolution of human altruism is that individual selection is a prominent driver of evolution. It is hard to imagine that altruistic groups would not be invaded by free-riders outcompeting them and driving them to extinction. In response, genetic group selectionists invoke assortative interaction (Sober and Wilson 1998, p. 135) or correlated interaction (Okasha 2005). They rightly argue that if there are mechanisms in place so that altruists only interact with other altruists, they can avoid being ‘suckered’ and outcompeted by free-riders. Such altruistic clusters would then have a marked evolutionary advantage over less altruistic groups and their genetic endowment would spread in the human gene pool. There is good evidence that humans did evolve cognitive faculties devoted to the detection of ‘cheaters’ (Cosmides and Tooby 2005) and to reputation tracking (Mealey et al. 1996; Oda 1997) together with altruistic dispositions. This would have protected altruists against the exploitation of free-riders and explains why reciprocal altruism occurs (not only in human groups but also in groups of other species such as certain bird species, vampire bats and meerkats, as pointed out above).

As critics have pointed out, however, postulating that there was genetic group selection of human traits requires us to make a series of additional assumptions that are problematic. First and foremost, it assumes that there was substantial genetic variation between human groups *and* that there was limited migration between groups (which is necessary to sustain genetic variation between groups). Moreover, it assumes that there was a considerable rate of group extinction and that successful groups split up to form more groups (reproducing or replicating as organisms and genes do) (Maynard Smith 1976; Pinker 2012; Richerson et al 2016). Therefore, the majority of evolutionary scientists are highly skeptical of theories advocating genetic group selection of human traits. When the famous biologist Edward Wilson (not to be confused with David Wilson mentioned above) wrote an article in which he defended genetic group selection with colleagues Nowak et al. (2010), 137 scientists responded in a joint paper strongly contesting their views (Abbot et al. 2011).<sup>2</sup> Reviewing the arguments and counter-arguments is beyond the scope of this paper. For our purposes, it suffices to say that—while the jury is still out—the majority of evolutionary scientists reject

<sup>2</sup> The response paper by Abbot et al. (2011) is situated in the debate about the evolution of eusociality. Nowak et al. (2010) argued against Hamilton (1964) that the evolution of eusociality (for instance among bees and ants belonging to the same hive or colony— see above) cannot be satisfactorily explained by invoking kin-selection and inclusive fitness. They believe that eusociality evolved primarily through genetic group selection. Abbot and colleagues (2011) respond that Hamilton’s (1964) theory stands and that the evolution of eusociality can and should be explained in terms of inclusive fitness.

explanations of human altruism in terms of genetic group selection. What I will argue below is that we do not need to invoke this controversial evolutionary mechanism to explain human altruism.

Sober and Wilson (1998) are right—I believe—in claiming that between-group dynamics are the architect of certain remarkable human altruistic dispositions. We cannot explain the evolution of human altruistic dispositions solely in terms of inclusive individual fitness (given that it is often directed at non-kin) and reciprocity (given that it is often directed at people who cannot reciprocate). This, however, does not entail that we need to go up a level of natural selection (the group level). Contra Sober and Wilson (1998), I will argue that it is standard individual natural selection that selected for altruistic dispositions in humans. How is this possible? Doesn't altruism reduce individual fitness and shouldn't it therefore be selected against at the individual level? To answer this question, we must insert culture and cultural evolution into the equation.

### 3.2 Cultural group selection (and gene-culture coevolution)

Evolutionary processes do not only shape the genome of organisms, they also shape features of human cultures (such as beliefs, customs and norms). In previous work (Vlerick 2016, 2020a, b), I have developed a model of cultural evolution in which I identify within and between group dynamics as the main drivers of cultural selection. Within group dynamics select for cultural features that are psychologically attractive (or beliefs that are memorable) and are therefore taken up and transmitted by group members. Between group dynamics select for cultural features that provide the group with an advantage over other groups that do not possess these cultural features (or possess them to a lesser extent).

Between group dynamics select—among other things—for prosocial norms and punishments. This enhances the (altruistic) cooperation within the group (Vlerick 2020a, b). In particular, competition between groups selects for norms (and punishments) that reduce conflict and enable and protect altruistic cooperation within groups (Aviles 2002; Boyd et al 2003; West et al. 2007; Puurtinen and Mappes 2009). The selective pressure arising from group competition is what Boyd, Richerson, Henrich and others refer to as 'cultural group selection' (Boyd and Richerson 1985, 2002; Richerson and Boyd 1999; Henrich 2004; Richerson et al. 2016).

Several important factors underlie the cultural selection or proliferation of group beneficial social norms and punishments. In direct conflict between groups, other things being equal, the most cooperative group is more likely to be victorious and conquer the other group. In competition between groups over scarce resources, cooperative groups are likely to outcompete less cooperative groups (and survive while the other groups perish). More cooperative groups are also more likely to produce more wealth which throughout human history (until very recently) correlated with demographic expansion and can lead to the demographic swamping of less successful groups. Finally, individuals from less wealthy groups often migrate to wealthier groups and the customs and norms of successful groups are often imitated by less successful neighboring groups (Bowles and Gintis 2011, p. 50).

For all of these reasons, social norms underlying extensive (and altruistic) in-group cooperation are likely to proliferate. Ethnographic analogues suggest that Pleistocene hunter-gatherer groups possessed such complex sets of rules regulating the interaction of individuals within the group, that there were substantial differences (with respect to these sets of rules) between different groups and that there was frequent (competitive) interaction between groups (Hill et al. 2014; Richerson et al. 2016). In such a context, between group dynamics must have been a prominent driver of cultural evolution.

Because of this cultural evolutionary process driven by between group competition, societies emerged that were increasingly governed by prosocial norms and punishments. Such a social environment did in turn have a strong effect on the biological evolution of ancestral humans. It (naturally) selected for cooperative, norm-abiding and altruistic individuals. Prosocial norms and punishment in ancestral societies did not only ensure that free-riders did not get away with their cooperation eroding behavior (they are being punished) and that consequently altruism could be sustained within groups (Vlerick 2016, 2020a), over time they also shaped the genome of the individuals inhabiting those societies. Because with such a normative framework in place, the egoists and the sociopaths are reliably punished (which included banishment and murder) for their anti-social behavior and would be less likely than norm abiding altruists to spread their (antisocial) genes.

In short, a culturally evolved highly cooperative niche radically changed the social environment in which human genetic evolution took place. It produced what Henrich (2016, p. 185) refers to as a process of ‘self-domestication’. Humans did not only domesticate animal species (e.g. turning wolves into dogs by selectively breeding with the most docile animals), they inadvertently did something similar to themselves. By consequently punishing egotistical, unruly and overly aggressive individuals and preventing them from spreading their sociopathic genes, humans were selectively ‘bred’ with those individuals that happened to have an inclination to follow social norms and behave altruistically.

This process is an instance of what Richerson and Boyd (1985, 2005) call ‘culture-led gene-culture coevolution’. A culturally evolved social environment steered human genetic evolution. Human culture and biology co-evolved, leading to ever more altruistic humans. The key to explaining the strong altruistic dispositions of many people—inciting to them to behave altruistically towards strangers without expecting anything in return—lies in the uniqueness of this behavior in the animal kingdom. It evolved in response to an equally unique feature of human life: complex culture with prosocial norms and punishments, which in turn had the power to shape the human genome. Any account of the evolution of human altruism and human moral psychology in general that doesn’t take into account the cultural context in which this evolution took place—such as explanations solely in terms of genetic group selection—misses the central cause.

### 3.3 A culturally evolved highly cooperative niche

Human altruistic dispositions, I have argued, were naturally selected in social environments characterized by ever more stringent prosocial norms, extensive monitoring

of group members and harsh punishment of those not abiding by these norms (harsh enough to decrease their reproductive success). In this culturally evolved context, norm abiding altruists had an evolutionary advantage over their more selfish and unruly peers. Any attempt to reverse engineer the environmental context in which traits evolved, however, invites the criticism of being ‘just-so-stories’. Therefore, in this section, I will discuss the evidence supporting my hypothesis.

### 3.4 Evidence

‘Sanctions for crimes against the collectivity’ features on Brown’s (1991) famous list of human universals. All current human societies have such (formal and/or informal) sanctions, including the few remaining hunter-gatherer societies. While anthropological evidence for the universality of prosocial punishment is of course no guarantee that ancestral human (hunter-gatherer) societies would have possessed such prosocial punishments, it is nevertheless a good indication that they had. According to Boehm (1997) and Bowles and Gintis (2011, p. 5), such prosocial punishments were facilitated by the possession of projectile weapons which enabled groups of people to collectively punish norm violators (e.g. by banishing or murdering them) at relatively low risk to each individual punisher.

However, as Henrich (2010, pp. 187–188) points out, punishment of norm violators in small-scale societies doesn’t typically take on this harsh form. It often starts with gossip and ridicule and—if the norm violator doesn’t redeem him or herself—punishment is ramped up leading to exclusion from marital prospects and from trading partners. Only as a last resort does it escalate to banishment, physical violence and coordinated group executions. Henrich (2010) finds support for the universality of such prosocial punishments in small-scale societies in studies on a wide range of different ethnic groups (see Boehm 1993; Chudek and Henrich 2011; Bowles et al. 2012; Mathew and Boyd 2011; and Wiessner 2005).

This is not surprising. Developmental research has brought to light that children are prone to punish rule breakers and free-riders at a very young age (Melis et al. 2013). This points at an innate human desire to punish rule breakers. Moreover, in all cultures people are socially reprimanded for violations of rules of conduct that do not actually harm anybody (such as violating a dietary taboo or ignoring a social convention). Demanding that others conform to the social rules and punishing (often in subtle ways) those who do not, seems to be deeply ingrained in human nature. The prevalence of such punishments combined with effective monitoring of social behavior would have reliably disadvantaged individuals less prone to follow social norms and individuals who repeatedly put their own interests before those of others.

Effective monitoring, in turn, is facilitated by reputation tracking and by exchanging social information. There is equally good evidence for the prevalence of these activities in all human societies. According to Trivers (1971) and Panchanathan and Boyd (2004), reputation tracking is another human universal. It is a common occurrence in all human societies and there is no reason to believe that it wasn’t equally prevalent in pre-agricultural societies inhabited by biologically modern human beings. According to Dunbar (1996) language evolved (gradually) in the human lineage for this very

purpose. He proposes a so-called ‘gossip theory’ of the evolution of language, in which he argues that language evolved for social bonding and to exchange social information. Language enabled our ancestors to form close ties with a relatively large number of individuals (about 150 individuals according to Dunbar) and enabled them to acquire and transmit information to others. This protected them against the exploitation of free-riders (see also Enquist and Leimar 1993). In support of his hypothesis, Dunbar points out that ‘gossiping’ (exchanging social information) is still language’s most prominent function. A whopping sixty percent of casual human conversations are about other people (Dunbar et al. 1995).

From these strands of evidence emerges a picture of the societal context in which our recent evolutionary history took place: a context characterized by demanding (pro)social norms, incessant monitoring whether or not individuals abide by these norms and hard to escape punishments for those breaking the rules.

### 3.5 Maladaptive in modern contexts?

Tooby and Cosmides (1996, p. 122) and Dawkins (1976, p. 220) have argued that there is a mismatch between our ancestral social context—in which most interactions took place between genetically related individuals or closely acquainted, reciprocating individuals—and our modern social context, characterized by its many interactions between total strangers. Therefore, they argue, human altruistic dispositions were adaptive in our ancestral context (which is why they evolved), but are actually maladaptive in the modern context. They no longer increase the (long-term, inclusive) fitness of the individuals engaging in altruistic behavior but decrease it, since they lead people to behave altruistically towards total strangers with no chance of reciprocation.

By this rationale, people donate blood and funds to strangers and behave altruistically in one-shot prisoner’s dilemma and dictator games, because natural selection has hardwired them to behave altruistically (see proximate explanations of human altruism—Sect. 2). While this was adaptive in ancestral times, it is maladaptive in modern times. In other words, the reason why humans engage in these peculiar (and allegedly maladaptive) forms of altruistic behavior, is because their evolved social nature ‘misfires’ in a modern context. Compare it with our craving for sweet tasting food and drinks. These cravings were adaptive in ancestral times, where they motivated humans to consume ripe fruit containing the necessary carbohydrates and vitamin C, but are maladaptive in modern environments filled with cheap and unhealthy candy and soft drinks.

Under scrutiny, however, the mismatch hypothesis to explain human altruism does not hold up. Firstly, as Hill and colleagues (2011, 2014) have pointed out, hunter-gatherer societies are relatively open social systems. In all likelihood, our ancestors would have interacted with an important number of people outside of their tribe (e.g. to trade). In other words, (paleo) anthropological evidence seems to refute the premise that our ancestors only interacted with kin and people to whom they were closely acquainted.

Moreover, I discern two important problems with Tooby and Cosmides (1996, p. 122) and Dawkins' (1976, p. 220) mismatch hypothesis to explain human altruism. The first is that it seems to assume that people evolved to be indiscriminate altruists (leading them to behave altruistically towards non-reciprocating strangers today). This is not the case. As pointed out above, humans have evolved a range of cognitive skills and dispositions—such as a ‘cheater detection module’, the ability and desire to track the reputation of others and to exchange social information with others—precisely to be discriminate altruists. When people behave altruistically towards total strangers, they are not ‘fooled’ by a confusing modern context. They typically do so because they empathize with these strangers and decide it is the morally right thing to do.

This brings me to the second problem with the mismatch hypothesis (and most other evolutionary explanations of human altruism). It assumes that altruistic behavior is solely the result of evolved, ‘hardwired’ psychological mechanisms adapted to the ancestral social environment. As I will argue in the next section, underlying actual altruistic behavior are not merely evolved intuition and emotion-based dispositions but also conscious and voluntary reasoning processes. Many scientific accounts of human altruism ignore<sup>3</sup> the important role of these reasoning processes (or at least, the causal role of these reasoning processes remains underdeveloped in said accounts). They often look no further than the evolutionary rationale underlying altruistic behavior and miss a very important piece of the puzzle.

## 4 The role of reasoning

Moral decisions—such as the decision to cooperate in one-shot prisoner’s dilemma, propose fair divisions in dictator games and donate blood—are not merely the outcome of hard-wired emotion and intuition-based processes. They also involve reasoning processes. In a landmark experimental study subjecting participants to brain scans while presenting them with moral dilemmas, Greene and colleagues (2001) found that next to an emotional cognitive subsystem, we employ a reason-based cognitive subsystem in moral evaluation and decision-making. Whereas the emotional system often floods our moral thinking automatically and subconsciously, the reasoning system can in some cases override its output and generally takes over when presented with moral problems for which we have no ready-made, automatic, intuition or emotion-based response (see also Greene 2013 and Vlerick 2017).

Therefore, if we want to explain human altruistic behavior we should not only take into account the evolution of the intuition and emotion-based psychological dispositions (which I have described in Sect. 2 on ‘proximate explanations’). We must also take into account conscious and voluntary reasoning processes involved in moral decision-making. These reasoning processes, I will argue below, have a major impact on moral behavior in general and altruism towards out-group strangers in particular.

<sup>3</sup> With the notable exception of Darwin (1871, p. 185–186) who identifies the following causes of the advance of morality: “the approbations of our fellow men—the strengthening of our sympathies by habit—example and imitation—*reason*—experience, and even self-interest—instruction during youth, and religious feelings” (my italics).

## 4.1 The ‘escalator’ effect of reasoning on morality

The moral behavior some people engage in is far-removed from the kind of behavior we would expect given the adaptive rationale of the psychological dispositions underlying this behavior. Our moral psychology, as argued above, evolved as an adaptation to a highly cooperative niche characterized by strong prosocial norms and punishments that orchestrated *in-group* interaction. In other words, our moral psychology evolved for altruistic cooperation within the groups in which we live. Yet humans routinely engage in altruistic acts directed at obvious out-group members (and even go so far as to behave altruistically towards non-human animals and future, unborn generations). This is puzzling.

Tooby and Cosmides (1996) and Dawkins’ (1976) attempt to explain this by arguing that our moral psychology ‘misfires’ in modern multicultural contexts is—as argued above—problematic. Western people who donate funds to starving Africans know very well that they are doing so for the benefit of ‘out-group’ individuals. They are not fooled by a confusing modern context, but consciously decide to help those in need, regardless of their culture or ethnicity (Vlerick 2017). This kind of moral behavior is not rooted in (intuition or emotion-based) psychological mechanisms which evolved for in-group (altruistic) cooperation. It is the outcome of conscious reasoning processes.

Peter Singer (1995, p. 226) refers to this as the ‘escalator of reason’. Reasoning about morality can lead to behavior and moral norms that are far-removed from the behavior for which our moral psychology evolved. Altruism towards strangers with no chance of reciprocation in the absence of any social expectation or potential reputation gain—such as anonymous charity donations, anonymous (and unadvertised) blood donations, cooperating in anonymous single shot prisoner dilemma’s with strangers and anonymous fair offerings in dictator games—is an instance of such behavior. These moral actions are not merely the output of hardwired psychological dispositions (which explains why many people do not engage in these altruistic acts). They often involve moral reasoning. Interestingly in this regard, a study has brought to light that altruistic behavior correlates with level of education (Westlake et al. 2019). The authors of the study surmise that people who benefited from a higher level of education might be better at internalizing prosocial norms. I would add that people who benefited from a higher level of education might also be better trained in reasoning about moral issues and reflecting on their moral behavior.

## 4.2 Norm abidance or reasoning?

A rival explanation for altruistic behavior that goes beyond the kind of behavior we would expect from an evolutionary perspective is that people just follow social norms that happen to impose or at least encourage this kind of altruistic behavior. So, rather than behaving altruistically after autonomous moral reasoning or reflection, people could simply be abiding by social norms or social expectations. This is a valid point. Norm abidance is indeed a major cause of altruistic behavior (see Sect. 2 on proximate explanations). As pointed out, data gathered from behavioral game-theoretic experi-



ments in different cultural contexts shows that people tend to follow the social norms that govern their societies in these experiments (Gintis 2006).

However, social norm following does not explain all altruistic deeds. There is no social norm that requires people to donate blood in contemporary societies (people are not socially reprimanded for not donating blood), yet some people regularly volunteer to do so. While they might do so for a variety of reasons—including virtue signaling—moral reasoning is likely to be an important factor. Campaigns for blood donation typically try to persuade people to donate by presenting the public with arguments (e.g. ‘you can save lives’). In other words, these campaigns trigger moral reasoning processes in potential donors, hoping they will make a conscious moral decision to donate.

Moreover, even if many people engage in altruistic acts directed at non-kin with no chance of reciprocation because they abide by social norms or expectations, conscious reasoning processes are still part of the explanation of these altruistic acts. Most of these norms saw the light because individuals challenged the status quo through moral reasoning and because (many) others accepted the new moral imperative after evaluating the reasons offered in support of this imperative. It is only once a social norm is ‘established’ that people abide by it without reflection. Even in cases of social norm following, reasoning processes (albeit of others) are therefore still part of the picture. They explain why these norms arose in the first place.

### 4.3 An evolved moral compass powered by reason

So, in answering the question why humans routinely engage in altruistic behavior towards non-kin and with no chance of reciprocation, the evolution of altruistic dispositions only provides us with half of the explanation. In addition to evolved moral intuitions and emotions (such as empathy and norm abidance), we must take into account reasoning processes that underlie moral decisions and behavior. This however does not diminish the importance of the evolution of these altruistic dispositions in explanations of human altruism. Reasoning processes—which are content-free—will not lead to moral behavior by themselves. They must latch onto ‘moral’ and ‘altruistic’ psychological dispositions such as a sense of fairness (Binmore 2005) and empathy.<sup>4</sup>

These evolved psychological dispositions provide our moral reasoning processes with a direction. They provide us with what I have called ‘an evolved moral compass’ in previous work (Vlerick 2017). Such a moral compass powered by reason—I have argued—is the driver of moral progress. Without reasoning processes there would be no way to challenge the moral status quo. Singer’s ‘escalator’ would disappear. Without an innate (intuition and emotion-based) moral compass, reasoning would not lead to moral or altruistic behavior. In the absence of these prosocial dispositions, it is safe to assume that we would apply our reasoning processes in our self-interest and the interest of close kin. What explains the uniqueness of human altruism—the fact

---

<sup>4</sup> Tomasello and colleagues (2005) argue that human empathy (rooted in a theory of mind) leads to ‘self-other equivalence’. In contrast to other primates, humans view their conspecifics as ‘other selves’ which are fundamentally no different than oneself, rather than viewing them as mere elements of the social environment.

that it is often directed at non-kin with no chance of reciprocation—is precisely this powerful combination of a highly prosocial nature (adapted to a highly cooperative social context) and our ability to take our prosocial behavior to the next level by reflecting on moral norms, decisions and behavior.

## 5 Conclusion

Human altruism is exceptional in the animal kingdom. In no other species has widespread (biological) altruism directed at non-kin, with no chance of reciprocation, been observed. This remarkable behavior has puzzled evolutionists since Darwin and attempts to explain human altruism have created a lot of confusion and debate. It has led many scholars to develop group selection theories, which in turn have been heavily criticized. Explanations of human altruism are still the subject of much (and heated) debate today, but often the debate suffers from a lack of clarity. It is not always clear what exactly ‘group selection’ refers to and different scholars use it in different ways. As Maynard Smith (1998) rightly points out in his review of Sober and Wilson’s (1998) ‘Unto others’—in which they develop their group selection account of human altruism—the discussion has often turned semantic, with quarreling parties mainly disagreeing on the appropriate terminology rather than the underlying processes they describe.

In response, I set out to create some much needed clarity to this incendiary debate by clearly distinguishing genetic from cultural group selection. The latter does not face the difficulties associated with the former and (together with gene-culture coevolution) provides us with an empirically supported hypothesis of the evolution of the strong altruistic dispositions of humans. Evolved psychological dispositions, however, do not suffice to explain many instances of actual human altruistic behavior. The final aim of this paper, therefore, was to complete extant scientific explanations of human altruism that have focused solely on its evolutionary underpinning. If we want to make sense of human altruism, we must take into account conscious and voluntary reasoning processes, creating—as Singer (1995) has called it—an ‘escalator’ effect on moral behavior and norms. Underlying the uniqueness of human altruism are two equally unique human attributes: the social norms and punishments that govern our societies and the reasoning processes we unleash on the evaluation of moral norms and decisions.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abbot, P., et al. (2011). Inclusive fitness theory and eusociality. *Nature*, 471(7339), E1–E4.
- Ananth, M. (2005). Psychological altruism vs. biological altruism: Narrowing the gap with the Baldwin effect. *Acta Biotheoretica*, 53, 217–239.
- Aviles, L. (2002). "Solving the freeloader paradox: Genetic associations and frequency dependent selection in the evolution of cooperation among nonrelatives. *PNAS*, 99(22), 14268–14273.
- Bateson, C. D. (2011). *Altruism in humans*. Oxford: Oxford University Press.
- Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. New York: Cambridge University Press.
- Binmore, K. (2005). *Natural justice*. Oxford: Oxford University Press.
- Boehm, C. (1993). Egalitarian behavior and reverse dominance hierarchy. *Current Anthropology*, 34(3), 227–254.
- Boehm, C. (1997). Impact of the human egalitarian syndrome on Darwinian selection mechanics. *The American Naturalist*, 150(Suppl.), 100–121.
- Bowles, S., Boyd, R., Mathew, S., & Richerson, P. J. (2012). The punishment that sustains cooperation is often coordinated and costly. *Behavioral and Brain Sciences*, 35(1), 20–21.
- Bowles, S., & Gintis, H. (2011). *A cooperative species: Human reciprocity and its evolution*. Princeton: Princeton University Press.
- Boyd, R., Gintis, H., Bowles, S., & Richerson, P. (2003). The evolution of altruistic punishment. *PNAS*, 100(6), 3531–3535.
- Boyd, R., & Richerson, P. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.
- Boyd, R., & Richerson, P. (2002). Group beneficial norms can spread rapidly in a structured population. *Journal of Theoretical Biology*, 215, 287–296.
- Bravetti, A., & Padilla, P. (2018). An optimal strategy to solve the Prisoner's Dilemma. *Nature, Scientific reports*, 8, 1948.
- Bruening, L. (2016). *Habits of a happy brain: Retrain your brain to boost your serotonin, dopamine, oxytocin and endorphin levels*. Avon, MA: Adams Media.
- Brown, D. (1991). *Human universals*. New York: McGraw-Hill.
- Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton: Princeton University Press.
- Camerer, C., & Thaler, R. (1995). Anomalies: Ultimatums, dictators and manners. *Journal of Economic Perspectives*, 9(2), 209–219.
- Chomsky, N. (1955). Logical syntax and semantics: Their linguistic relevance. *Language*, 31(1), 36–45.
- Chudek, M., & Henrich, J. (2011). Culture-gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences*, 15(5), 218–226.
- Cosmides, L., & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. In D. M. Buss (Ed.), *Handbook of evolutionary psychology* (pp. 584–627). Hoboken, NJ: Wiley.
- Darwin, C. (1871). *The descent of man and selection in relation to sex*. London: John Murray.
- Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.
- Dawkins, R. (1994). Burying the vehicle commentary on Wilson & Sober: Group selection. *Behavioral and Brain Sciences*, 17(4), 616–617.
- Dennett, D. C. (1994). E Pluribus Unum? Commentary on Wilson & Sober: Group selection". *Behavioral and Brain Sciences*, 17(4), 617–618.
- Dunbar, R. (1996). *Grooming, gossip and the evolution of language*. London: Faber and Faber.
- Dunbar, R. I. M., Duncan, N. D. C., & Nettle, D. (1995). Size and structure of freely forming conversational groups. *Human Nature*, 6, 67–78.
- Dunn, E., Aknin, L., & Norton, M. (2008). Spending money on other promotes happiness. *Science*, 319, 1687.
- Enquist, M., & Leimar, O. (1993). The evolution of cooperation in mobile organisms. *Animal Behaviour*, 45(4), 747–757.
- Fehr, E., & Gächter, S. (2002a). Altruistic punishment in humans. *Nature*, 415, 137–140.
- Fehr, E., & Gächter, S. (2002b). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, 13(1), 1–25.
- Fehr, E., & Rockenbach, B. (2004). Human altruism: economic, neural, and evolutionary perspectives. *Current Opinion in Neurobiology*, 14, 784–790.

- Filkowski, M. M., Cochran, R. N., & Haas, B. W. (2016). Altruistic behavior: Mapping responses in the brain. *Neuroscience and Neuroeconomics*, 5, 65–75.
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford: Clarendon Press.
- Fletcher, J., & Doebeli, M. (2009). A simple and general explanation for the evolution of altruism. *Proceedings of the Royal Society B*, 276, 13–19.
- Gächter, S., & Herrmann, B. (2009). Reciprocity, culture and human cooperation: Previous insights and a new cross-cultural experiment. *Proceedings of the Royal Society B*, 364, 791–806.
- Gintis, H. (2006). Behavioral game theory and sociology. Retrieved from: <https://www.umass.edu/preferen/gintis/behavi~1.pdf>
- Gintis, H., Bowls, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior*, 24, 153–172.
- Greene, J. (2013). *Moral tribes: Emotions, reason, and the gap between us and them*. New York: Penguin Press.
- Green, J., Sommersville, R., Nystrom, L., Darley, J., & Cohen, J. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105–2108.
- Haldane, J. B. S. (1932). *The causes of evolution*. London: Longmans, Green & Co.
- Hamilton, W. (1964). The Genetical evolution of social behaviour I and II. *Journal of Theoretical Biology*, 7, p. 1–16, and p. 17–52.
- Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization*, 53(1), 3–35.
- Henrich, J. (2010). *The secret of our success: How culture is driving human evolution, domesticating our species and making us smarter*. Princeton: Princeton University Press.
- Henrich, J. (2016). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.
- Henrich, J., Boyd, R., Bowles, S., Gintis, H., Camerer, C., Fehr, E., et al. (2001). In search of Homo economicus: Experiments in 15 small-scale societies. *American Economic Review*, 91, 73–78.
- Hill, K. R., Walker, R., Božičević, M., Eder, J., Headland, T., Hewlett, B., et al. (2011). Co-residence patterns in Hunter-Gatherer societies show unique human social structure. *Science*, 331(6022), 1286–1289.
- Hill, K. R., Wood, B. M., Baggio, J., Hurtado, A. M., & Boyd, R. T. (2014). Hunter-gatherer inter-band interaction rates: Implications for cumulative culture. *PLoS ONE*, 9(7), e102806.
- Hrdy, S. (2009). *Mothers and others: The evolutionary origins of mutual understanding*. Cambridge: Harvard University Press.
- Mathew, S., & Boyd, R. (2011). Punishment sustains large-scale cooperation in prestate warfare. *Proceedings of the National Academy of Sciences, USA*, 108(28), 11375–11380.
- Maynard Smith, J. (1976). Group selection. *Quarterly Review of Biology*, 51(2), 277–283.
- Maynard Smith, J. (1998). Book review: Sober, E., Wilson, D. (1998). *Unto others: The evolution and psychology of unselfish behavior* Harvard University Press: 1998. *Nature*, 393, p. 639–640.
- Mealey, L., Daoud, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology*, 17, 119–128.
- Melis, A., Altricher, A., Schneider, A., & Tomasello, M. (2013). Allocation of resources to collaborators and free-riders by 3-years-old. *Journal of Experimental Child Psychology*, 114, 364–370.
- Nowak, M. A., Tarnita, C. E., & Wilson, E. O. (2010). The evolution of eusociality. *Nature*, 466(7310), 1057–1062.
- Oda, R. (1997). Biased face recognition in the prisoner's dilemma games. *Evolution and Human Behavior*, 18, 309–315.
- Okasha, S. (2005). Altruism, group selection and correlated interaction. *British Journal for the Philosophy of Science*, 56(4), 703–725.
- Panchanathan, K., & Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free-rider problem. *Nature*, 432, 499–502.
- Pinker, S. (2012). The false allure of group selection. Edge June 19, 2012. Retrieved from: [https://www.edge.org/conversation/steven\\_pinker-the-false-allure-of-group-selection](https://www.edge.org/conversation/steven_pinker-the-false-allure-of-group-selection)
- Puurtinen, N., & Mappes, T. (2009). Between-group competition and human cooperation. *Proceedings of the Royal Society B*, 276, 355–360.
- Richerson, P., & Boyd, R. (1999). Complex societies: The evolutionary origins of a crude superorganism. *Human Nature*, 10(3), 253–289.
- Richerson, P., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.

- Richerson, P., et al. (2016). Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behavioral and Brain Sciences*, 39, e30.
- Ruse, M. (1979). *Sociobiology: Sense or nonsense?*. Dordrecht: Reidel.
- Schmidt, M., Rakoczy, H., & Tomasello, M. (2013). Young children understand and defend the entitlements of others. *Journal of Experimental Child Psychology*, 116(4), 930–944.
- Singer, P. (1995). *How are we to live? Ethics in an age of self-interest*. New York: Prometheus Books.
- Sober, E. (1988). What Is Evolutionary Altruism? In B. Linsky and M. Matthen (eds.), *New Essays on Philosophy and Biology*, Canadian Journal of Philosophy, supplementary volume 14, p. 75–99.
- Sober, E., & Wilson, D. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge MA: Harvard University Press.
- Sonne, J., & Gash, D. (2018). Psychopathy to altruism: Neurobiology of the selfish–selfless spectrum. *Frontiers in Psychology*, 9, 575.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28, 675–735.
- Tooby, J., & Cosmides, L. (1996). Friendship and the banker’s paradox: Other pathways to the evolution of adaptations for altruism. In W. G. Runciman, J. Maynard Smith, & R. I. M. Dunbar (Eds.), *Evolution of social behaviour patterns in primates and man* (Vol. 88, pp. 119–143). Proceedings of the British Academy.
- Trivers, R. (1971). The Evolution of Reciprocal Altruism. *Quarterly Review of Biology*, 46, 35–57.
- Vakoch, D. A. (Ed.). (2013). *International and cultural psychology. Altruism in cross-cultural perspective*. Berlin: Springer.
- Vlerick, M. (2016). Explaining universal social institutions: A game-theoretic approach. *Topoi*, 35(1), 291–300.
- Vlerick, M. (2017). ‘Better than our nature? Evolution and moral realism, justification, and progress’. In M. In Ruse & R. Richards (Eds.), *Handbook of evolutionary ethics* (pp. 226–239). Cambridge: Cambridge University Press.
- Vlerick, M. (2020a). The evolution of social contracts. *Journal of Social Ontology*, 5(2), 181–203.
- Vlerick, M. (2020b). The cultural evolution of institutional religions. *Religion, Brain & Behavior*, 10(1), 18–34.
- Warneken, F., Lohse, K., Melis, A., & Tomasello, M. (2011). Young children share the spoils after collaboration. *Psychological Science*, 22, 267–273.
- Warneken, F., & Tomasello, M. (2009). Varieties of altruism in children and Chimpanzees. *Trends in Cognitive Sciences*, 13(9), 397–402.
- West, S., Griffin, A., & Gardner, A. (2007). Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, 20(2), 415–432.
- Westlake, G., Coall, D., & Grueter, C. (2019). Educational attainment is associated with unconditional helping behaviour. *Evolutionary Human Sciences*, 1, E15.
- Wiessner, P. (2005). Norm enforcement among the Jul’hoansi Bushmen—A case of strong reciprocity? *Human Nature*, 16(2), 115–145.
- Wilson, D. S. (1975). A theory of group selection. *Proceedings of the National Academy of Sciences (USA)*, 72(1), 143–146.
- Wilson, D. S. (2005). *Does altruism exist? Culture, genes, and the welfare of others*. New Haven: Yale University Press.
- Wilson, D. S., & Kniffin, K. M. (1999). Multilevel selection and the social transmission of behavior. *Human Nature*, 10(3), 291–310.