

Opinion

The hidden costs of human cooperation

Jörg Gross ^{1,2,*}, Caroline Graf ¹, and Charlotte S.L. Rossetti ¹

Cooperation enables humans to reshape entire environments and build complex societies. Although often celebrated, cooperation also has hidden costs. By presenting core mechanisms behind its emergence, we demonstrate that maintaining cooperation frequently relies on social control and coercion, which can lead to extortion and discrimination. Group cooperation further necessitates defining who belongs to the group, fostering exclusion and intergroup conflict. Free-rider concerns fuel scapegoating and polarization. These downsides challenge the notion of cooperation as a simple success story. The resulting conundrum for scientists is not just to explain cooperation but to identify institutions that harness its benefits while limiting its risks. Understanding these complexities is crucial to ensuring that human cooperation serves the common good rather than deepening social divides.

The two sides of cooperation

Humans are regarded as champions of **cooperation** (see [Glossary](#)). Our cognitive abilities, empathetic concerns, and language faculties not only enable unparalleled levels of coordination and collaboration but also may have evolved in part to achieve this unique capability, making cooperation the ‘secret of our success’ [1–3].

Sustaining cooperation, however, is far from trivial and poses an evolutionary puzzle [4]. Cooperation means incurring personal costs – time, effort, or resources – to create benefits for another individual or group. While cooperation generates benefits that exceed what individuals could achieve alone, it also invites free-riding [4,5]. If others are cooperating, why not enjoy the benefits without sharing the burden? Yet, if others are free-riding, why pay the costs to benefit others if those efforts are not reciprocated? Because of this logic, comprehensively captured in the Prisoner’s Dilemma and Public Goods Game ([Figure 1](#)), a central question is how cooperation can emerge in the first place.

Highlighted as one of the 25 big open questions in science [6], cross-disciplinary research in recent decades has made remarkable progress in unraveling this puzzle and identifying the mechanisms and institutions that can overcome free-riding. We now have a solid understanding of how humans maintain cooperation [3,7–10]. Core reviews, perspectives, and meta-analyses focus on explaining humans’ extraordinary capacity for cooperation [1,2,4,7–12] while typically portraying it as a net positive that should be maximized or remaining agnostic regarding its broader consequences for both individuals and society (with some notable exceptions [3,13–17]). However, by focusing on mechanistic explanation rather than evaluating wider societal implications, previous research has largely overlooked important questions about whether and how cooperation can and should serve the common good without producing serious side effects.

Here, we provide a different perspective. Cooperation can sometimes be a curse rather than a cure for humanity’s problems. To outline this duality, we discuss six central mechanisms that

Highlights

Cooperation is a cornerstone of human evolutionary success, enabling individuals to achieve more than they could alone.

Core mechanisms have been identified that maintain cooperation and mitigate the free-rider problem associated with it, such as indirect fitness, direct and indirect reciprocity, partner choice, or punishment.

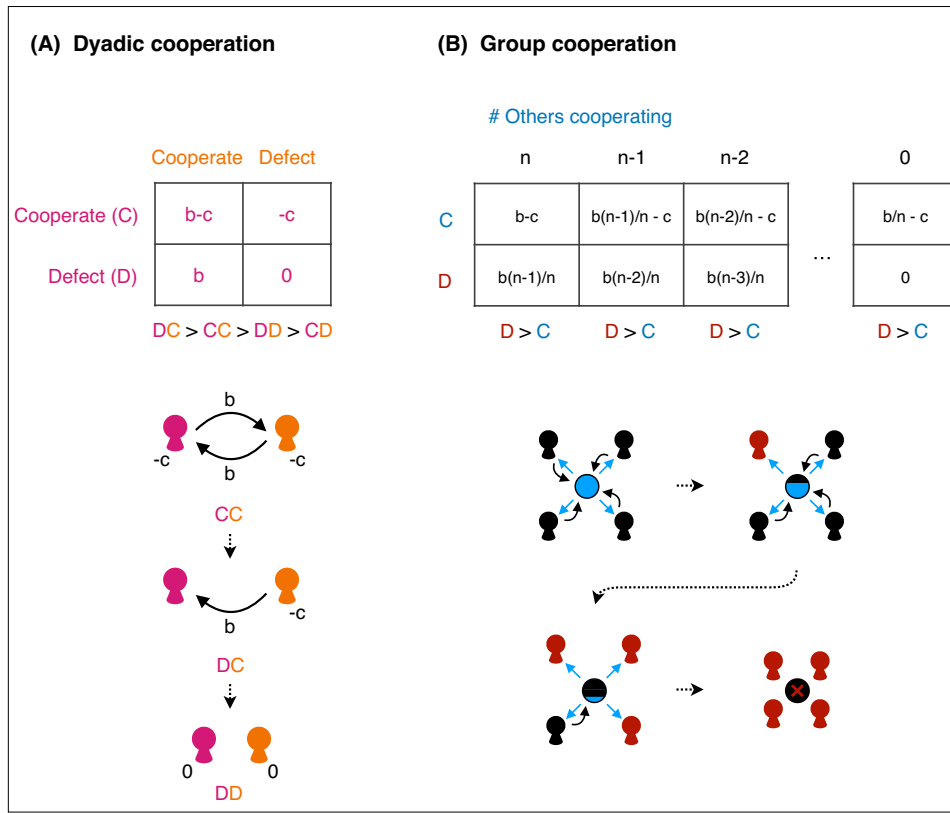
Yet, cooperation based on these mechanisms can also incur significant social costs that are often overlooked in the literature, such as increasing inequality and perpetuating social divisions.

Group cooperation also enforces sharp group boundaries and paves the way for intergroup conflict, while ambiguity and free-rider narratives can lead to intragroup polarization and identity-based conflicts.

Mechanisms that evolved to support cooperation are not ‘designed’ to foster intergroup cooperation, which, however, is needed to address some of humanity’s most pressing challenges.

¹Department of Psychology, Social and Economic Psychology, University of Zurich, 8050 Zurich, Switzerland
²www.psychology.uzh.ch/en/areas/sob/sowipsy.html

*Correspondence:
gross@psychologie.uzh.ch (J. Gross).



Trends in Cognitive Sciences

Figure 1. The puzzle of cooperation. The challenge of maintaining cooperation between two agents is captured by the Prisoner's Dilemma Game (A). In this game, two agents independently decide whether to cooperate – by paying a cost c to create a benefit b for the other agent (with $b > c > 0$) – or to defect, paying no cost and creating no benefit. In the case of mutual cooperation (upper-left cell), an agent receives a benefit b from the other agent's cooperation minus the cost of their own cooperation ($b-c$). However, if the other agent is cooperating, it is advantageous for the agent to defect (bottom-left cell), allowing them to enjoy the benefit b without incurring the cost of reciprocating cooperation (also referred to as free-riding). Conversely, if the other agent defects, it is still better to defect, as one-sided cooperation results in a net cost ($-c$). In group settings, the puzzle of cooperation is captured by the Public Goods Game (B). Here, cooperation creates a marginal benefit b/n for each member of the group while imposing a cost c on the cooperator (with $b > c > b/n > 0$). Regardless of how many members in the group choose to cooperate (table columns), it is always individually advantageous to defect rather than to cooperate, because the marginal individual benefit of cooperation is less than the cost of cooperating. As in the Prisoner's Dilemma, mutual cooperation generates the highest collective benefit, but the temptation to defect and exploit the cooperation of others can quickly undermine cooperation (bottom). In fact, both games, despite having different labels, capture the same cooperation dilemma that arises from the underlying payoff structure, in which the marginal personal return from cooperation is lower than the personal cost. This holds regardless of the number of beneficiaries of cooperation.

explain the emergence of cooperation at different levels of social organization (Figure 2; see also [3,4,7,9,18]) and illustrate that the same mechanisms that facilitate cooperation can also give rise to extortion, discrimination, and group divisions, which in turn can amplify polarization and conflict. We further discuss how group cooperation can impose costs on third parties, thereby creating more harm than benefits. Finally, we demonstrate how small-scale cooperation can undermine large-scale cooperation, jeopardizing global efforts to solve some of humanity's most pressing problems, such as climate change. Our aim is to show that many of the negative side effects of cooperation emerge precisely because cooperation poses a difficult social dilemma that requires (imperfect) mechanisms to sustain it. Nonetheless, not all cooperative mechanisms are created equal with respect to such side effects, which explains why human societies developed institutions that curtail certain forms of cooperation while promoting others.

Glossary

Cooperation: a costly, deliberate action to create a benefit for another individual or group.

Direct reciprocity: form of cooperation that relies on repeated interactions between the same individuals where one individual cooperates as long as cooperation is reciprocated.

Exit costs: the material or psychological losses incurred when losing group membership.

Free-rider narratives: stories about individuals or groups who allegedly take advantage of a public good without contributing enough, often invoking fairness concerns to justify responses such as punishment, exclusion, or social control.

Gossip: sharing information (such as trustworthiness or cooperativeness) about others, typically not present.

Group markers: observable traits or behaviors – such as language, clothing, symbols, or rituals – that signal membership to a particular group and help distinguish members from nonmembers.

Indirect fitness: reproductive success gained by helping relatives pass on shared genes.

Indirect reciprocity: form of cooperation where people can gain a positive reputation from helping others and condition their actions on the reputation of others, rather than expecting a direct return from the recipient.

Negative externalities: unintended costs affecting other people, groups, or entities that are not directly involved in the actions that led to these costs.

Opportunity costs: the trade-off that arises when limited cooperative capacity necessitates favoring one party over another.

Ostracism: excluding someone from the group. In the context of cooperation, ostracism is often modeled as a coordinated group action to exclude someone from the benefits of group goods.

Partner choice: the ability to abandon relationships and switch or search for partners to interact with.

Public goods cooperation: costly actions aimed at contributing to a shared resource. 'Pure' public goods are nonexcludable; no one can be omitted from their benefits. Many public goods do not satisfy this assumption (e.g., golf

Indirect fitness: 'I jump into the river to save two brothers or eight cousins'

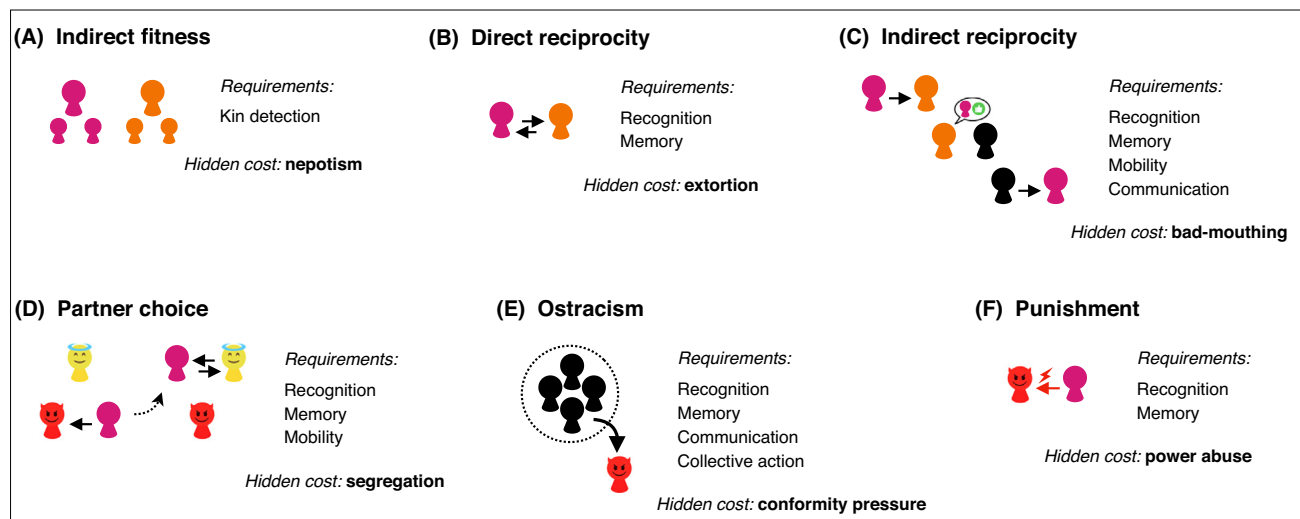
One of the earliest concepts explaining cooperation is **indirect fitness** [19] – the idea that evolutionary success is a function of not solely the number of offspring an individual produces but also the reproductive success of genetically related individuals (Figure 2A). According to Hamilton's rule, cooperation can emerge when the benefit b it produces for another individual, weighted by the genetic relatedness r to this individual (the indirect fitness benefit for helping), outweighs the cost c of a cooperative act: $rb > c$. This also implies that, all else being equal, animals, including humans, should be willing to incur greater costs to help relatives than nonrelatives [20].

Hamilton's rule provides a biological explanation for why parents are willing to sacrifice considerable amounts of time and energy to support their children or why family bands and dynasties play such an important role in hunter-gatherer societies and across human history [21–23]. Yet, kin-based cooperation can entail kin-based favoritism. For example, from the recorded 200 000 living organ donations in the USA between 1988 and 2024, over 117 000 were from genetically related donorsⁱⁱ. Although this kin altruism aids many patients, it disadvantages those without eligible or willing family donors, prolonging their wait for a transplant. In a different domain, the recent US college admissions scandal exposed parents bribing exam administrators to secure unfair advantages for their children in university entrance exams [24]. More broadly, a stronger inclination to help relatives can explain the cultural practice of passing wealth to children, perpetuating inequality across generations [25], as well as nepotism [26], the practice of favoring relatives for influential positions in business or politics.

It is important to note that, under Hamilton's rule, an action is classified as altruistic only if it results in a net loss in direct fitness. Practices such as nepotism may also directly benefit the actor. For example, autocrats, who often face internal threats to their power, do well by appointing

clubs that require membership) and should be referred to as 'club goods.' Nevertheless, we use the term 'public good' because it is often used even when the excludability criterion does not strictly hold.

Punishment: costly action to reduce the welfare of another individual. Also referred to as 'peer punishment' when enacted by one individual toward another.



Trends In Cognitive Sciences

Figure 2. Mechanisms for sustaining cooperation and their drawbacks. Indirect fitness theory (A) proposes that costly, cooperative acts can indirectly benefit an actor's fitness if they are directed toward genetically related individuals. This provides an explanation for nepotism, the tendency to privilege close kin over nonkin. Direct reciprocity (B) explains the emergence of dyadic cooperation over time among nonkin. Recent theories, however, also suggest that close dyadic relationships can give rise to extortionist strategies. Indirect reciprocity (C) considers the indirect benefits of having a reputation as a cooperator. However, reputation systems can be unreliable and susceptible to manipulation (e.g., through bad-mouthing). Partner choice (D) considers the possibility of leaving uncooperative partners in search of more cooperative ones, enabling endogenous sorting. Yet, if cooperation abilities are unevenly distributed, partner choice can lead to segregation and perpetuate inequalities. Ostracism (E) enables groups to exclude uncooperative members. The fear of exclusion can create strong conformity pressures, forcing individuals to align their behavior with sometimes excessively strict group norms. Last, direct punishment by peers (F) can result in wrongful harm, take on disproportionate severity, and lead to power abuse.

trustworthy acolytes [27]. Assuming relatives are more trustworthy because of kin relations, their loyalty can reduce the risk of betrayal.

Direct reciprocity: 'I (sometimes) scratch your back if you scratch mine'

Although genetic relatedness plays a role, human cooperation is often considered special because it extends far beyond kinship, encompassing close bonds with unrelated individuals. Such dyadic relationships can be explained by **direct reciprocity** theory [28] (Figure 2B), according to which cooperation can evolve if the benefit from mutual cooperation b , weighted by the expectation of future interactions p , exceeds the cost of cooperation: $pb > c$. Simply, when the likelihood of meeting again is high enough, it is more beneficial to establish a mutually cooperative relationship than to exploit a cooperator and forgo the long-term benefits. Classic work in evolutionary game theory has shown how strategies that cooperate in the first encounter but defect if the partner exploited them in a previous encounter can sustain cooperation, as long as the long-term expected payoff of cooperation is higher than the one-time payoff for exploitation [7,29,30].

However, an influential paper demonstrated the existence of a class of extortionist strategies that can systematically exploit cooperative players [31]. These strategies force cooperators into receiving a lower (or equal) payoff by reciprocating cooperation only imperfectly, occasionally exploiting the partner's cooperation while resuming cooperation if the other player cooperates [32,33]. In such 'abusive' relationships, continuing to cooperate can be the cooperator's best option, even when being taken advantage of.

Indirect reciprocity: nice guys finish first

To avoid extortionist players or outright defectors, people may exchange information about each other through gossip. **Gossip** can provide an 'early warning system' to distinguish noncooperators from cooperators without having to rely on direct experiences. This can create endogenous pressure to be seen as cooperative because acting prosocially creates a good reputation in the group and may be rewarded by prompting cooperation of others (viz., **indirect reciprocity**; Figure 2C) [34–37]. Indeed, humans gossip frequently, with estimates suggesting that 65% to 90% of day-to-day conversations include gossip [38,39].

Although gossip can facilitate cooperation [40,41], many societies condemn it [42]. This could be because in the competition for a good reputation, having a 'better' reputation than someone else can be sufficient [43], creating incentives to systematically damage the reputation of others through bad-mouthing [44–47]. Because deceptive gossip is often not easy to identify, it may be left uncorrected, leading to bullying, harassment, or even violence. Witch hunts, which have occurred across different cultures and periods [48], were driven by bad-mouthing and false accusations, resulting in thousands of wrongful deaths and widespread fear. A similar pattern emerged in the USA during the 1950s, when fear of communism, stoked by Senator Joseph McCarthy, sparked a wave of baseless accusations. Much like historical witch hunts, these allegations lacked evidence but created a climate of fear, forcing people to prove their loyalty and cooperation to avoid suspicion and escape shunning.

Partner choice: searching for the right partner

Indirect reciprocity enables one to preemptively avoid exploitation through reputation. **Partner choice**, another core mechanism supporting the emergence of cooperation (Figure 2D) [49–51], works retroactively. When encountering a defector, free partner choice enables individuals to exit the relationship and seek a new, more cooperative partner. This facilitates endogenous sorting, where cooperators eventually find one another. At the same time, free partner choice puts pressure on defectors to adopt cooperative strategies, because mutual defection

is inferior to mutual cooperation. Theoretical models and experimental research have demonstrated how partner choice can effectively address the free-rider problem and promote cooperation [49,52–58].

However, these models often assume that it is within an individual's capacity to choose whether to cooperate or not. But what if it is not always possible for an individual to spend resources on cooperation? What if people simply cannot reciprocate cooperation to the same extent as others even if they would like to? If people differ not in their willingness but in their ability to cooperate, partner choice can become a potent driver of social segregation and stratification [59–63]. Individuals with greater resources or skills – qualities that make them more attractive cooperation partners – may cluster together, forming exclusive networks of highly 'able' individuals. Meanwhile, those with fewer resources or lower abilities are excluded. In humans, inequality in physical capabilities, wealth, or education can introduce systematic differences in the ability to cooperate. Under such structural differences, partner choice does not lead to segregation along defectors and cooperators. Rather, it reinforces social divides and inequality [59].

Ostracism: getting thrown out of the group

A coordinated approach to isolating free-riders is to exclude them from the group (Figure 2E). The ability to exclude members from the benefits of group cooperation can successfully deter free-riding [64,65]. Anthropologists have documented both subtle and overt forms of exclusion, ranging from being ignored by group members and excluded from social gatherings to permanent expulsion. In hunter-gatherer societies, the consequences of **ostracism** can be devastating, because it not only comes with the psychological distress of social isolation but also threatens subsistence [66,67].

As with gossip or partner choice, the fear of exclusion can act as a powerful mechanism to foster group cooperation and uphold certain norms [67]. However, it can also lead to excessive social coercion, pressuring individuals to conform to strict rules, which may not always be within their control. Jean Briggs, in her fieldwork with Inuit communities [68], describes a vivid example of this dynamic. Niqi, an Inuit woman, was often perceived by the group as unusually emotional and volatile. Briggs suspected that Niqi was mentally handicapped, unable to fully control her temper to meet the expectations of the community, which led to repeated shunning and exclusion of her and her family. Because cooperation can be fragile, even subtle signs of deviance can provoke excessive social control to maintain group harmony. Similarly, close-knit groups such as cults may create high **exit costs** for their members by requiring them to cut social ties to outsiders [69]. As a result, the fear of ostracism can exert significant conformity pressure by threatening social isolation.

For these reasons, psychologists have frequently highlighted the psychological toll ostracism can impose [70,71] and advocate for restricting such practices, particularly with the rise of social media, which enables gossip and ostracism on a global scale [72,73].

Punishment: keeping order at a cost

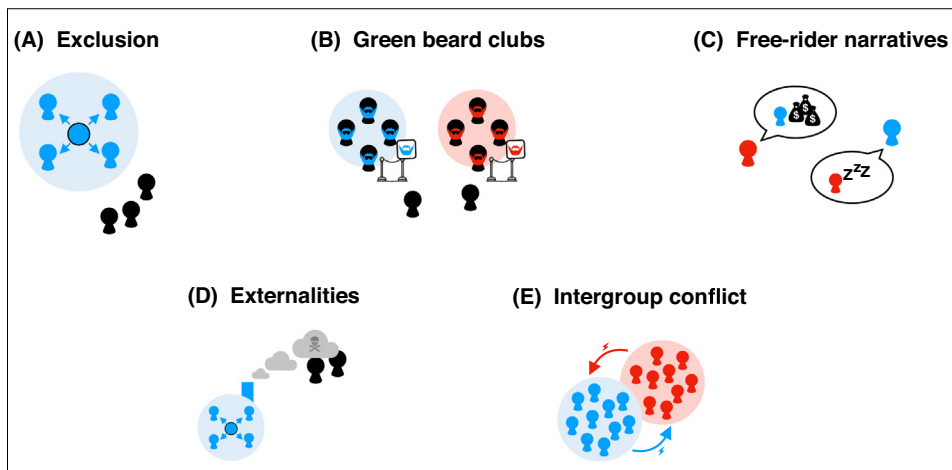
A milder form of sustaining cooperation is to punish rather than completely exclude others (Figure 2F). Research has documented that people are willing to impose costs on themselves to punish others in order to deter free-riding [74–76]. Unfortunately, **punishment** can easily be misused as a tool to control and oppress others [77–79]. Antisocial punishment, the act of punishing cooperators, is observed in some societies, indicating that many different motivations for punishment exist outside of deterring free-riding [80]. Moreover, unregulated peer punishment can trigger retaliation or escalate into punishment vendettas [81,82], directly undermining the benefits of cooperation [83].

Punishment can be particularly effective if publicly visible but also excessively harsh. Cultural practices of public humiliation or shaming have often been used to punish alleged perpetrators and deter others [84]. For example, during the Edo period in Japan, thieves were marked with tattoos, creating a visible, unremovable stigma. During the First and Second World Wars, women in the UK distributed white feathers in public to men not wearing military uniform, aiming to expose cowardice and free-riding to pressure them into enlisting for the war. Often this public punishment was misdirected at veterans or soldiers not wearing their uniforms.

Theoretical and experimental work allows us to cleanly disentangle punishment from gossip or ostracism. Outside of the laboratory, distinctions between being the target of gossip, getting excluded from the group, or getting verbally or physically attacked are often blurry [66]. Yet, what these mechanisms have in common is that they exert social control that endogenously emerges from within groups and can take a life of its own [84].

Group cooperation: who is part of the club?

By mastering the above-mentioned mechanisms, cooperation, with all its benefits and downsides, permeates every aspect of life, from families and friendships to extended social networks and large groups. In groups, humans cooperate not to provide a direct benefit to another person but to create shared resources, referred to as a **public good** (although, from a theoretical perspective, it is irrelevant if a benefit is created for one person, a group of people, or, e.g., a charity organization, as long as the personal cost is greater than the personal benefit; Figure 1). The creation of large-scale public goods can be regarded as a foundation of modern societies. Healthcare and retirement systems, public education, and infrastructure are all provided through group efforts that rely on individually costly contributions, such as taxes. To sustain these public goods, groups must ensure that members sufficiently cooperate within the group and minimize free-riding (e.g., tax evasion) to avoid cooperation breakdown.



Trends in Cognitive Sciences

Figure 3. Negative consequences of group cooperation. Group cooperation requires defining who benefits from shared public goods and who is excluded, necessitating the establishment of clear group boundaries (A). To distinguish between members and nonmembers, groups may develop cultural markers and group tags (similar to 'green beards'; B). Within groups, differing fairness standards and ambiguity surrounding actions and intentions can lead to in-group conflicts and the emergence of free-rider narratives (C), fueling intragroup divisions and identity conflicts. Although group cooperation is beneficial for the group itself, it may impose negative externalities on the environment or out-group members (D). Finally, within-group cooperation can sometimes be directly aimed at fighting, hurting, or oppressing other individuals or groups (E).

An important but often overlooked implication of **public goods cooperation** is that groups must define who belongs to the group and has access to public goods and who does not (Figure 3A). The very nature of group cooperation necessitates the creation or reinforcement of sharp group boundaries, possibly explaining why humans organize themselves into social groups in the first place. Importantly, such group boundaries can lay the foundation for intergroup conflict. Groups may envy the public goods of other groups, compete for ownership of public goods, or expand into the territory of neighboring groups [13,85–88]. Moreover, groups may invest heavily in ‘securing’ their group borders through border patrols (also observed in chimpanzees [89]), the construction of physical barriers, or developing group-targeted paranoia [90]. Concerns for one’s own group can systematically blunt social concerns for outsiders [91].

To distinguish in-group members from out-group members, groups also develop rituals or markers that uphold in-group cohesion while allowing them to identify and exclude outsiders (Figure 3B) [92–95]. **Group markers** are akin to the ‘green beards’ theory, proposing that cooperation can be maintained among individuals that share a gene determining their cooperative trait (acting altruistically) together with a perceptible trait (having a ‘green beard’). Although green beard dynamics are widely regarded as implausible for human cooperation on the genetic level [18,96,97], humans may employ similar social discrimination mechanisms on a cultural level through group markers.

For group markers to be effective, attaining them – and thus group membership – must involve costs. Without such costs, individuals could freely enter or leave groups, which would allow defectors to easily infiltrate cooperative groups. Examples of costly group markers are religious affiliations, college fraternities, academic titles, or simply an expensive membership to a golf club. Although group markers can help groups distinguish between cooperators and defectors, they can also result in people being classified as insiders and outsiders. This can pave the way for hierarchy formation, segregation, in-group favoritism, and stereotyping, ultimately leading to the emergence of exclusive ‘clubs’ and the fostering of ‘old boy networks’ that entrench divisions along class or ethnic lines [91,98–100].

Group cooperation can create tensions not only between but also within groups. There is often uncertainty about who should contribute how much to public goods (varying fairness standards) but also whether a lack of cooperation is a signal of free-riding or the inability to cooperate (ambiguity about intentions). This uncertainty can give rise to **free-rider narratives** (Figure 3C). Public discourse is permeated by free-rider narratives, such as whether ‘the wealthy pay their fair share’ or to what extent immigrants or the unemployed are unable to contribute and deserve support or are taking advantage of public goods and ‘exploit the system.’ Free-rider narratives also extend to intergroup relations with claims such as ‘one group is taking advantage of another,’ which can undermine intergroup exchange and fuel populism and scapegoating. Groups that depend more on group cooperation and public goods may be more sensitive to free-riding risks than less interdependent ones [98,101]. This heightened vigilance can make them more susceptible to free-rider narratives and increase identity-based polarization.

Cooperation with externalities

As argued, cooperation and the mechanisms upholding it can lead to a myriad of negative social consequences. Importantly, the very definition of cooperation – creating a benefit for someone else – is often incomplete. In many cases, although cooperation creates benefits for one party, it (unintentionally) imposes external costs (Figure 3D). This is particularly evident in light of climate change and the ongoing Anthropocene extinction event [102]. Humanity’s extraordinary capacity for cooperation, enabling the exploitation of resources and transforming entire ecosystems, has placed immense strain on the environment, threatening the very foundations of human existence. Human cooperation may be a success story until it isn’t.

In fact, hurting another individual or group can also be the explicit aim of cooperation. Military conflicts require cooperation, as does the coordination of terrorist attacks or the building of Mafia organizations (Figure 3E) [99,100,103]. Collusion, corruption, and human trafficking are forms of cooperation or rely on some of the cooperation mechanisms outlined above [104–110]. An example of this was Silk Road, a black-market platform that enabled the trading of drugs and weapons. Despite the lack of formal enforcement to punish cheating, trust between sellers and buyers was maintained through a reputation-rating system and an online forum where users shared gossip about the trustworthiness of market participants [15,16]. This demonstrates how the same mechanisms that sustain benign forms of cooperation can also enable such an illegal marketplace.

The mechanisms essential for maintaining cooperation also reveal their limitations in scope and scale. Whether acts of cooperation are perceived as altruism or corruption is often a matter of perspective [14,104,111–113]¹. Many forms of corruption can be seen as cooperation turned inward, benefiting a few while burdening the many. Moreover, the same mechanisms support both corruption and cooperation: indirect fitness explains kin altruism and nepotism; direct reciprocity supports mutual aid and cronyism; and indirect reciprocity accounts for group cooperation but also for failures to solve global challenges through intergroup cooperation.

In the face of problems such as climate change, cooperation that transcends group boundaries is required [114] – something the above-mentioned mechanisms arguably did not ‘evolve for’ [115–117]. How, then, can we promote global cooperation while limiting cooperation on smaller scales? Encouragingly, universal cooperation is possible under certain conditions [114,118] (Box 1). Yet, there is always the threat of local groups eroding global cooperation, because cooperation at the global scale can be conceived as defection at the local scale [115–117]. This is also exemplified by the observation that whistleblowers, exposing fraudulent activities in organizations or governments, often face ostracism and shunning, similar to free-riders, even if their actions ultimately benefit society. It may be one of humanity’s greatest tests to determine how to build large-scale, intergroup cooperation capable of addressing current global challenges.

When and why do the benefits of cooperation outweigh its costs?

To evaluate the hidden costs of cooperation, we must ask: Are these unavoidable? Do they reflect aspects of our evolved psychology, or are they direct products of social institutions? How should we weigh them against the benefits of cooperation? To this end, it is helpful to distinguish between distinct types of negative consequences.

First, cooperation can produce **negative externalities**, whether voluntary or unintended. Yet, individuals, too, can harm others, deliberately or inadvertently. What cooperation adds is a scale to these consequences that exceeds what individuals are capable of. Although some have suggested that the capacity for cooperation evolved under conditions of intergroup conflict [13,17,119–121], we see little empirical evidence that cooperation depends on competition as a psychological motivator [122]. Negative externalities associated with cooperation are not inevitable; conflict and harm are not necessary for cooperation to function.

Second, incurring the costs of cooperation for one party always means withholding cooperation from another (**opportunity costs**). Both kin altruism and tag-based cooperation, likely evolutionarily prepared and mediated by selective empathy, bias prosocial behavior toward keeping the costs and benefits of cooperation within the group or kin, a tendency further reinforced by the more frequent opportunities for direct and indirect reciprocity with close rather than distant others [123]. Cooperation, to some extent, gravitates toward such small scales with nepotism and corruption as naturally emerging manifestations.

Box 1. Emergence of intergroup cooperation

Group cooperation is often maintained and reinforced by strong cooperative bonds between individual group members (Figure 1, top; [114–118]). Because of physical, cultural, or psychological closeness, such dyadic relationships are more likely to develop within groups than between them. However, throughout human history, members of distinct groups have also often developed mutually beneficial relationships (e.g., through trade), even across large geographical distances, language barriers, and cultural divides (Figure 1, second panel). This can set off a gradual transition in which group members start to interact more frequently with out-group members (Figure 1, third panel). If such intergroup contact is frequent enough, group-transcending public goods can emerge [114]. These dyadic relationships across group boundaries and shared goods can reinforce each other, progressively leading to group fusion, whereby two formerly distinct groups form a shared higher-order identity and institutions. Over the course of human history, group fusion has enabled humans to create increasingly large and complex societies. However, maintaining cooperation within larger societies also poses unique challenges. If these challenges are not addressed, the process depicted in Figure 1 may reverse, triggering fission dynamics that can lead to a collapse into smaller groups again.

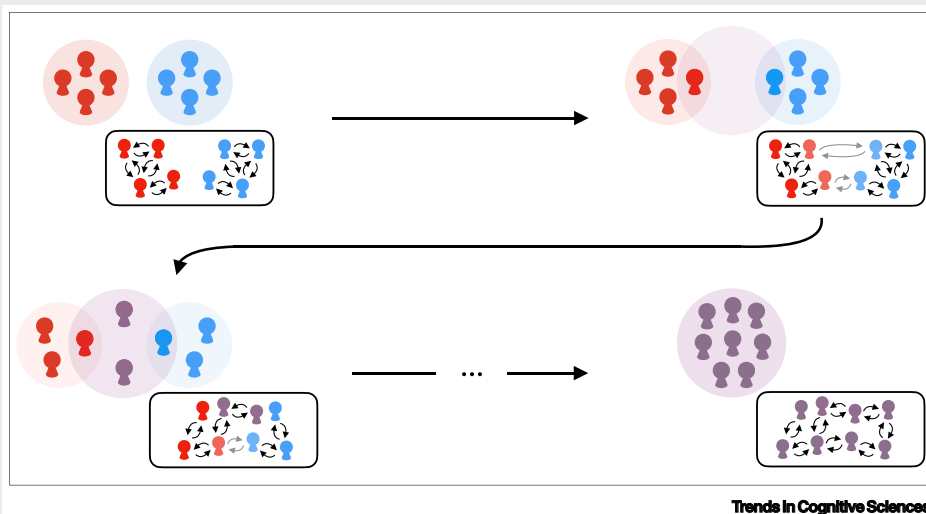


Figure 1. Intergroup cooperation through intergroup exchange.

Third, people can fail to cooperate because of limitations in ability, circumstances, or resources. This creates a fundamental epistemological problem, leading to an enforcement dilemma in distinguishing between actions versus abilities and intentions. Groups can err on the side of caution at the cost of falsely accusing, punishing, or ostracizing noncontributors whose lack of cooperation may not have been in their control [124]. Or they can tolerate noncontributors, thereby inviting genuine free-riders and risking the collapse of cooperation. This uncertainty is, to some degree, irreducible, often fueling ideological conflicts and free-rider narratives. Such narratives may be rooted in an evolved psychology that predisposes people to be suspicious of noncontributors [125]. However, people also exhibit some tolerance and do take intentions and mitigating circumstances into account when judging free-riding, if accessible [126–129]. The crucial point is that this dilemma logically follows from (i) the threat free-riders pose to cooperation and (ii) the need to limit temptations to free-ride under (iii) imperfect information about underlying abilities or intentions.

Fourth, even if someone willfully free-rides, disciplinary measures can be overly harsh and counterproductive. A single lapse can provoke disproportionate stigma or retaliation. For that reason, punishment has often been calibrated by societies, moving from an eye-for-an-eye approach to allowing forgiveness, punishing repeat offenders more than one-time ‘slips,’ and reducing social

pressure and the psychological toll of punishment. Although it is difficult to quantify when punitive costs outweigh their benefits, at least some of these negative consequences can be mitigated.

Over the last centuries, some societies have developed a tacit understanding of the hidden costs of cooperation, giving rise to institutions promoting certain forms of cooperation while curtailing others. For example, policies restrict collusion, combat corruption, tax inherited wealth, or encourage organ donation among nonkin to counteract kin favoritism. Principles of impartiality, formulated as the moral duty to value the well-being of all individuals equally regardless of kinship or group affiliation, provide a cultural counterweight to kin- or group-based favoritism. Policies are also increasingly put in place to price environmental damage or social harm. Public education and social housing initiatives aim to reduce the hidden costs of unrestricted partner choice and inequalities in the ability to cooperate, leveling the playing field between advantaged and disadvantaged members of society and curbing segregation and inequality. Modern legal systems emphasize proof of intentions over consequences of actions when determining liability, highlighting efforts to mitigate detection errors. Frequent partner rotation is sometimes employed as a measure to prevent the emergence of corrupt collaborations [105], and the balance of power in modern democracies is designed to foster cooperation that benefits all members of society. Many of these advances are relatively recent in human history.

One guiding principle for weighing the downsides against the benefits of mechanisms that support cooperation is to consider the scale of cooperation they enable while reducing the downsides listed above. Modern institutions, which are based on norms that promote impartiality, highlight common identity, distinguish between action and ability, and differentiate between luck and merit-based inequality, seem to be the most promising arrangement for promoting large-scale, inclusive cooperation. Although costly to maintain and difficult to transition to, they may provide leveraging effects [130] that compensate for ‘cheaper’ mechanisms such as free partner choice, unregulated gossip, or peer punishment. Although all-encompassing cooperation may be utopian, given the risks of free-riding and the costs of maintaining complex institutions, the above examples aim to illustrate that scaling up cooperation toward larger, more inclusive groups is both possible and desirable (see also Box 1). At the same time, complex institutions pose a second-order free-rider problem and can be highly fragile, requiring constant checks and balances to counteract the erosive forces of kin or group favoritism. From this perspective, the main threat to large-scale cooperation is not primarily free-riding but the constant danger of parochial forms of cooperation or the polarizing force of free-rider narratives that lead to group fission and the fragmentation of large-scale into small-scale cooperation.

Concluding remarks

Compared with other animals, our physical abilities are mediocre at best. We lost the climbing ability of our closest relatives, we are not particularly fast (though enduring) runners, we lack claws, our canine teeth are small, and we rely on cooked food [131]. Rather, what made humans so successful is our ability to cooperate, enabling human groups to become more than the sum of their parts.

Yet, cooperation is a mixed blessing. Many of the mechanisms at humanity’s disposal for maintaining cooperation rely on social control, which can sometimes lead to excessive pressures of conformity and social cruelties. Moreover, selective cooperation – such as kin favoritism or partner choice – can systematically increase inequality and foster social segregation. Free-rider narratives can further polarize societies, fueling internal conflicts based on identity politics.

Outstanding questions

What factors increase the risk that cooperation amplifies inequality, segregation, unfairness, polarization, or negative externalities?

Under what conditions does enforcing norms through punishment, exclusion, or gossip promote cooperation without undermining trust, autonomy, or social cohesion?

How can we find a sustainable balance between relational mobility and individual liberties, considering their potential negative roles in segregation and the perpetuation of inequality?

How can groups develop sustained intergroup cooperation and overcome parochialism?

How can group cooperation be maintained without relying on group markers, high entry or exit costs, identity polarization, or intergroup conflict?

How can institutions be designed to resist kin favoritism and corruption?

Is there a relationship between education and economic development and the mechanisms and scope of cooperation that groups develop?

What social norms and values support the development of universalist institutions that prioritize impartiality, equality, and large-scale cooperation?

How can societies address the ‘dual-use’ problem in cooperation? How can they promote cooperation while minimizing its risks of exploitation for harmful purposes?

How should societies weigh obligations to family and close others against the larger societal benefits of global cooperation?

What strategies help expand group boundaries to include more socially, culturally, or geographically distant others?

For social scientists, the challenge lies in identifying mechanisms that not only support cooperation but also make it sustainable (see [Outstanding questions](#)). Instead of simply maximizing cooperation or minimizing free-riding, the issue of cooperation becomes a complex, multiobjective optimization problem that also entails reducing externalities, intergroup tensions, inequality, and unfairness, as well as limiting excessive social control and maintaining individual liberties.

Cooperation may no longer be much of a puzzle. The puzzle is rather how to foster sustainable, inclusive cooperation that leverages the unique capabilities of humanity without creating losers and laying the foundation for inequality, polarization, conflict, and environmental destruction. On the bright side, humans are capable of shaping their environment and defining the rules of the games they play. Creating institutions that foster people's virtues while curbing their vices offers humanity its best chance at addressing current challenges.

Acknowledgments

This work has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI) to J.G. We thank the Social and Economic Psychology group at the University of Zurich for their valuable comments and feedback.

Declaration of interests

The authors have no interests to declare.

Resources

ⁱwww.theguardian.com/books/2022/oct/24/the-big-idea-is-cooperation-always-a-force-for-good

ⁱⁱ<https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/>

References

- Melis, A.P. and Semmann, D. (2010) How is human cooperation different? *Philos. Trans. R. Soc. B Biol. Sci.* 365, 2663–2674
- Boyd, R. and Richerson, P.J. (2009) Culture and the evolution of human cooperation. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 3281–3288
- Henrich, J. and Muthukrishna, M. (2020) The origins and psychology of human cooperation. *Annu. Rev. Psychol.* 72, 1–34
- Rand, D.G. and Nowak, M.A. (2013) Human cooperation. *Trends Cogn. Sci.* 17, 413–425
- Fischbacher, U. and Gächter, S. (2010) Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *Am. Econ. Rev.* 100, 541–556
- Pennisi, E. (2005) How did cooperative behavior evolve? *Science* 309, 93
- Nowak, M.A. (2006) Five rules for the evolution of cooperation. *Science* 314, 1560–1563
- West, S.A. *et al.* (2007) Evolutionary explanations for cooperation. *Curr. Biol.* 17, R661–R672
- Tomasello, M. and Vaish, A. (2013) Origins of human cooperation and morality. *Annu. Rev. Psychol.* 64, 231–255
- Fehr, E. and Schurtenberger, I. (2018) Normative foundations of human cooperation. *Nat. Hum. Behav.* 2, 458–468
- Dietz, T. *et al.* (2003) The struggle to govern the commons. *Science* 302, 1907–1912
- Jin, S. *et al.* (2025) Institutions and cooperation: a meta-analysis of structural features in social dilemmas. *J. Pers. Soc. Psychol.* 129, 286–312
- Choi, J.-K. and Bowles, S. (2007) The coevolution of parochial altruism and war. *Science* 318, 636–640
- Muthukrishna, M. *et al.* (2017) Corrupting cooperation and how anti-corruption strategies may backfire. *Nat. Hum. Behav.* 1, 0138
- Przepiorka, W. *et al.* (2017) Order without law: reputation promotes cooperation in a cryptomarket for illegal drugs. *Eur. Sociol. Rev.* 33, 752–764
- Macanovic, A. and Przepiorka, W. (2023) The moral embeddedness of cryptomarkets: text mining feedback on economic exchanges on the dark web. *Socio Econ. Rev.* 22, 1705–1732
- Bowles, S. (2008) Being human: conflict: altruism's midwife. *Nature* 456, 326–327
- Henrich, J. (2004) Cultural group selection, coevolutionary processes and large-scale cooperation. *J. Econ. Behav. Organ.* 53, 3–35
- Wild, G. (2023) Pillars of biology: 'the genetical evolution of social behaviour, I and II'. *J. Theor. Biol.* 562, 111430
- Stewart-Williams, S. (2007) Altruism among kin vs. nonkin: effects of cost of help and reciprocal exchange. *Evol. Hum. Behav.* 28, 193–198
- Walker, R.S. and Hill, K.R. (2014) Causes, consequences, and kin bias of human group fissions. *Hum. Nat.* 25, 465–475
- Bó, E.D. *et al.* (2009) Political dynasties. *Rev. Econ. Stud.* 76, 115–142
- Duindam, J. (2016) *Dynasties: A Global History of Power, 1300–1800*, Cambridge University Press
- Korn, M. and Levitz, J. (2020) *Unacceptable: Privilege, Deceit & the Making of the College Admissions Scandal*, Penguin
- Piketty, T. (2011) On the long-run evolution of inheritance: France 1820–2050. *Q. J. Econ.* 126, 1071–1131
- Pearce, J.L. (2015) Cronyism and nepotism are bad for everyone: the research evidence. *Ind. Organ. Psychol.* 8, 41–44
- Brownlee, J. (2007) Hereditary succession in modern autocracies. *World Polit.* 59, 595–628
- Trivers, R.L. (1971) The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57
- Rossetti, C.S.L. and Hilbe, C. (2024) Direct reciprocity among humans. *Ethology* 130, e13407
- Axelrod, R. and Hamilton, W.D. (1981) The evolution of cooperation. *Science* 211, 1390–1396
- Press, W.H. and Dyson, F.J. (2012) Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. *Proc. Natl. Acad. Sci.* 109, 10409–10413
- Hilbe, C. *et al.* (2014) Extortion subdues human players but is finally punished in the prisoner's dilemma. *Nat. Commun.* 5, 3976

33. Milinski, M. (2022) Extortion – a voracious prosocial strategy. *Curr. Opin. Psychol.* 44, 196–201
34. Roberts, G. *et al.* (2021) The benefits of being seen to help others: indirect reciprocity and reputation-based partner choice. *Philos. Trans. R. Soc. B* 376, 20200290
35. Ohtsuki, H. and Iwasa, Y. (2006) The leading eight: social norms that can maintain cooperation by indirect reciprocity. *J. Theor. Biol.* 239, 435–444
36. Kawakatsu, M. *et al.* (2024) A mechanistic model of gossip, reputations, and cooperation. *Proc. Natl. Acad. Sci. U. S. A.* 121, e2400689121
37. Wu, J. *et al.* (2016) Reputation, gossip, and human cooperation. *Soc. Personal. Psychol. Compass* 10, 350–364
38. Beersma, B. and Kleef, G.A.V. (2012) Why people gossip: an empirical analysis of social motives, antecedents, and consequences. *J. Appl. Soc. Psychol.* 42, 2640–2670
39. Dunbar, R.I.M. *et al.* (1997) Human conversational behavior. *Hum. Nat.* 8, 231–246
40. Nieper, A.S. *et al.* (2022) When and why does gossip increase prosocial behavior? *Curr. Opin. Psychol.* 44, 315–320
41. Pan, X. *et al.* (2024) Explaining the evolution of gossip. *Proc. Natl. Acad. Sci.* 121, e2214160121
42. Foster, E.K. (2003) Research on gossip: taxonomy, methods, and future directions. *Rev. Gen. Psychol.* 8, 78–99
43. Antonioni, A. *et al.* (2016) Cooperation survives and cheating pays in a dynamic network structure with unreliable reputation. *Sci. Rep.* 6, 27160
44. Hess, N.H. and Hagen, E.H. (2006) Psychological adaptations for assessing gossip veracity. *Hum. Nat.* 17, 337–354
45. Peters, K. and Fonseca, M.A. (2020) Truth, lies, and gossip. *Psychol. Sci.* 31, 702–714
46. Wu, J. *et al.* (2021) Honesty and dishonesty in gossip strategies: a fitness interdependence analysis. *Philos. Trans. R. Soc. B* 376, 20200300
47. Cruz, T.D.D. *et al.* (2019) The bright and dark side of gossip for cooperation in groups. *Front. Psychol.* 10, 1374
48. Levack, B.P. (2013) *The Witch-Hunt in Early Modern Europe*, Pearson Education
49. Barclay, P. (2016) Biological markets and the effects of partner choice on cooperation and friendship. *Curr. Opin. Psychol.* 7, 33–38
50. Noë, R. and Hammerstein, P. (1995) Biological markets. *Trends Ecol. Evol.* 10, 336–339
51. Barclay, P. (2013) Strategies for cooperation in biological markets, especially for humans. *Evol. Hum. Behav.* 34, 164–175
52. Barclay, P. and Willer, R. (2007) Partner choice creates competitive altruism in humans. *Proc. R. Soc. B Biol. Sci.* 274, 749–753
53. Brekke, K.A. *et al.* (2011) Playing with the good guys. A public good game with endogenous group formation. *J. Public Econ.* 95, 1111–1118
54. Hardy, C.L. and Vugt, M.V. (2006) Nice guys finish first: the competitive altruism hypothesis. *Personal. Soc. Psychol. Bull.* 32, 1402–1413
55. Roberts, G. (1998) Competitive altruism: from reciprocity to the handicap principle. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 265, 427–431
56. Baumard, N. *et al.* (2013) A mutualistic approach to morality: the evolution of fairness by partner choice. *Behav. Brain Sci.* 36, 59–78
57. Efferson, C. *et al.* (2016) Sustained cooperation by running away from bad behavior. *Evol. Hum. Behav.* 37, 1–9
58. Helbing, D. and Yu, W. (2009) The outbreak of cooperation among success-driven individuals under noisy conditions. *Proc. Natl. Acad. Sci.* 106, 3680–3685
59. Stallen, M. *et al.* (2023) Partner choice and cooperation in social dilemmas can increase resource inequality. *Nat. Commun.* 14, 6432
60. Raihani, N.J. and Barclay, P. (2016) Exploring the trade-off between quality and fairness in human partner choice. *R. Soc. Open Sci.* 3, 160510
61. Snijder, L.L. *et al.* (2024) Decision-makers self-servingly navigate the equality-efficiency trade-off of free partner choice in social dilemmas among unequals. *J. Econ. Psychol.* 105, 102758
62. Hauser, O.P. *et al.* (2019) Social dilemmas among unequals. *Nature* 572, 524–527
63. Melamed, D. *et al.* (2022) Inequality and cooperation in social networks. *Sci. Rep.* 12, 6789
64. Feinberg, M. *et al.* (2013) Gossip and ostracism promote cooperation in groups. *Psychol. Sci.* 25, 656–664
65. Hirshleifer, D. and Rasmusen, E. (1989) Cooperation in a repeated prisoners' dilemma with ostracism. *J. Econ. Behav. Organ.* 12, 87–106
66. Boehm, C. (1999) *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*, Harvard University Press
67. Ouwerkerk, J.W. *et al.* (2005) Avoiding the social death penalty – ostracism and cooperation in social dilemmas. In *The Social Outcast*, pp. 321–332, Psychology Press
68. Briggs, J.L. (1971) *Never in Anger: Portrait of an Eskimo Family*, Harvard University Press
69. Lindenberg, S. (1984) Normen und die Allokation sozialer Wertschätzung. In *Schriften des Vereins für Sozialpolitik*, Duncker & Humblot
70. Williams, K.D. (2007) Ostracism. *Psychology* 58, 425–452
71. Chen, Z. *et al.* (2025) Psychosocial and behavioural consequences of ostracism across contexts. *Nat. Rev. Psychol.* 4, 112–126
72. Brady, W.J. *et al.* (2021) How social learning amplifies moral outrage expression in online social networks. *Sci. Adv.* 7, eabe5641
73. Crockett, M.J. (2017) Moral outrage in the digital age. *Nat. Hum. Behav.* 1, 769–771
74. Gross, J. *et al.* (2016) Building the Leviathan – voluntary centralisation of punishment power sustains cooperation in humans. *Sci. Rep.* 6, 20767
75. Egas, M. and Riedl, A. (2008) The economics of altruistic punishment and the maintenance of cooperation. *Proc. Biol. Sci.* 275, 871–878
76. Fehr, E. and Gächter, S. (2002) Altruistic punishment in humans. *Nature* 415, 137–140
77. Wylie, J. and Gantman, A. (2024) Cooperation, domination: twin functions of third-party punishment. *Soc. Pers. Psychol. Compass* 18
78. Raihani, N.J. and Bshary, R. (2019) Punishment: one tool, many uses. *Evol. Hum. Sci.* 1, e12
79. Hoeft, L. *et al.* (2025) Normative acceptance of power abuse. *Kyklos* 78, 1375–1386
80. Herrmann, B. *et al.* (2008) Antisocial punishment across societies. *Science* 319, 1362–1367
81. Engelmann, D. and Nikiforakis, N. (2015) In the long-run we are all dead: on the benefits of peer punishment in rich environments. *Soc. Choice Welf.* 45, 561–577
82. Fehl, K. *et al.* (2012) I dare you to punish me – vendettas in games of cooperation. *PLoS One* 7, e45093
83. Han, T.A. *et al.* (2024) Evolutionary mechanisms that promote cooperation may not promote social welfare. *J. R. Soc. Interface* 21, 20240547
84. Billingham, P. and Parr, T. (2020) Enforcing social norms: the morality of public shaming. *Eur. J. Philos.* 28, 997–1016
85. Snijder, L.L. *et al.* (2024) Prosocial preferences can escalate intergroup conflicts by countering selfish motivations to leave. *Nat. Commun.* 15, 9009
86. Dreu, C.K.W.D. *et al.* (2024) Group formation and the evolution of human social organization. *Perspect. Psychol. Sci.* 19, 320–334
87. Dreu, C.K.W.D. *et al.* (2021) Pro-sociality as a foundation for intergroup conflict. *Curr. Opin. Psychol.* 44, 112–116
88. Dreu, C.K.W.D. *et al.* (2020) Group cooperation, carrying-capacity stress, and intergroup conflict. *Trends Cogn. Sci.* 24, 760–776
89. Massaro, A.P. *et al.* (2022) Correlates of individual participation in boundary patrols by male chimpanzees. *Philos. Trans. R. Soc. B* 377, 20210151
90. Raihani, N.J. and Bell, V. (2019) An evolutionary perspective on paranoia. *Nat. Hum. Behav.* 3, 114–121
91. Dorison, C.A. and Kteily, N.S. (2024) Group-based reputational incentives can blunt sensitivity to societal harms and benefits. *J. Exp. Psychol. Gen.* 153, 2605–2625

92. Cohen, E. (2012) The evolution of tag-based cooperation in humans: the case for accent. *Curr. Anthropol.* 53, 588–616
93. Jansen, V.A.A. and van Baalen, M. (2006) Altruism through beard chromodynamics. *Nature* 440, 663–666
94. Axelrod, R. *et al.* (2004) Altruism via kin-selection strategies that rely on arbitrary tags with which they coevolve. *Evolution* 58, 1833–1838
95. Efferson, C. *et al.* (2008) The coevolution of cultural groups and ingroup favoritism. *Science* 321, 1844–1849
96. West, S.A. and Gardner, A. (2010) Altruism, spite, and greenbeards. *Science* 327, 1341–1344
97. Heath, J. and Rioux, C. (2018) Recent trends in evolutionary ethics: greenbeards! *Biol. Philos.* 33, 16
98. Cavallé, C. and Ferwerda, J. (2023) How distributional conflict over in-kind benefits generates support for far-right parties. *J. Polit.* 85, 19–33
99. Gambetta, D. (1993) *The Sicilian Mafia: The Business of Protection*, Harvard University Press
100. Meier, S. *et al.* (2016) Trust and in-group favoritism in a culture of crime. *J. Econ. Behav. Organ.* 132, 78–92
101. Köhler, L. and Gollwitzer, M. (2024) How victim sensitivity affects our attitudes and behaviour towards immigrants. *Br. J. Psychol.* 115, 406–436
102. Ceballos, G. and Ehrlich, P.R. (2023) Mutilation of the tree of life via mass extinction of animal genera. *Proc. Natl. Acad. Sci.* 120, e2306987120
103. Nese, A. *et al.* (2018) Cooperation, punishment and organized crime: a lab-in-the-field experiment in southern Italy. *Eur. Econ. Rev.* 107, 86–98
104. Dorrough, A.R. *et al.* (2023) Conditional bribery: Insights from incentivized experiments across 18 nations. *Proc. Natl. Acad. Sci.* 120, e2209731120
105. Gross, J. *et al.* (2018) Ethical free riding: when honest people find dishonest partners. *Psychol. Sci.* 29, 1956–1968
106. Thielmann, I. *et al.* (2021) Buying unethical loyalty: a behavioral paradigm and empirical test. *Soc. Psychol. Personal. Sci.* 12, 363–370
107. Szekely, A. *et al.* (2020) Aggression, conflict, and the formation of intimidating group reputation. *Soc. Psychol. Q.* 83, 70–87
108. Weisel, O. and Shalvi, S. (2015) The collaborative roots of corruption. *Proc. Natl. Acad. Sci.* 112, 10651–10656
109. Raihani, N. (2021) *The Social Instinct: How Cooperation Shaped the World*, Random House
110. Leib, M. *et al.* (2021) Collaborative dishonesty: a meta-analytic review. *Psychol. Bull.* 147, 1241–1268
111. Muthukrishna, M. (2017) Corruption, cooperation, and the evolution of prosocial institutions. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.3082315>
112. Schram, A. *et al.* (2022) Corruption: a cross-country comparison of contagion and conformism. *J. Econ. Behav. Organ.* 193, 497–518
113. Spadaro, G. *et al.* (2023) Corrupt third parties undermine trust and prosocial behaviour between people. *Nat. Hum. Behav.* 7, 46–54
114. Gross, J. *et al.* (2023) The evolution of universal cooperation. *Sci. Adv.* 9, eadd8289
115. Schnell, E. and Muthukrishna, M. (2024) Indirect reciprocity undermines indirect reciprocity destabilizing large-scale cooperation. *Proc. Natl. Acad. Sci.* 121, e2322072121
116. Gross, J. *et al.* (2024) Indirect reciprocity can foster large-scale cooperation. *Proc. Natl. Acad. Sci.* 121, e2409894121
117. Schnell, E. and Muthukrishna, M. (2024) Reply to Gross *et al.*: Indirect reciprocity undermines large-scale cooperation under realistic conditions. *Proc. Natl. Acad. Sci.* 121, e2410085121
118. Gross, J. *et al.* (2025) Free mobility across group boundaries promotes intergroup cooperation. *Commun. Psychol.* 3, 10
119. Bowles, S. (2009) Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science* 324, 1293–1298
120. Efferson, C. *et al.* (2024) Super-additive cooperation. *Nature* 626, 1034–1041
121. Bauer, M. *et al.* (2016) Can war foster cooperation? *J. Econ. Perspect.* 30, 249–274
122. Yamagishi, T. and Mifune, N. (2016) Parochial altruism: does it explain modern human group psychology? *Curr. Opin. Psychol.* 7, 39–43
123. Yamagishi, T. and Kiyonari, T. (2000) The group as the container of generalized reciprocity. *Soc. Psychol. Q.* 63, 116
124. Cushman, F. *et al.* (2009) Accidental outcomes guide punishment in a 'trembling hand' game. *PLoS ONE* 4, e6699
125. Rosset, E. (2008) It's no accident: our bias for intentional explanations. *Cognition* 108, 771–780
126. Delton, A.W. *et al.* (2012) The psychosemantics of free riding: dissecting the architecture of a moral concept. *J. Pers. Soc. Psychol.* 102, 1252–1270
127. Cushman, F. (2008) Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition* 108, 353–380
128. Cushman, F. *et al.* (2013) The development of intent-based moral judgment. *Cognition* 127, 6–21
129. Alicke, M.D. (2000) Culpable control and the psychology of blame. *Psychol. Bull.* 126, 556–574
130. Lie-Panis, J. *et al.* (2024) The social leverage effect: Institutions transform weak reputation effects into strong incentives for cooperation. *Proc. Natl. Acad. Sci.* 121, e2408802121
131. Henrich, J. (2016) *The Secret of Our Success*, Princeton University Press