

Taking charge and stepping in: Individuals who punish are rewarded with prestige and dominance

Daniel Redhead¹  | Nathan Dhaliwal² | Joey T. Cheng³

¹Department of Human Behaviour, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

²Sauder School of Business, University of British Columbia, Vancouver, Canada

³Department of Psychology, York University, Canada

Correspondence

Daniel Redhead, Department of Human Behaviour, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

Email: daniel_redhead@eva.mpg.de

Joey T. Cheng, Department of Psychology, York University, Canada.

Email: chengjt@york.ca

Funding information

Open access funding enabled and organized by Projekt DEAL.

Abstract

A hallmark of human societies is the scale at which we cooperate with many others, even when they are not closely genetically related to us. One proposed mechanism that helps explain why we cooperate is punishment; cooperation may pay and proliferate if those who free ride on the cooperation of others are punished. Yet this ‘solution’ raises another puzzle of its own: Who will bear the costs of punishing? While the deterrence of free-riders via punishment serves collective interests, presumably any single individual—who has no direct incentive to punish—is better off letting others pay the costs of punishment. However, emerging theory and evidence indicate that, while punishment may at times be a costly act, certain individuals are better able to ‘afford’ to pay the price of punishment and are often consequentially rewarded with fitness-enhancing reputation benefits. Synthesizing across these latest lines of research, we propose a novel framework that considers how high status individuals—that is, individuals with greater prestige or dominance—enjoy lower punishment costs. These individuals are thus more willing to punish, and through their punitive action can in turn reap reputational rewards by further gaining more prestige or dominance. These reputational gains, which

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. Social and Personality Psychology Compass published by John Wiley & Sons Ltd.

work in concert to promote the social success of punishers, recoup the costs of punishing. Together, these lines of work suggest that while punishment is often assumed to be altruistic, it need not always depend on altruism, and motivations to punish may at times be self-interested and driven (whether consciously or unconsciously) by reputational benefits.

1 | INTRODUCTION

As a social species, humans face the problem of balancing their own individual goals and desires with the expectations and motivations of the group in which they live. Groups of two or more individuals best function together when they are able to effectively work together, coordinating their actions and resolving conflicts between individual motivations and those of the collective good (often termed *collective action problems*; Dugatkin, 1997; Smith, 2003). However, in such group settings, individuals may act in self-interest (i.e., become defectors or free-riders) and reap the benefits of the cooperative endeavours of other group members without contributing to the collective effort (Hardin, 1982; Hooper, Kaplan, & Boone, 2010; Olson, 1965). Although this tension pervades most settings, the success of human societies is largely derived from our remarkable ability to cooperate and solve problems of collective action (Fehr & Gächter, 2000; Ostrom, 2000). Indeed, formal-theoretical models—that is, mathematical or computational models that express relations among variables and complex processes (for an introduction, see Smaldino, 2020)—suggest that the widespread tendency to cooperate can be explained by considering the ways in which contributing to the collective good can be profitable for the contributor (reviewed in Nowak, 2006). These putative mechanisms that promote cooperation include, for example, direct reciprocity (an individual will act in kind with a partner who has cooperated with them; Axelrod & Hamilton, 1981; Trivers, 1971), indirect reciprocity (an individual gains a positive reputation through cooperation: Alexander, 1987), genetic or kin relatedness (termed ‘kin selection’; an individual gains indirect fitness benefits by cooperating with relatives; Fromhage & Jennions, 2019; Hamilton, 1964), and partner choice (individuals compete for the most cooperative social partners and the most cooperative individuals are chosen; Barclay, 2016).

Although many of these mechanisms facilitate cooperation in small group settings, they often fail to explain the scale and complexity of cooperation characterising large human groups (Boyd & Richerson, 1988; Gardner & West, 2004). In such groups, cooperation may be effectively maintained by punishment (Boyd & Richerson, 1992). Punishment describes the event when an individual invests resources to reduce the well-being of a free-riding partner (Clutton-Brock & Parker, 1995). Cooperation can be maintained regardless of whether the punishing individual is a (a) ‘second-party punisher’, who has personally suffered harm; (b) ‘third-party punisher’ who, despite not having been personally harmed, chooses to inflict harm on a defector who has inflicted cost on others (Fehr & Fischbacher, 2004; Fehr & Gächter, 2000; Fehr & Gächter, 2002; Hirshleifer & Rasmusen, 1989); or (c) ‘centralized punisher’ who is designated to punish free-riders on behalf of other group members (e.g., police, government). This latter form of coordinated, institutional punishment, often exercised by a leader or authority figure who may or may not receive financial reimbursement for their services, has been shown to be, especially, effective in maintaining cooperation (Boyd, Gintis, & Bowles, 2010).

It is, of course, easy to see why second-party punishers, motivated by retaliation and revenge, would seek to punish violators. But why third parties—who have not been personally harmed (including centralized punishers who

do not receive a wage for their services)—would care to punish is puzzling. In fact, this situation wherein any observer would be better off by letting others punish produces the *second-order free-rider problem*, where individuals profit from the success of improved collective action by letting others pay the costs of, but personally refrain from, punishing.

In recent years, a promising body of research supplies an answer to this puzzle: Third-party punishment can be explained by considering the reputational gains that punitive action accrues (Brandt, Hauert, & Sigmund, 2003; Panchanathan & Boyd, 2004; Rockenbach & Milinski, 2006; Santos et al., 2011, 2013; Sigmund, Hauert, & Nowak, 2001). Increasing work has, for example, shown that punishers gain a trustworthy reputation and that this reputational benefit fuels future willingness to punish (Barclay, 2006; Raihani & Bshary, 2015a). Meanwhile, separate lines of evidence are beginning to shed light on other reputational consequences of punishment beyond trust. Consequently, a comprehensive understanding of the reputational gains to punishment must consider diverse domains alongside trust. Toward this end, we review research on the reputational gains to punishment in domains of trust, *prestige* (the perceived ability and willingness to confer benefits to others) and *dominance* (the perceived ability and willingness to inflict harm)—the latter of which are two distinct forms of social rank that form the foundations of human social hierarchy. We conclude that understanding why people punish requires a thorough consideration of how punishment contributes to the resolution of collection action problems and the diverse ways in which punishment is both facilitated by, and improves, one's reputation.

2 | WHY IS PUNISHMENT OFTEN INSTITUTIONALIZED? WHEN DOES NON-INSTITUTIONAL PUNISHMENT OCCUR?

To begin, it is important to recognize that across societies, many instances of punishment tend to operate in the form of institutional punishment (Guala, 2010; Sigmund, De Silva, Traulsen, & Hauert, 2010). Under institutional punishment, an individual (or a group of individuals, in the case of legal, or policing institutions) is designated as a representative who is tasked with punishing free-riders on behalf of group members using investments extracted from the group (e.g., paid taxes; Hilbe, Wu, Traulsen, & Nowak, 2014; Kuwabara & Yu, 2017; O'Gorman, Henrich, & Van Vugt, 2009; Traulsen, Röhl, & Milinski, 2012). For example, in a 'lab-in-field' experiment among Ugandan farmers, Baldassarri and Grossman (2011) found that centralized sanctioning effectively promoted cooperation, and that such sanctions were perceived with increased legitimacy—and were received with greater responsiveness—if they were enacted by individuals elected to be 'monitors' (i.e., those with the sole power to sanction) as opposed to individuals randomly allocated as punishers. Institutional punishment may be crucial for explaining the maintenance of cooperation because, relative to individually administered punishment (e.g., gossip, social exclusion, or withholding help; Balafoutas, Nikiforakis, & Rockenbach, 2014; Molho, Tybur, Van Lange, & Balliet, 2020), it is less vulnerable to problems of second-order free-riding and counter-punishment (Balafoutas, Nikiforakis, & Rockenbach, 2016).

By contrast, most people may be disinclined towards personally meting out individually administered punishment, due to costs associated with a high risk of counter-punishment from punished individuals (Cushman, 2015). Though less prevalent than institutional punishment, individually administered punishment remains an important mechanism that explains the emergence and maintenance of cooperation in many daily contexts. In particular, it tends to be exercised more frequently by certain individuals who, through possessing privileged status or other attributes, are able to reduce the transaction costs of punishment. Below, we turn to how individuals who possess prestige- and dominance-based status are often more willing to punish as their existing status reduces the net costs of punishment.

3 | THE COSTS AND BENEFITS OF THIRD-PARTY PUNISHMENT IN A HIERARCHICAL CONTEXT

In most human societies, individuals within social groups are informally arranged within social hierarchies. In these hierarchies, individuals have differential influence over group decision-making and privileged access to a group's resources, and deference is freely conferred to those who have respect, esteem and admiration (Anderson & Kilduff, 2009; Bales, Strodtbeck, Mills, & Roseborough, 1951; Berger, Rosenholtz, & Zelditch, 1980; Henrich & Gil-White, 2001; Redhead et al., 2019a). Those who are most willing and able to confer benefits to group members –such as offering valued knowledge, information, or contributing to the collective good—gain *prestige*. Evidence suggests that prestige is associated with an agreeable, prosocial and conscientious personality profile (Cheng, Tracy, & Henrich, 2010), alongside other cues to an ability and willingness to transfer culturally or evolutionary-relevant knowledge, expertise and skill (Henrich & Gil-White, 2001; Jiménez & Mesoudi, 2019).

Meanwhile, co-existing with prestige-based status is influence that derives from *dominance*. For example, individuals with egocentric, aggressive and coercive personality dispositions have been shown to be influential in newly formed North American student task groups (Cheng et al., 2010, 2013; Redhead et al., 2019), in mixed-age social groups in the United Kingdom (Brand & Mesoudi, 2019), and among employees and business managers in modern workplaces (Anderson, Sharps, Soto, & John, 2020). Similar findings have been made in a range of small-scale societies, where peer perceptions of dominance (e.g., fear) and dominance-related traits (e.g., physical formidability and position within a strong coalition) are linked to greater group-wide influence and control over resources among men (Garfield & Hagen, 2020; von Rueden, 2014) (for a review, see: Cheng et al., 2013; Cheng, 2020; Henrich & Gil-White, 2001)¹.

Below, we discuss the bi-directional association between these two forms of status and punishment. Individuals with either form of status may be more willing to punish, owing to their increased capital to pay the costs of punishing and thus can equalize these costs with relative ease (von Rueden et al., 2012); and in turn, their punitive action further increases their prestige or dominance.

3.1 | When individuals can 'afford' to punish: high status individuals have the unique resources to balance the costs of punishment

Punishment of lower-status members by higher-status members within a social group is common, as high-status punishers tend to have access to the capital necessary to bear the costs of punishing (Gavrilets, 2015), and may in turn enjoy higher net payoff from punitive action. Indeed, high status individuals are *expected* to make costly contributions to the public good across a multitude of social dilemmas (Halevy, Chou, Cohen, & Livingston, 2012), and predicted by others to be more effective punishers.

3.1.1 | Why do those high in prestige experience lower punishment costs?

Individuals high in prestige harbour important and distinct resources that reduce the costs of third-party punishment. This includes resources in the form of embodied capital (e.g., knowledge, physical strength), material capital (e.g., financial wealth) and relational capital (e.g., cooperative partners or allies, number of friends in privileged positions: von Rueden, 2014; von Rueden, Gurven, & Kaplan, 2008). For instance, an individual who has attracted a large number of cooperative partners or allies (a form of relational capital) through their high prestige will incur lower punishment costs, given their lower risk of retaliation, and greater perceived legitimacy as well as ease and efficiency in coordinating punitive action with allies (Nikiforakis, 2008).

Those high in prestige are also more capable at bearing the negative reputational repercussions that can sometimes follow punitive action. Beyond paying the material costs associated with sanctioning, punishers can occasionally face backlash from other group members when the severity of their punitive action is questioned, or when their action causes social unrest that undermines group unity (Adams & Mullen, 2012; Atwater, Waldman, Carey, & Cartier, 2001; van den Berg, Molleman, & Weissing, 2012). Prestigious individuals, with their high regard and positive reputation within a community, may receive more support and less scrutiny than less prestigious others. Consistent with this logic, studies indicate that prestigious individuals and their behaviours are more likely to be perceived by observers as fair, morally just and humble (Cheng et al., 2010; Effron & Miller, 2012; Price & Van Vugt, 2014; Weidman, Cheng, & Tracy, 2018). Even when revealed to have made transgressions, prestigious people are penalized less harshly than those with dominance-based rank (Fragale, Rosen, Xu, & Merideth, 2009; Kakkar, Sivanathan, & Gobel, 2020). This greater trust and loyalty that prestigious individuals receive can foster the perceived legitimacy of, and support for, any punitive action that they carry out (Baldassarri & Grossman, 2011). Consequently, prestigious individuals are able to more effectively coordinate and distribute the costs of any punitive action (de Waal-Andrews & Van Vugt, 2020; Magee & Langner, 2008).

Another mechanism through which prestigious individuals may experience higher punishment efficiency stems from their privileged status as cultural models for social learners. First, their punitive action may be readily emulated by other group members, thus elevating overall rates of punishment and leading to greater group-wide cooperation. Formal modelling work shows that followers readily retain the cooperation-enhancing behaviours they acquire from prestigious leaders, and influence other co-followers to adopt the same behaviours (Henrich, Chudek, & Boyd, 2015). Experimental and observational evidence further supports this notion that prestigious individuals' punishment decisions are likely emulated by others (FeldmanHall, Otto, & Phelps, 2018; Henrich & Boyd, 2001; Salali, Juda, & Henrich, 2015), and that the prosocial behaviours of such individuals are often copied (Gächter & Renner, 2018). In one study, for example, group members made higher donations to a public good in a voluntary contributions game when individuals high in prestige visibly made the first donation, compared to when an individual low in prestige was the first mover (Kumru & Vesterlund, 2010). Second, prestigious individuals who are sought after by learners who wish to acquire their knowledge can readily punish norm violators by simply withholding knowledge transmission (Bhui, Chudek, & Henrich, 2019), effectively serving a low-cost punishment. Overall, the possession of a positive reputation or capital in these domains thus allows one to lower or better weather any uncertain and indirect long-term costs that can follow punitive action.

3.1.2 | Why do individuals high in dominance also experience lower punishment costs?

Individuals high in dominance also harbour other distinct traits and motivations that effectively lower the costs of engaging in punishment. For any punisher, cultivating a formidable reputation is among the most effective ways of reducing the risk of retaliation, lowering the effort needed to enforce cooperation, and deterring any future cooperation partner from free-riding (Deutchman, Bračić, Raihani, & McAuliffe, 2020; Raihani & Bshary, 2015b). With a tough and threatening reputation, dominant punishers are also more able to threaten would-be free-riders into compliance by instilling a fear of *severe* repercussions. Empirical evidence indicates that individuals with a dominant reputation generally harbour traits—such as increased physical strength and size, raised levels of testosterone, a heightened disposition towards aggression, large coalitional support networks—that promote their ability and willingness to inflict harm or punish others (Capps, 2002; Cheng et al., 2010; Cheng & Kornienko, 2020; Dreher et al., 2016; Pfattheicher, Landhäußer, & Keller, 2014; Redhead et al., 2019b; von Rueden, Gurven, & Kaplan, 2011).

Although a dominant reputation may privilege individuals to engage in group-beneficial third-party punishment, there is reason to believe that dominant individuals may simultaneously be prone to *anti-social punishment*, which promotes personal benefits at the expense of the group. Anti-social punishment refers to the punishment of

cooperators by non-cooperators (Nikiforakis, 2008), often driven by the punisher's desire to protect one's self-interests by lowering the well-being of potential competitors (Pleasant & Barclay, 2018; Rand & Nowak, 2011). Consistent with this, studies have shown that people high in dominance are more likely to prioritize their own needs over the needs of the group (French & Raven, 1959; Maner & Mead, 2010; McClelland, 1970), and punish others in an attempt maintain a vacuum of power between themselves and other group members, and assert their control (Ellis, 1995; Maner & Mead, 2010). Similarly, other evidence indicates that those high in power and control—but not prestige—likely abuse their power, are motivated to punish as a form of retaliation, and behave in a demeaning manner towards peers, all of which, if revealed, increase one's dominant reputation but lowers one's prestige (Fast & Chen, 2009; Kuwabara & Yu, 2017; Zelditch & Walker, 2000).²

3.2 | Why punishment pays: the reciprocal benefits of punishment for increasing and maintaining status

We have outlined the distinct motivations, traits and capital that lower the costs of engagement in punishment for those high in prestige or dominance. Yet, the links that prestige and dominance have with punishment are likely bi-directional, such that punishing can in turn further promote either form of status. While research has yet to directly assess the effects of punishment on prestige or dominance, we next outline the theory and provide an overview the extant evidence suggestive of such dynamics.

3.2.1 | Why punishing can promote one's prestige

Punitive action may bolster an individual's prestigious reputation by being seen as a trustworthy and moral protector. Paying the cost of punishment to confer benefits to the group efficiently broadcasts an individual's ability and willingness to provide benefits to a potentially large number of group members. Theory suggests that prestige is founded on a market-like exchange relationship (Henrich & Gil-White, 2001; Price & Van Vugt, 2014), buttressed by evidence that whenever an individual pays a cost to provide a public good, group members receiving the good confer prestige to the producer. In a similar vein, in a longitudinal study von Rueden, Redhead, O'Gorman, Kaplan, and Gurven (2019) found that men with highly prestigious reputations enjoyed an increased probability of receiving help from others, and were more likely chosen as future cooperation partners, highlighting how those who contribute to public goods reap social and material rewards. According to this logic, given that third-party punishment can signal an individual's cooperative intent, such forms of prosocial punishment may also similarly promote prestige.

Just as punishing can elevate prestige, *failing* to punish under appropriate circumstances can lead to a *loss* of prestige. Studies indicate that those high in prestige who fail to punish norm violators can lose social support and influence, while those with lower prestige are not expected to 'right the wrong' and do not suffer status loss when they fail to intervene (Gordon & Lea, 2016). Thus, those high in prestige are not only more able and likely to pay the price of third-party punishment, but are also expected to. In turn, such individuals may be able to increase and maintain their prestige by signalling their cooperative intent, and promoting norms that benefit both the group and themselves through punishment of norm violators.

3.2.2 | Why punishing can promote one's dominance

Engaging in punishment not only promotes a prestigious reputation, but can also cultivate a dominant reputation. Indeed, it has been argued that a primary reputation that punishers cultivate in observers is fear (Raihani &

Bshary, 2015a, 2015b). To our knowledge, Gordon, Madden, and Lea (2014) and Gordon and Lea (2016) are the most direct evidence demonstrating that punishers acquire reputations of high dominance. Across a series of survey-based vignette studies administered to undergraduate students in the United Kingdom, participants rated third-party punishers as being high in dominance, even if their invention was unsuccessful (Gordon et al., 2014; Gordon & Lea, 2016). Alongside this, experimental evidence suggests that individuals strategically engage in third-party punishment to signpost their formidability, which effectively fosters enough fear to deter members of a group from mistreating the individual in future interactions (Krasnow, Delton, Cosmides, & Tooby, 2016). While these findings indicate that demonstrating one's willingness to harm the well-being of another—especially, if it entails the potential of incurring a personal cost—could promote dominant reputations, future research is necessary to determine whether these effects extend to real world settings.

As discussed above, however, despite indirectly raising levels of group cooperation, punishment meted out by dominantly inclined individuals may well be selfishly motivated by personal interests. Across a series of repeated Prisoner's Dilemma games with costly punishment, Dreber, Rand, Fudenberg, and Nowak (2008) found that the highest earners were not among those who punished, and that punishment did not increase average payoffs. While congruent evidence within human populations is scarce, there are a multitude of examples where non-human animals use punishment as a means of *coercion*, where by dominant individuals force subordinate individuals to partake in costly acts, or relinquish coveted resources, which would only increase the dominant individual's fitness (reviewed in Clutton-Brock & Parker, 1995; Raihani & Bshary, 2015a). This body of evidence suggests that punishment may also be a tool used for suppression and coercion, which increases the perceived dominance of stronger individuals, allowing them to dominate their weaker counterparts and control access to resources, and at times escalate within-group conflicts or lower the well-being of rival cooperators.

3.3 | Disentangling the reputational benefits of punishment

Throughout this review, we have highlighted several unique associations that prestige and dominance may have with third-party punishment. There are many reasons to believe that prestige and dominance may have variable connections to punishment. For example, antithetical to what has been outlined above for dominance, for example, individuals high in prestige are not likely to explicitly engage in punishment that may easily be perceived as serving self-interest at a detriment to the group, as evidence has shown negative associations between egocentric behaviours and prestige (Cheng et al., 2010). Rather, individuals high in prestige are likely to engage in forms of punishment that increase their appeal as a social partner, and build reputations of being trustworthy (Barclay, 2006; Raihani & Bshary, 2015a). Formal theory suggests that the costs of punishing may be recouped in the long-run because punishers effectively signal their trustworthiness and ability to deter future harm against others (Jordan, Hoffman, Bloom, & Rand, 2016). Consistent with this, laboratory evidence indicates that punishers are compensated for their incurred costs as they become entrusted by others, preferred as interaction partners and receive support (Barclay, 2006; Horita, 2010; Nelissen, 2008). However, such results are only uncovered when punishment is regarded as group-oriented, rather than driven by any personal gains or desire for personal revenge (Wang & Murnighan, 2017). Further highlighting the impact of punitive action on shaping one's trustworthiness, recent evidence indicates that people strategically punish more when they are unable to create a more trustworthy reputation through prosocial means (Jordan & Rand, 2019).

It is important to emphasise that the distinction between prestige and dominance (and their associations with punishment) cannot always be easily disentangled. There are a multitude of factors that can produce correlated motivations, actions and outcomes between these two domains. Both prestige and dominance are platforms that foster enough comparable forms of capital for an individual to either effectively balance the costs—or reap the benefits—of third-party punishment. In the absence of such capital, the barriers-to-entry are much too high for many individuals to access any substantial benefits from punishment. Alongside this, the role of

third-party punishment by those high in dominance can be conceived as services rendered for a group or coalition. Such services made possible by the capital harboured by those high in dominance may, in fact, produce positive reputations associated with prestige (Price & Van Vugt, 2014). On the other hand, any punishment of others by individuals high in prestige will most likely feed into perceptions of dominance, as these acts of punishment will signal both their ability and willingness to impart costs on others—albeit through different motivations and enacted through more legitimate, low-cost forms. While we have only provided a handful of examples that touch upon the complex relationships that prestige and dominance have with punishment, the ways in which prestige and dominance may interact in cooperative and punitive contexts are manifold. Understanding the settings and behaviours that bring about the directional associations between prestige and dominance—especially, in relation to punishment and collective action more generally—is an important avenue for future research.

4 | CONCLUSIONS AND FUTURE RESEARCH

The primary aim of this review has been to outline the ways in which punitive action may shape, and be shaped by, one's reputation. We hope that this review will help integrate disparate empirical and theoretical lines of work under a broadened umbrella that considers the direct and indirect benefits that third-party punishment has on three domains of reputation: trust, prestige and dominance. In closing, we highlight several important questions for future research.

Future research should examine how the effects of punishment on reputation may vary across cultures (Henrich et al., 2001, 2005; Herrmann, Thöni, & Gächter, 2008). Some evidence indicates that prosocial punishment may fail to effectively increase cooperation in some societies (Wu et al., 2009), and its cooperation-enhancing effect may be weak or non-existent in contexts that lack trust (Balliet & Van Lange, 2013). In contexts where punishment may be less effective at maintaining cooperation, punishers may accordingly receive fewer reputational rewards. Cultures that place a greater emphasis on interpersonal warmth and harmonious relationships may also be less inclined to reward punishment (Torelli, Leslie, Stoner, & Puente, 2014).

Gender may also be an important moderating factor. Men and women may differ in the domains of conflict in which they most readily intervene. In traditional societies, for example, women are found to mediate conflicts over sexual affairs, disputes over negligence in child care, and friendship conflicts, whereas men are reported to mediate in conflicts over land, theft and free-riding (von Rueden, Alami, Kaplan, & Gurven, 2018). Other evidence from small-scale societies indicates that women play more important roles than men in bringing attention to norm violations and conflict mediation, perhaps owing to women's relatively lower risks of harsh physical retaliation (Bowser & Patton, 2010; Wiessner, 2005). Finally, men may engage in punishment driven by status or mating motivations (Griskevicius et al., 2009), rather than out of any prosocial orientation, which could lower the reputational gains they experience following punitive action. Future comparative research is needed to assess the psychological, evolutionary and cultural mechanisms that may guide the forms and functions of third-party punishment across groups and societies.

Beyond differences across-cultural and ecological context, punishment may also have variable implications on reputation depending upon timescales. For instance, in some contexts, the value of punishers may become increasingly apparent to group members as threats to collective action intensify, which would lead to more (rather than less) trust and prestige to be conferred to cooperation-enhancing punishers over time. In light of this, longitudinal design (or long-run iterative behavioural games) are needed to assess the directional associations between third-party punishment and reputation. For example, future research may tease apart when prestige and dominance predict third-party punishment, when punishment may increase reputations of prestige and dominance, and in what stages of group formation punishment pays and proliferates. While prestige and dominance may have different temporal associations with third-party punishment, the degree to which

engagement in punishment is beneficial to the individual—at any point in time—derives from how such acts are received by their peers (Blau, 1964; Zelditch & Walker, 2000). The structural within-group conditions (e.g., the characteristics of the social networks in which individuals operate) that govern acts of third-party punishment, and when such acts are perceived as abuses of power and produce negative reputations, is a fruitful avenue for future research.

The foregoing discussion would appear to imply that punishment always results in a net positive gain for the punisher. However, this is not the case. While individuals can sometimes recoup the costs of punishment through reputational gains, punishment does not always pay for the individual. For example, punishment may incur costs beyond the direct price of enforcement, such as risk of retaliation and reduction of fitness through partner choice. Formal models indicate that introducing the possibility of retaliation from punished defectors reduces both prosocial punishment and cooperation, especially, when retaliation is equally or more costly than punishment itself (Janssen & Bushman, 2008). Indeed, when surveyed about their reaction to norm violations, 60% of individuals reported that they would not punish the violator due to fear of counter-punishment (i.e., the fear of being reciprocally harmed by those who one had punished; Balafoutas et al., 2016). Similarly, in experiments that do introduce the threat of feuds, prosocial punishment is greatly reduced (Nikiforakis & Engelmann, 2011). Beyond direct retaliation, punished violators are also less likely to choose the punisher as a social partner in future interactions. These potential costs notwithstanding, punishers may nevertheless experience net gains in cooperative partners so long as the gains in new or strengthened cooperative ties with other group members exceed the losses resulting from the withdrawal of cooperation from the punished party. Overall, determining the net costs of punishment must involve a complex tabulation of the myriad costs against any reputational gains. When the costs of punishment and risk of counter-punishment are relatively low, punish may pay in the long run.

Finally, punishment is, of course, not the only means to maintaining cooperation; contributing to collective action through alternative means can also result in reputational gains. For example, rather than punish violators, third-party individuals may also compensate those who have been the victims of acts of defection (Heffner & FeldmanHall, 2019). Experiment evidence indicates that, when given a free choice, individuals in fact prefer to use a mix of *both* punishment and compensation as means to enforce cooperative norms (Lotz, Okimoto, Schlösser, & Fetchenhauer, 2011). Moreover, third-party helpers are rewarded with trust to an even greater degree than punishers, emphasising the utility of conflict resolution by non-punitive means (Anderson et al., 2020; Raihani & Bshary, 2015b). Thus, people may be more willing to reward cooperators—and punish those who do not reward cooperators—than to punish those that do not cooperate (Kiyonari & Barclay, 2008). By conferring aid and benefiting other group members through these prosocial means, individuals can increase and sustain both group-level cooperation, and their individual levels of prestige (Henrich et al., 2015; von Rueden et al., 2019).

By answering these questions, future research will produce a deeper understanding of the individual differences and group processes that bind social status and third-party punishment. In turn, such research will elucidate the role of social hierarchy in promoting collective action and increasing cooperation.

ACKNOWLEDGEMENTS

We would like to thank Professor Sebastien Brion, and two anonymous reviewers for their helpful feedback. Daniel Redhead was supported by the Department of Human Behaviour, Ecology and Culture at the Max Planck Institute for Evolutionary Anthropology.

Open access funding enabled and organized by Projekt DEAL.

ORCID

Daniel Redhead  <https://orcid.org/0000-0002-2809-8121>

ENDNOTES

- ¹ Despite the existence of these distinct forms of status in humans (prestige and dominance), their prevalence and effects on an individual's position within a group may vary across contexts (Powers & Lehmann, 2014; von Rueden, 2020; Cheng, 2020). For instance, prestige may play a more important role in environments where survival requires a greater degree of interdependence, such as when groups are in particular need of skilled others to aid in the production of food and other resources for survival (Kaplan et al., 2009; Hooper et al., 2010), or knowledgeable peers for cultural learning (Henrich & Gil-White, 2001). Meanwhile, dominance may be more muted in groups that are better able to form levelling coalitions against would-be dominants (Gavrilets et al., 2008; Boehm et al., 1993). Such context-dependencies have likely paved the path for a greater reliance on prestige as an avenue for social status in humans (von Rueden et al., 2019), and a greater fragility of dominance-based approaches to status (Cheng, 2020).
- ² Several factors may cause variation in the frequency and effectiveness of anti-social forms of punishment between groups. Although anti-social punishment has been observed across diverse cultures, including small-scale societies (Boehm, 1999), substantial cross-cultural variation exists (Herrmann et al., 2008). Individuals high in dominance may be most motivated to anti-socially punish highly skilled and competent group members to protect their self-interest when their position in a hierarchy is threatened or unstable (Maner & Mead, 2010). Other factors that may moderate anti-social punishment and its link with dominance are the veracity of a group's authoritative institutions, and their norms of civic cooperation (Herrmann et al., 2008).

REFERENCES

- Adams, G. S., & Mullen, E. (2012). The social and psychological costs of punishing. *Behavioral and Brain Sciences*, 35(1), 15–16.
- Alexander, R. D. (1987). *The biology of moral systems*. New Jersey, NJ: Transaction Publishers.
- Anderson, C., & Kilduff, G. J. (2009). The pursuit of status in social groups. *Current Directions in Psychological Science*, 18(5), 295–298.
- Anderson, C., Sharps, D. L., Soto, C. J., & John, O. P. (2020). People with disagreeable personalities (selfish, combative, and manipulative) do not have an advantage in pursuing power at work. *Proceedings of the National Academy of Sciences*, 117(37), 22780–22786.
- Atwater, L. E., Waldman, D. A., Carey, J. A., & Cartier, P. (2001). Recipient and observer reactions to discipline: Are managers experiencing wishful thinking? *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 22(3), 249–270.
- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211(4489), 1390–1396.
- Balafoutas, L., Nikiforakis, N., & Rockenbach, B. (2014). Direct and indirect punishment among strangers in the field. *Proceedings of the National Academy of Sciences*, 111(45), 15924–15927.
- Balafoutas, L., Nikiforakis, N., & Rockenbach, B. (2016). Altruistic punishment does not increase with the severity of norm violations in the field. *Nature Communications*, 7(1), 1–6.
- Baldassarri, D., & Grossman, G. (2011). Centralized sanctioning and legitimate authority promote cooperation in humans. *Proceedings of the National Academy of Sciences*, 108(27), 11023–11027.
- Bales, R. F., Strodtbeck, F. L., Mills, T. M., & Roseborough, M. E. (1951). Channels of communication in small groups. *American Sociological Review*, 16(4), 461–468.
- Balliet, D., & Van Lange, P. A. (2013). Trust, punishment, and cooperation across 18 societies: A meta-analysis. *Perspectives on Psychological Science*, 8(4), 363–379.
- Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior*, 27(5), 325–344.
- Barclay, P. (2016). Biological markets and the effects of partner choice on cooperation and friendship. *Current opinion in psychology*, 7, 33–38.
- Berger, J., Rosenholtz, S. J., & Zelditch, Jr, M. (1980). Status organizing processes. *Annual Review of Sociology*, 6(1), 479–508.
- Bhui, R., Chudek, M., & Henrich, J. (2019). How exploitation launched human cooperation. *Behavioral Ecology and Sociobiology*, 73(6), 78.
- Blau, P. M. (1964). Justice in social exchange. *Sociological Inquiry*, 34(2), 193–206.
- Boehm, C. (1999). *Hierarchy in the forest: The evolution of egalitarian behavior*. Massachusetts, MA: Harvard University Press.
- Boehm, C., Barclay, H. B., Dentan, R. K., Dupre, M.-C., Hill, J. D., Kent, S., ... Rayner, S. (1993). Egalitarian behavior and reverse dominance hierarchy [and comments and reply]. *Current Anthropology*, 34(3), 227–254.
- Bowser, B., & Patton, J. (2010). *Women's leadership: Political alliance, economic resources, and reproductive success in the Ecuadorian amazon. The evolution of leadership: Transitions in decision making from small-scale to middle-range societies*. 51–71.
- Boyd, R., Gintis, H., & Bowles, S. (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science*, 328(5978), 617–620.

- Boyd, R., & Richerson, P. J. (1988). The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology*, 132(3), 337–356.
- Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, 13(3), 171–195.
- Brand, C. O., & Mesoudi, A. (2019). Prestige and dominance-based hierarchies exist in naturally occurring human groups, but are unrelated to task-specific knowledge. *Royal Society open science*, 6(5), 181621.
- Brandt, H., Hauert, C., & Sigmund, K. (2003). Punishment and reputation in spatial public goods games. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1519), 1099–1104.
- Capps, J. S. (2002). Explaining punitiveness: Right-wing authoritarianism and social dominance. *North American Journal of Psychology*, 4(2), 263–278.
- Cheng, J. T. (2020). Dominance, prestige, and the role of leveling in human social hierarchy and equality. *Current Opinion in Psychology*, 33, 238–244.
- Cheng, J. T., & Kornienko, O. (2020). The neurobiology of human social behavior: A review of how testosterone and cortisol underpin competition and affiliation dynamics. *Salivary bioscience* (pp. 519–553) Springer.
- Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Two ways to the top: Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *Journal of Personality and Social Psychology*, 104(1), 103.
- Cheng, J. T., Tracy, J. L., & Henrich, J. (2010). Pride, personality, and the evolutionary foundations of human social status. *Evolution and Human Behavior*, 31(5), 334–347.
- Clutton-Brock, T. H., & Parker, G. A. (1995). Punishment in animal societies. *Nature*, 373(6511), 209–216.
- Cushman, F. (2015). Punishment in humans: From intuitions to institutions. *Philosophy Compass*, 10(2), 117–133.
- de Waal-Andrews, W., & van Vugt, M. (2020). The triad model of follower needs: Theory and review. *Current opinion in psychology*, 33, 142–147.
- Deutchman, P., Bračić, M., Raihani, N., & McAuliffe, K. (2020). Punishment is strongly motivated by revenge and weakly motivated by inequity aversion. *Evolution and Human Behavior*, 41(1), 12–20.
- Dreber, A., Rand, D. G., Fudenberg, D., & Nowak, M. A. (2008). Winners don't punish. *Nature*, 452(7185), 348–351.
- Dreher, J.-C., Dunne, S., Pazderska, A., Frodl, T., Nolan, J. J., & O'Doherty, J. P. (2016). Testosterone causes both prosocial and antisocial status-enhancing behaviors in human males. *Proceedings of the National Academy of Sciences*, 113(41), 11633–11638.
- Dugatkin, L. A. (1997). *Cooperation among animals: An evolutionary perspective*. United Kingdom: Oxford University Press.
- Effron, D. A., & Miller, D. T. (2012). How the moralization of issues grants social legitimacy to act on one's attitudes. *Personality and Social Psychology Bulletin*, 38(5), 690–701.
- Ellis, L. (1995). Dominance and reproductive success among nonhuman animals: A cross-species comparison. *Ethology and Sociobiology*, 16(4), 257–333.
- Fast, N. J., & Chen, S. (2009). When the boss feels inadequate: Power, incompetence, and aggression. *Psychological Science*, 20(11), 1406–1413.
- Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, 25(2), 63–87.
- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *The American Economic Review*, 90(4), 980–994.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137–140.
- FeldmanHall, O., Otto, A. R., & Phelps, E. A. (2018). Learning moral values: Another's desire to punish enhances one's own punitive behavior. *Journal of Experimental Psychology: General*, 147(8), 1211.
- Fragale, A. R., Rosen, B., Xu, C., & Merideth, I. (2009). The higher they are, the harder they fall: The effects of wrongdoer status on observer punishment recommendations and intentionality attributions. *Organizational Behavior and Human Decision Processes*, 108(1), 53–65.
- French, J. R., & Raven, B. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies in social power* (pp. 150–167). Ann Arbor, MI: University of Michigan Press.
- Fromhage, L., & Jennions, M. D. (2019). The strategic reference gene: An organismal theory of inclusive fitness. *Proceedings of the Royal Society B*, 286(1904), 20190459.
- Gächter, S., & Renner, E. (2018). Leaders as role models and 'belief managers' in social dilemmas. *Journal of Economic Behavior & Organization*, 154, 321–334.
- Gardner, A., & West, S. A. (2004). Cooperation and punishment, especially in humans. *The American Naturalist*, 164(6), 753–764.
- Garfield, Z. H., & Hagen, E. H. (2020). Investigating evolutionary models of leadership among recently settled ethiopian hunter-gatherers. *The Leadership Quarterly*, 31(2), 101290.
- Gavrilets, S. (2015). Collective action problem in heterogeneous groups. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1683), 20150016.

- Gavrilets, S., Duenez-Guzman, E. A., & Vose, M. D. (2008). Dynamics of alliance formation and the egalitarian revolution. *PLoS One*, 3(10), e3293.
- Gordon, D. S., & Lea, S. E. (2016). Who punishes? The status of the punishers affects the perceived success of, and indirect benefits from, "moralistic" punishment. *Evolutionary Psychology*, 14(3), 1474704916658042.
- Gordon, D. S., Madden, J. R., & Lea, S. E. (2014). Both loved and feared: Third party punishers are viewed as formidable and likeable, but these reputational benefits may only be open to dominant individuals. *PLoS One*, 9(10), e110045.
- Griskevicius, V., Tybur, J. M., Gangestad, S. W., Perea, E. F., Shapiro, J. R., & Kenrick, D. T. (2009). Aggress to impress: Hostility as an evolved context-dependent strategy. *Journal of Personality and Social Psychology*, 96(5), 980.
- Guala, F. (2010). *Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate*. University of Milan Department of Economics, Business and Statistics Working Paper.
- Halevy, N., Chou, E. Y., Cohen, T. R., & Livingston, R. W. (2012). Status conferral in intergroup social dilemmas: Behavioral antecedents and consequences of prestige and dominance. *Journal of Personality and Social Psychology*, 102(2), 351.
- Hamilton, W. D. (1964). The genetical evolution of social behaviour. ii. *Journal of Theoretical Biology*, 7(1), 17–52.
- Hardin, R. (1982). *Collective action*. Resources for the future.
- Heffner, J., & FeldmanHall, O. (2019). Why we don't always punish: Preferences for non-punitive responses to moral violations. *Scientific Reports*, 9(1), 1–13.
- Henrich, J., & Boyd, R. (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology*, 208(1), 79–89.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In search of homo economicus: Behavioral experiments in 15 small-scale societies. *The American Economic Review*, 91(2), 73–78.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., ... Tracer, D. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28(6), 795–815. <https://doi.org/10.1017/S0140525X05000142>
- Henrich, J., Chudek, M., & Boyd, R. (2015). The big man mechanism: How prestige fosters cooperation and creates prosocial leaders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1683), 20150013.
- Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, 22(3), 165–196.
- Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, 319(5868), 1362–1367.
- Hilbe, C., Wu, B., Traulsen, A., & Nowak, M. A. (2014). Cooperation and control in multiplayer social dilemmas. *Proceedings of the National Academy of Sciences*, 111(46), 16425–16430.
- Hirshleifer, D., & Rasmusen, E. (1989). Cooperation in a repeated prisoners' dilemma with ostracism. *Journal of Economic Behavior & Organization*, 12(1), 87–106.
- Hooper, P. L., Kaplan, H. S., & Boone, J. L. (2010). A theory of leadership in human cooperative groups. *Journal of Theoretical Biology*, 265(4), 633–646.
- Horita, Y. (2010). Punishers may be chosen as providers but not as recipients. *Letters on Evolutionary Behavioral Science*, 1(1), 6–9.
- Janssen, M. A., & Bushman, C. (2008). Evolution of cooperation and altruistic punishment when retaliation is possible. *Journal of Theoretical Biology*, 254(3), 541–545.
- Jiménez, Á. V., & Mesoudi, A. (2019). Prestige-biased social learning: Current evidence and outstanding questions. *Palgrave Communications*, 5(1), 1–12.
- Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature*, 530(7591), 473–476.
- Jordan, J. J., & Rand, D. G. (2019). Signaling when no one is watching: A reputation heuristics account of outrage and punishment in one-shot anonymous interactions. *Journal of Personality and Social Psychology*, 118(1), 57–88.
- Kakkar, H., Sivanathan, N., & Gobel, M. S. (2020). Fall from grace: The role of dominance and prestige in the punishment of high-status actors. *Academy of Management Journal*, 63(2), 530–553.
- Kaplan, H. S., Hooper, P. L., & Gurven, M. (2009). The evolutionary and ecological roots of human social organization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1533), 3289–3299.
- Kiyonari, T., & Barclay, P. (2008). Cooperation in social dilemmas: Free riding may be thwarted by second-order reward rather than by punishment. *Journal of Personality and Social Psychology*, 95(4), 826.
- Krasnow, M. M., Delton, A. W., Cosmides, L., & Tooby, J. (2016). Looking under the hood of third-party punishment reveals design for personal benefit. *Psychological Science*, 27(3), 405–418.
- Kumru, C. S., & Vesterlund, L. (2010). The effect of status on charitable giving. *Journal of Public Economic Theory*, 12(4), 709–735.
- Kuwabara, K., & Yu, S. (2017). Costly punishment increases prosocial punishment by designated punishers: Power and legitimacy in public goods games. *Social Psychology Quarterly*, 80(2), 174–193.

- Lotz, S., Okimoto, T. G., Schlösser, T., & Fetchenhauer, D. (2011). Punitive versus compensatory reactions to injustice: Emotional antecedents to third-party interventions. *Journal of Experimental Social Psychology, 47*(2), 477–480.
- Magee, J. C., & Langner, C. A. (2008). How personalized and socialized power motivation facilitate antisocial and prosocial decision-making. *Journal of Research in Personality, 42*(6), 1547–1559.
- Maner, J. K., & Mead, N. L. (2010). The essential tension between leadership and power: When leaders sacrifice group goals for the sake of self-interest. *Journal of Personality and Social Psychology, 99*(3), 482.
- McClelland, D. C. (1970). The two faces of power. *Journal of International Affairs, 24*(1), 29–47.
- Molho, C., Tybur, J. M., Van Lange, P. A., & Balliet, D. (2020). Direct and indirect punishment of norm violations in daily life. *Nature Communications, 11*(1), 1–9.
- Nelissen, R. M. (2008). The price you pay: Cost-dependent reputation effects of altruistic punishment. *Evolution and Human Behavior, 29*(4), 242–248.
- Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics, 92*(1–2), 91–112.
- Nikiforakis, N., & Engelmann, D. (2011). Altruistic punishment and the threat of feuds. *Journal of Economic Behavior & Organization, 78*(3), 319–332.
- Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science, 314*(5805), 1560–1563.
- O’Gorman, R., Henrich, J., & Van Vugt, M. (2009). Constraining free riding in public goods games: Designated solitary punishers can sustain human cooperation. *Proceedings of the Royal Society B: Biological Sciences, 276*(1655), 323–329.
- Olson, M. (1965). *The logic of collective action: Public goods and the theory of Groups*, (2nd ed.). Harvard University Press.
- Ostrom, E. (2000). Collective action and the evolution of social norms. *The Journal of Economic Perspectives, 14*(3), 137–158.
- Panchanathan, K., & Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature, 432*(7016), 499–502.
- Pfattheicher, S., Landhäußer, A., & Keller, J. (2014). Individual differences in antisocial punishment in public goods situations: The interplay of cortisol with testosterone and dominance. *Journal of Behavioral Decision Making, 27*(4), 340–348.
- Pleasant, A., & Barclay, P. (2018). Why hate the good guy? Antisocial punishment of high cooperators is greater when people compete to be chosen. *Psychological Science, 29*(6), 868–876.
- Powers, S. T., & Lehmann, L. (2014). An evolutionary model explaining the neolithic transition from egalitarianism to leadership and despotism. *Proceedings of the Royal Society B: Biological Sciences, 281*(1791), 20141349.
- Price, M. E., & Van Vugt, M. (2014). The evolution of leader–follower reciprocity: The theory of service-for-prestige. *Frontiers in Human Neuroscience, 8*, 363.
- Raihani, N. J., & Bshary, R. (2015a). The reputation of punishers. *Trends in Ecology & Evolution, 30*(2), 98–103.
- Raihani, N. J., & Bshary, R. (2015b). Third-party punishers are rewarded, but third-party helpers even more so. *Evolution, 69*(4), 993–1003.
- Rand, D. G., & Nowak, M. A. (2011). The evolution of antisocial punishment in optional public goods games. *Nature Communications, 2*(1), 1–7.
- Redhead, D., Cheng, J., & O’Gorman, R. (2019a). Higher Status in Group. In: T. Shackelford & V. Weekes-Shackelford (Eds.), *Encyclopedia of Evolutionary Psychological Science*. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-16999-6_3495-1
- Redhead, D., Cheng, J., & O’Gorman, R. (2019b). Individuals that Impose Costs. In T. Shackelford & V. Weekes-Shackelford (Eds.), *Journal: Encyclopedia of Evolutionary Psychological Science*. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-16999-6_3572-1
- Redhead, D., Cheng, J. T., Driver, C., Foulsham, T., & O’Gorman, R. (2019). On the dynamics of social hierarchy: A longitudinal investigation of the rise and fall of prestige, dominance, and social rank in naturalistic task groups. *Evolution and Human Behavior, 40*(2), 222–234.
- Rockenbach, B., & Milinski, M. (2006). The efficient interaction of indirect reciprocity and costly punishment. *Nature, 444*(7120), 718–723.
- Salali, G. D., Juda, M., & Henrich, J. (2015). Transmission and development of costly punishment in children. *Evolution and Human Behavior, 36*(2), 86–94.
- Santos, M. d., Rankin, D. J., & Wedekind, C. (2011). The evolution of punishment through reputation. *Proceedings of the Royal Society B: Biological Sciences, 278*(1704), 371–377.
- Santos, M. d., Rankin, D. J., & Wedekind, C. (2013). Human cooperation based on punishment reputation. *Evolution, 67*(8), 2446–2450.
- Sigmund, K., De Silva, H., Traulsen, A., & Hauert, C. (2010). Social learning promotes institutions for governing the commons. *Nature, 466*(7308), 861–863.
- Sigmund, K., Hauert, C., & Nowak, M. A. (2001). Reward and punishment. *Proceedings of the National Academy of Sciences, 98*(19), 10757–10762.

- Smaldino, P. E. (2020). How to translate a verbal theory into a formal model. *Social Psychology*, 51(4), 207–218.
- Smith, E. (2003). Human cooperation: Perspectives from behavioral ecology. In P. Hammerstein (Ed), *Genetic and cultural evolution of cooperation* (pp. 401–428). MIT Press.
- Torelli, C. J., Leslie, L. M., Stoner, J. L., & Puente, R. (2014). Cultural determinants of status: Implications for workplace evaluations and behaviors. *Organizational Behavior and Human Decision Processes*, 123(1), 34–48.
- Traulsen, A., Röhl, T., & Milinski, M. (2012). An economic experiment reveals that humans prefer pool punishment to maintain the commons. *Proceedings of the Royal Society B: Biological Sciences*, 279(1743), 3716–3721.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46(1), 35–57.
- van den Berg, P., Molleman, L., & Weissing, F. J. (2012). The social costs of punishment. *Behavioral and Brain Sciences*, 35(1), 42–43.
- Von Rueden, C. (2014). The roots and fruits of social status in small-scale human societies. *The psychology of social status* (pp. 179–200). Springer.
- von Rueden, C. (2020). Making and unmaking egalitarianism in small-scale human societies. *Current opinion in psychology*, 33, 167–171.
- von Rueden, C., Alami, S., Kaplan, H., & Gurven, M. (2018). Sex differences in political leadership in an egalitarian society. *Evolution and Human Behavior*, 39(4), 402–411.
- von Rueden, C., Gurven, M., & Kaplan, H. (2008). The multiple dimensions of male social status in an amazonian society. *Evolution and Human Behavior*, 29(6), 402–415.
- von Rueden, C., Gurven, M., & Kaplan, H. (2011). Why do men seek status? Fitness payoffs to dominance and prestige. *Proceedings of the Royal Society B: Biological Sciences*, 278(1715), 2223–2232.
- von Rueden, C. R., Gurven, M., & Guala, F. (2012). When the strong punish: Why net costs of punishment are often negligible. *Behavioral and Brain Sciences*, 35(1), 43.
- von Rueden, C. R., Redhead, D., O’Gorman, R., Kaplan, H., & Gurven, M. (2019). The dynamics of men’s cooperation and social status in a small-scale society. *Proceedings of the Royal Society B*, 286(1908), 20191367.
- Wang, L., & Murnighan, J. K. (2017). The dynamics of punishment and trust. *Journal of Applied Psychology*, 102(10), 1385.
- Weidman, A. C., Cheng, J. T., & Tracy, J. L. (2018). The psychological structure of humility. *Journal of Personality and Social Psychology*, 114(1), 153.
- Wiessner, P. (2005). Norm enforcement among the ju/’hoansi bushmen. *Human Nature*, 16(2), 115–145.
- Wu, J.-J., Zhang, B.-Y., Zhou, Z.-X., He, Q.-Q., Zheng, X.-D., Cressman, R., & Tao, Y. (2009). Costly punishment does not always increase cooperation. *Proceedings of the National Academy of Sciences*, 106(41), 17448–17451.
- Zelditch, M., & Walker, H. A. (2000). The normative regulation of power. *Advances in Group Processes*, 17, 155–178.

How to cite this article: Redhead D, Dhaliwal N, Cheng JT. Taking charge and stepping in: Individuals who punish are rewarded with prestige and dominance. *Soc Personal Psychol Compass*. 2021;15:e12581. <https://doi.org/10.1111/spc3.12581>