



Policing and punishment across the domains of social evolution

Manvir Singh and Jacobus J. Boomsma

M. Singh (manvir Singh@fas.harvard.edu) and J. J. Boomsma, Centre for Social Evolution, Dept of Biology, Univ. of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark. Present address for MS: Dept of Human Evolutionary Biology, Harvard Univ., Peabody Museum, 11 Divinity Avenue, Cambridge, MA 02318, USA.

Several decades of research in humans, other vertebrates, and social insects have offered fascinating insights into the dynamics of punishment (and its subset, policing), but authors have only rarely addressed whether there are fundamental joint principles underlying the maintenance of these behaviors. Here we present a punisher/bystander approach rooted in inclusive fitness logic to predict which individuals should take on punishing roles in animal societies. We apply our scheme to societies of eusocial Hymenoptera and nonhuman vertebrate social breeders, and we outline potential extensions for understanding conflict regulation among cells in metazoan bodies and unrelated individuals in human societies. We highlight that: 1) no social unit is expected to express punishment behavior unless it collects positive inclusive fitness benefits that surpass alternative benefits of bystanding; 2) punishment with public good benefits can be maintained through either direct fitness benefits (coercion) or indirect fitness benefits (correction) or both; 3) differences across social systems in the distributions of power, relatedness, and reproductive options drive variation in the extent to which individuals actively punish; and 4) inclusive fitness logic captures many punishment-relevant evolutionary and ecological variables in a single framework that appears to apply across very different types of social arrangements.

Synthesis

Researchers have long observed that individuals in animal societies punish (and by extension, police) each other, but they have rarely investigated whether general principles underlie this behavior across social arrangements. In this paper, we present a punisher/bystander approach rooted in inclusive fitness logic to predict which individuals should take on punisher roles in animal societies. We apply the approach to eusocial insects and cooperatively breeding vertebrates and outline extensions towards the control of cancer cell lineages and punishment in human groups. We highlight how variation in specific social variables may drive differences in punishing/policing across the social domains.

Over the past two decades, questions concerning punishment have surged to the forefronts of research in behavior, evolution, and economic decision-making. Although punishment was long considered to be uniquely human, contributions in the late eighties and early nineties showed that this behavior also plays a role in minimizing social conflict in non-human societies. Seminal papers, including Ratnieks' (1988) description of worker-policing in honeybees, Hauser's (1992) account of punishment in macaques, and Frank's (1995) model of competition repression, collectively underscored the prevalence and evolvability of these behaviors in natural social groups. Clutton-Brock and Parker (1995a) consolidated these approaches, emphasizing punishment's role in both disrupting and maintaining cooperation, and presented a simple model to demonstrate how punishment could evolve and remain stable in two-player dyads. Together with Ratnieks' and Frank's models, these insights provided much of the general groundwork for later policing and punishment research.

The pioneering work of the 1980 and 1990s inspired an impressively diverse literature, with novel insights in

eusocial insects (Ratnieks et al. 2006), social vertebrates (Flack et al. 2006), human psychology (Fehr et al. 2002, Chaudhuri 2010), intragenomic conflict (Hurst et al. 1996), and cancer-suppression (Dunn et al. 2004), but some fundamental general questions concerning the evolutionary ecology of punishment across these domains have remained unexplored. Most ambiguities surround punishment behaviors which involve an actor (often third-party) paying a cost to penalize a defector and in turn producing a valuable public good, such as reduced intragroup aggression (see next section for definitions of 'punishment' and 'policing'). Although punishment with public good benefits (what we term 'policing') has been noted across the social domains, explicit attempts to synthesize the principles underlying these behaviors have not been attempted. This leaves open questions like: are there unifying principles that predict which individuals police across the different domains of social evolution? And, if these principles exist, should we expect policing to evolve in all social arrangements? This second question is important in light of the recent contention over whether the

evolution of punishment is constrained in some natural social groups, but not in others (Fehr et al. 2002, Jensen 2010, Riedl et al. 2012, Raihani et al. 2012).

Maintaining punishment (and thereby policing) as a public good by natural selection requires that individuals producing the good obtain sufficient fitness benefits that compensate for the associated costs. These benefits can be both direct (affect personal reproduction) and indirect (affect reproduction of relatives), so they are best considered in an inclusive fitness framework (West and Gardner 2013). This paradigm implies that natural selection appears to ‘design’ adaptations that maximize individuals’ inclusive fitness via the replication of genes that shape phenotypes of social behavior (Grafen 2006).

In this essay, we present an inclusive fitness-based approach for predicting which individuals should punish in different animal societies. We test its utility by applying it both to eusocial Hymenoptera and vertebrate social breeders before exploring the approach’s potential for understanding conflict suppression within metazoan bodies and human societies. In applying our framework, we briefly review public good punishment (or policing) across the domains of social evolution. We conclude by connecting our approach with recent conceptual advances based on the degrees of social “organismality” (Queller and Strassmann 2009) and parental commitment (Boomsma 2013) to address whether punishment should be coercive or corrective. We use the term ‘coercive’ to describe when individuals, typically dominants, punish to force group-mates into delivering them direct fitness benefits, while ‘corrective’ describes when individuals, usually non-reproductives, punish solely to maximize indirect fitness benefits, often in the form of group- or colony-level fitness.

Definitions of punishment and policing

We use a broad definition of punishment and define ‘policing’ as a subset of ‘punishment’. We will henceforth use punishment as a general term and mention policing only when we specifically address that behavior.

Our definition of punishment is based on Clutton-Brock and Parker’s (1995a) original use, emphasizing Grotius’ “an ill suffered for an ill done”, with the punisher reaping long-term benefits (largely direct fitness benefits, but here considered to be inclusive fitness benefits to allow for possible effects towards or from interacting relatives). We thus define ‘punishment’ as ‘a behavior that inflicts a net cost on, or removes net benefits from, a target individual in response to a specific behavior by that target individual’. Under an inclusive fitness approach, we can make two assumptions about the evolution of punishment. First, the imposed costs or removed benefits should on average be high enough to make the target behavior unprofitable, and second, as we elaborate in the next section, punishment should only evolve when it provides a punisher with net inclusive fitness benefits.

Our definition of ‘policing’ is based on three significant uses in the literature – policing in eusocial insects (Ratnieks et al. 2006), policing in primates (Flack et al. 2005a, 2006), and general policing theory (Frank 1995, El Mouden et al. 2010), all of which concur on policing being a behavior creating public good benefits. These benefits can include reduced aggression and enhanced stability of relationships between group members (Flack et al. 2005b, 2006) and a reduction in reproductive conflict (Ratnieks et al. 2006). We therefore define ‘policing’ as ‘a punishing behavior that reduces average inclusive fitness losses of group members resulting from within-group competition for direct fitness gains’. Table 1

Table 1. The main intra-specific punishment and policing behaviors studied and discussed in the literature (first column), the defaulting behavior being targeted (i.e. made unprofitable) (second column), the public good created or served, the Figure in the present paper that represents this situation (third column), and relevant references. Items are marked with an asterisk (*) when the average fitness payoffs to group-mates are unclear; these scenarios therefore might not qualify as true public goods.

Punishment behavior	Target behavior	Public good created	Figure	Citation
Bullying of subordinates by dominant breeder in naked mole rats and meerkats	subordinate inactivity/ reproduction	increased help by subordinates, reduced reproductive conflict	3A	Reeve 1992, Young et al. 2006
Conflict intervention by dominants in pig-tailed macaques	inter-individual aggression and dominance assertion	reduced aggression, increased reconciliation	3A	Flack et al. 2005a, b
Aggression by other subordinates in cooperatively-breeding cichlids	forsaking helping duties	increased help by subordinates	3B	Balshine-Earn et al. 1998
Aggression by low-ranking workers in ponerine ants	challenging the established dominant	reduced reproductive conflict	4A	Monnin et al. 2002
Worker aggression by sisters in honeybees	worker egg-laying	reduced reproductive conflict	4B	Visscher and Dukas 1995
Selfish egg-eating and egg-replacement in paper wasps, yellowjackets, and bumblebees	worker egg-laying	reduced reproductive conflict*	5	Saigo and Tsuchida 2004, Wenseleers et al. 2005, Zanette et al. 2012
Apoptosis by cancer cells in metazoan bodies	cancerous proliferation of somatic cells	cancer prevented	6A	Lowe and Lin 2000
Conflict intervention in human societies	inter-individual aggression and bullying	reduced intra-group conflict	6B	von Rueden and Curven 2012
Monetary punishment in laboratory economics games with humans	monetary selfishness	donations to public pool	6C	Fehr and Gächter 2002

presents different potential policing behaviors across the social domains and the public goods they create.

Our policing definition encompasses behaviors like coercion of subordinates through bullying or harassment (Reeve 1992), egg-eating (including selfish policing: Wenseleers et al. 2005), and forms of group-sanctioning in human societies (Boehm 1999). It includes impartial third-party intervention (Flack et al. 2006, von Rohr et al. 2012) and some instances of partial intervention, such as when punishment resolves conflicts and reduces future disputes (e.g. nepotistic coalitions by high-ranking hyena females: Smith et al. 2010). Our policing definition excludes punishing behaviors that disrupt group stability, like some instances of retaliatory or coalitionary violence, because these behaviors do not benefit average inclusive fitness within groups and can sometimes be subject to policing themselves (Ren et al. 1991, Watts 1997). Although our definitions encompass interspecific interactions (e.g. mutualistic sanctioning: Kiers et al. 2003, Jandér and Herre 2010), we will not discuss them here. Finally, we realize that many forms of punishment and policing can be defined as forms of short-term spiteful behavior (harm oneself to harm others to which the actor is less than average related), but we avoid the term as it might induce semantic confusion and would not make a difference for inclusive fitness.

The punisher/bystander approach: heuristic principles

We consider two effects on inclusive fitness: punisher effects (p_t) and bystander effects (b_t). We define punisher effects (p_t) as the change in an individual's inclusive fitness when (s)he actively punishes, while bystander effects (b_t) represent an individual's incurred change in inclusive fitness when a group-mate punishes (see Fig. 1 for explicit definitions of p_t and b_t). These fitness payoffs do not need to be realized immediately – costs and benefits typically emerge over a time scale longer than the act of punishing. Note also that punisher and bystander effects are not independent: the two in fact typically co-vary, because an individual who benefits from punishing a specific target behavior will also tend to benefit when group-mates punish that behavior. Moreover, the difference between an individual's punisher and bystander effects ($p_t - b_t$) represents the inclusive fitness payoff for actively punishing, the key variable in our approach.

We expect individuals to punish when this both leads to inclusive fitness benefits and is more profitable than bystanding (i.e. letting someone else punish):

$$p_t > 0 \\ p_t > b_t$$

By plotting the two fitness effects on separate axes, we can evaluate the values of p_t and b_t at which individuals benefit from actively punishing (Fig. 2). The horizontal line $p_t = 0$ separates individuals that profit from punishing (above the line) from those who suffer a net cost from taking on a punishing role (below the line). Individuals in the white zones (1 and 2) will always punish. Individuals in the blue regions (zones 3 and 4) should never punish, although those in zone 4 do prefer that a group-mate punishes (see below). Indi-

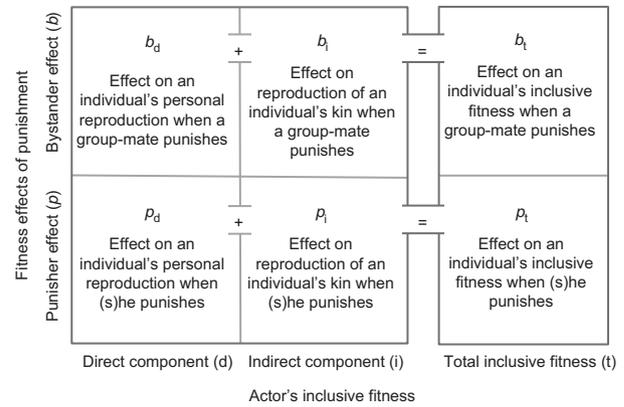


Figure 1. The relationship between the two components of inclusive fitness and the punisher (p_t) and bystander (b_t) effects. Inclusive fitness is the sum of direct fitness (personal reproduction) and indirect fitness (reproduction of relatives). The contribution of the four squares is arbitrary as any combination of direct and indirect effects is possible. Punishment scenarios assume group-living but group members can vary in relatedness, from being clones ($r = 1$) or full-sibling offspring ($r = 0.5$) to unrelated ($r = 0$; indirect fitness component becomes zero). Regardless of whether an individual is the punisher or a bystander, punishment of a group-mate should lead to either a gain in inclusive fitness (because reproductive success of self and/or kin increases), or a loss (because punishing harms the direct or indirect fitness interests of self and/or kin), or no difference. An individual should punish when the net inclusive fitness effect of punishing (p_t) is positive and greater than the net inclusive fitness effects of bystanding (b_t). When p_t is positive but less than b_t , an individual may punish if no other individual has $p_t > b_t$. The same logic applies to policing, which we have defined as a subset of punishment.

viduals in the purple region (zone 5) should refrain from punishing, unless their group-mates are unable or unwilling to punish. Here we elaborate on each zone:

Zone 1: individuals punish because punishment is the most profitable decision ($p_t > b_t > 0$). Individuals in zone 1 generally gain direct benefits from punishing, such as increases in dominance rank. This zone also encompasses policing behaviors driven exclusively by indirect fitness interests that are aligned with those of the colony as a whole.

Zone 2: individuals will punish ($p_t > 0$), but usually to the detriment of group-mates because bystanding effects are negative ($b_t < 0$). This implies that groups with all members in zone 2 can be characterized by recurrent harassment that harms group productivity (examples in Clutton-Brock and Parker 1995a). Examples of zone 2 behaviors include unpoliced sexual coercion (Clutton-Brock and Parker 1995b) and cycles of blood revenge in human societies (Beckerman et al. 2009). These behaviors are likely to be maintained by power asymmetries, which prevent low-ranking group-mates from policing behaviors that are costly to them. This zone also includes antisocial punishment as described by Herrmann et al. (2008), in which individuals, often low contributors themselves, pay a cost to discourage policing activities by other group members.

Zone 3: Individuals should never express punishment behaviors in this zone, because they reduce the actor's inclusive fitness, both when actively punishing and when bystanding ($p_t < 0$; $b_t < 0$). The antisocial punishment described

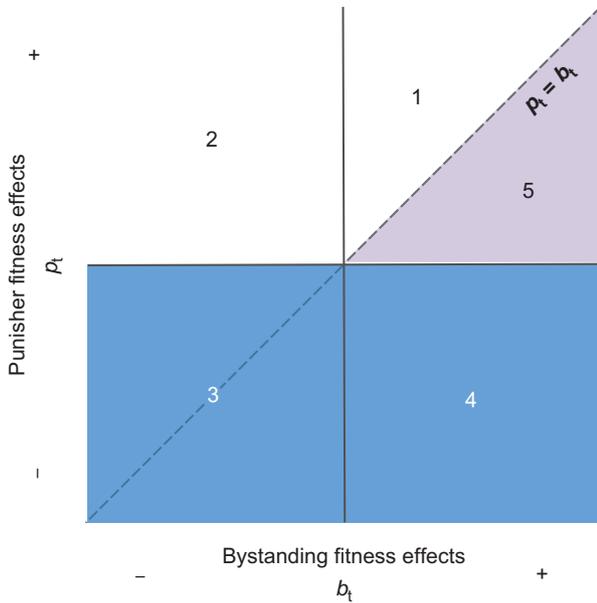


Figure 2. The costs (–) and benefits (+) of punishing and bystanding. Bystander fitness effects (b_t ; x-axis) refer to an individual's change in inclusive fitness when punishment occurs via the actions of one or more group members, whereas punisher effects (p_t ; y-axis) refer to changes in inclusive fitness when a focal individual expresses punishing behavior. Punishing and bystanding have the same effect on inclusive fitness on the dashed diagonal where $p_t = b_t$. Individuals in the white zones will punish, those in the blue zones will not, and those in the purple zone will punish only if group-mates are unwilling or unable. See text for definitions.

above does not appear to qualify, because it tends to be motivated by suspicion or revenge (Herrmann et al. 2008).

Zone 4: Individuals benefit from bystanding ($b_t > 0$) but suffer a net cost when punishing ($p_t < 0$). This zone can include subordinates in dominance hierarchies, queens in insect colonies, and other individuals who benefit from punishment but cannot afford the necessary costs to enact it because the net gains are insufficient. We expect individuals in this zone to ‘tattletale’ signal (analogous to children snitching on one another to a teacher) to punishers in zones 1 or 2 when a group-mate is defecting. Examples of tattletale signaling include calls to dominants in rhesus monkeys (Hauser 1992) and the tagging of defaulters in ponerine ants (Monnin et al. 2002).

Zone 5: Individuals benefit from punishing ($p_t > 0$) but benefit more when others do so ($b_t > p_t$), and should thus tattletale to competent punishers. However, if group-mates are unwilling or unable to punish, the bystander effects of these focal individuals become zero, shifting them horizontally to the central vertical axis (between zone 1 and 2) and motivating them to punish. Several outcomes are possible when multiple individuals are in zone 5: 1) A dominant may permit a subordinate to punish (allowing it to attain dominance over the defaulter with the dominants’ approval, which pushes subordinates into zone 1). 2) The cost of punishing relative to bystanding may be so marginal ($p_t - b_t \approx 0$) that it is more beneficial that the first individual to encounter a defaulter should punish irrespective of status. These individuals are on the line $p_t = b_t$. 3) Individuals may play a war of attrition game for others to punish, where the outcome will depend on

the rate at which p_t decreases (relative to b_t) with waiting time (similar to the Brave leader game: Shen et al. 2010).

Applications of the punisher/bystander approach

Vertebrate social breeders

The most straightforward means of developing punishment is through power asymmetries, because punishment is more profitable for individuals at the top of a dominance hierarchy. Theoretical models support this notion, finding that power or other asymmetries in access to resources motivate dominant individuals to take on punishing roles (Clutton-Brock and Parker 1995b, Frank 1996, Ruxton and van der Meer 1997, Flack et al. 2005a, Úbeda and Duéñez-Guzmán 2011). Our approach (Fig. 3A) captures this for groups where dominant breeders (D) hold considerably more power than subordinates (S). These breeders reap high benefits and pay very small costs when punishing, because ‘dove’-like subordinates will rarely retaliate against ‘hawk’-like dominants (Maynard Smith and Price 1973). Although dominant breeders often benefit from others taking on punishing roles (placing them in zone 5), they tend to be the only individuals for whom punishment is profitable. Their bystander benefits thus become zero, shifting them to the y-axis where we expect them to punish.

Studies of policing in vertebrate societies indeed find that dominants police, usually with near-zero costs. Flack et al. (2005a) found that in groups of pig-tailed macaques where power is highly-skewed, top-ranking individuals were much more likely to intervene in conflicts than lower-ranking ones, likely because punishment was too costly for subordinates. Punishment typically only involved the threat of aggression rather than overt physical violence, reducing the cost of policing to almost zero. Ren et al. (1991) observed similar dynamics in golden monkey ‘harems’ (i.e. single male, multi-female groups), where adult males intervened in almost 95% of the female-female conflicts. Of these interventions, 64% involved no aggression, illustrating once more how power asymmetries temper the costs of policing. Researchers have further noted punishment by dominant breeders in naked mole rats (Reeve 1992), superb fairy wrens (Mulder and Langmore 1993), gorillas (Watts 1997), meerkats (Young et al. 2006), fallow deer (Jennings et al. 2011), banded mongooses (Cant et al. 2010), and chimpanzees (Kahlenberg et al. 2008, von Rohr et al. 2012; but see Riedl et al. 2012).

When power is less skewed, subordinates may be strong enough to punish each other, especially when additional benefits, such as expected increases in dominance, incentivize punishment (Fig. 3B). This appears to be the case with policing in the cooperatively-breeding cichlid *Neolamprologus pulcher*. This fish lives in groups consisting of a breeding pair with several helpers, and while dominant breeders punish experimentally-induced ‘defectors’ in small groups, other helpers punished in large groups (Balshine-Earn et al. 1998, Fischer et al. 2014). The authors noted that aggressed defectors showed submissive behaviors towards their punishers, suggesting that punishment induces or helps to maintain dominance. However, additional studies in which hierarchical relationships are measured both before and after punishment would be helpful to better quantify the benefits

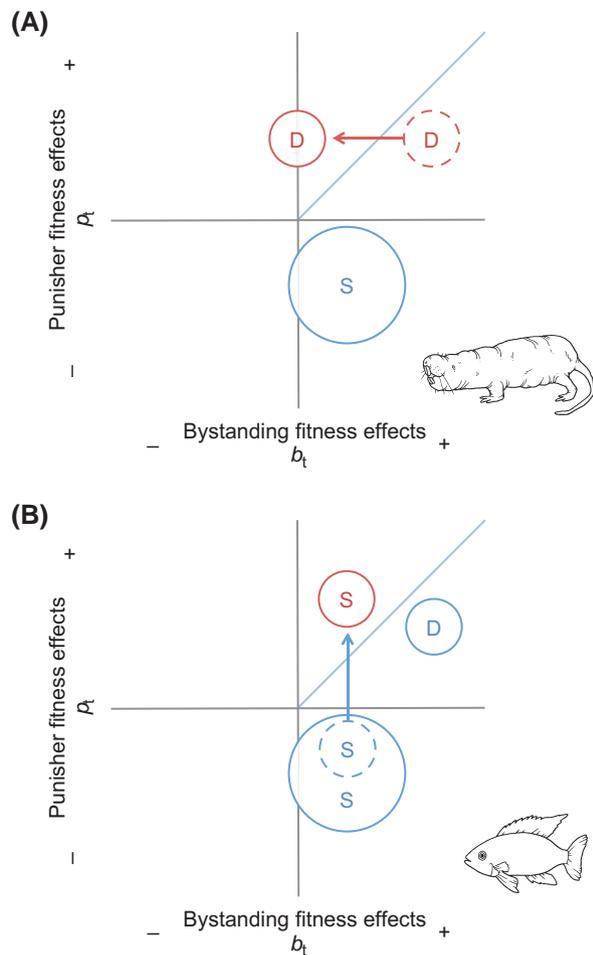


Figure 3. Punishment in non-human vertebrate social breeders. Axes are the same as those used in Fig. 2; red circles represent individuals expected to punish because p_i is both positive and $> b_i$, whereas blue circles are individuals that should not punish because these inclusive fitness conditions are not fulfilled. (A) Punishing a subordinate defector in a cooperative breeder group with the alpha breeder (D) stably dominant over a group of subordinates (S). D obtains net benefits from punishing while S normally pay higher costs when punishing each other and derive at best low bystander benefits. Because D is the only individual to benefit from punishing, his/her bystander benefits become zero, shifting D horizontally towards the y-axis. (B) When punishing is likely to increase their dominance, subordinates may collect punisher benefits, shifting a subset of such individuals (dashed circle) upwards along the vertical axis (to the red S).

to individual status. A subset of subordinates, rather than breeding dominants, also use punishment to stabilize mating hierarchies in the coral-dwelling goby *Paragobiodon xanthosomus* (Wong et al. 2007). Higher-ranked females queuing for mating status use threats of eviction to control any rise in status of their immediate subordinates, preventing those individuals from moving up in the hierarchy while also controlling reproductive conflict. Aggressive punishment also appears to re-exert dominance in some insect societies, as we discuss below.

Eusocial Hymenoptera

Punishment in the eusocial Hymenoptera is typically aimed at the products and perpetrators of a single behavior: viable

egg production. Many female workers, though unable to mate, retain functional ovaries and lay unfertilized, male-destined eggs that compete for nursing attention with the queen's eggs. Studies have documented aggression towards workers with developed ovaries, as well as towards the eggs produced by such workers, across the eusocial Hymenoptera (Ratnieks et al. 2006, van Zweden et al. 2007). Our coverage of the eusocial insects focuses on the Hymenoptera.

It is important to note two questions when considering punishment in eusocial insects: do colonies have an appreciable cohort of fully sterile workers (unable to lay any kind of viable egg during the time-window relevant for punishment/policing), and do they have dominance hierarchies? Because they lack direct fitness opportunities, fully sterile workers are driven solely by indirect fitness, so their interests tend to align completely with the reproductive interests of the entire colony. A substantial cohort of sterile workers leads to what we call 'organismal' policing arrangements, as seems to occur in honeybee and ponerine ant colonies. The size of a cohort of sterile workers can range from a subset of individuals to nearly all workers, and their ability to exert colony-level power depends on their relative numbers. Workers that are sterile may not necessarily remain sterile under all circumstances (Woyciechowski and Kuszewska 2012), hence our explicit time-window specification above. The second question to consider is whether colonies contain dominance hierarchies. Dominance hierarchies are normally absent in lineages where castes are physically differentiated but often appear when caste has remained phenotypically plastic, where they are typically correlated with mating and breeding status. Such hierarchies tend to be analogous with those of vertebrate cooperative breeders and converge on similar punishment dynamics (e.g. halictid bees: Michener and Brothers 1974; paper wasps: West-Eberhard 1986). Hierarchies have re-emerged in some lineages with physically differentiated castes where they normally regulate the production of unfertilized male eggs, such as in bumblebee colonies, where dominant workers compete with their mother queen late in the season (Bloch and Hefetz 1999), or when polygyny evolved as a secondary elaboration and was followed by reduction of colony size (Gill and Hammond 2011). In special evolutionary derived cases, such as in some ponerine ants, hierarchies determine mating opportunities as well (Higashi et al. 1994, Gobin et al. 2001).

Below we use the punisher/bystander approach to understand three classes of worker punishment: aggression in dominance hierarchies, aggression in evolutionarily derived 'organismal' colonies, and selfish egg-eating. Although our examples do not exhaustively review policing in the eusocial Hymenoptera, they widely sample the diversity of colony sizes and levels of conflict.

1) Aggression in dominance hierarchies

A number of studies have shown that eusocial queens (Q) may share the production of the colony's males with one or several workers (W_D) who hierarchically dominate their high-ranking (W_D^*) and low-ranked sisters (W_S) (Cole 1981, Bourke 1988, Ichinose and Lenoir 2009). In some colonies, subordinate workers at the bottom of dominance hierarchies (W_S) are older individuals who lack future drone-producing prospects. We thus consider them to be

sterile non-reproductives motivated purely by indirect fitness gains (in terms of Fig. 1, $b_d = p_d = 0$; therefore, $b_t = b_i$ and $p_t = p_i$).

When highly-ranked subordinate workers with future reproductive prospects (W_D^*) develop their ovaries, they generate reproductive conflict in the colony. We can use our punisher/bystander approach to evaluate which individuals should punish these workers (Fig. 4A). Our approach predicts that the queen (Q) and other subordinates with reproductive prospects (W_D^*) should not punish, because both would likely pay direct fitness costs from aggression without collecting appreciable benefits. Assuming the overall proportion of worker sons that a colony can raise is fixed, queen reproduction should remain unaffected by which workers lay most eggs, because the conflict occurs between workers for their mutually contested share of reproduction. The punishment payoffs to highly-ranked subordinate workers (W_D^*) are marginal, because they are related equally to a defaulting low ranking worker and the current dominant worker(s) as long as all workers are full-siblings. Benefits from punishment should therefore relate almost solely to increased colony efficiency for W_D^* individuals.

The dominant workers (W_D) and the sterile non-reproductives (W_S), on the other hand, both benefit from punishing an egg-laying worker, but need to sort out who punishes. Dominant workers gain direct fitness benefits, because punishment allows them to continue producing males, while sterile workers are motivated by indirect fitness benefits, and should act to maximize colony efficiency. Dominant workers should thus aggress competitors when punisher benefits exceed bystander benefits (e.g. punishment re-exerts dominance over a specific individual). Non-reproductive workers (W_S) should aggress whenever dominants do not, but to such a complementary degree that we expect subordinate non-reproductives and dominants to be mutually exclusive in their punishment.

Studies investigating ‘who polices’ support our prediction of mutual exclusivity in aggressive policing in these types of hierarchical colonies. Hartmann et al. (2003) found that, in the thelytokous ant species *Platythyrea punctata*, subordinate non-reproductives almost always aggressed new reproductives, while few established reproductives bothered doing so. Bourke (1988) explained how alpha workers, as opposed to unranked subordinates, attacked ovary-developing betas in the hierarchical slavemaker *Harpagoxenus sublaevis*. In some ponerine ant species, newly hatched workers can become mated reproductives (gamergates), though indirect fitness interests may motivate non-reproductive older workers to prevent their younger sisters from usurping a reproductive position (Higashi et al. 1994, Gobin et al. 2001). The non-reproductives’ behavior should change depending on who is currently heading the colony. When colonies are headed by the mother of the non-reproductives, the sterile workers should punish their ovary-developing sisters, because relatedness differences lead them to prefer that their mother, rather than sister, continue producing eggs. If a sister of the non-reproductives heads the colony, the workers are equally related to the defector and the current breeder, and punishment of ovulating defectors is maintained only because it reduces intra-colonial conflict. In both scenarios, these

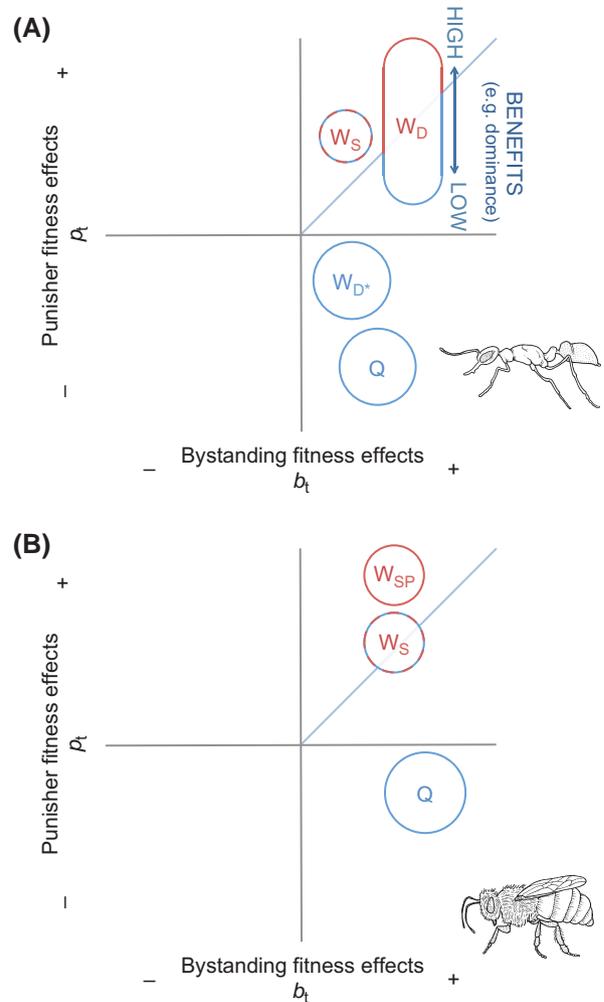


Figure 4. Punishment as aggression in eusocial insect colonies. Red and blue colors represent punishers and non-punishers as in Fig. 3, and bicolored ellipses are individuals whose response is conditional on social context (they should punish only when group-mates fail to do so because of lower or contingent inclusive fitness incentives). (A) Aggression against workers that have developed their ovaries in a eusocial insect colony in which a dominance hierarchy determines which worker(s) produce(s) the males. The queen (Q) will never punish: the conflict concerns worker reproduction, and the benefits of colony efficiency are unlikely to compensate for the potential costs of being injured by aggression. High ranking subordinate workers (W_D^*) should also refrain from punishing: in the time window considered, they benefit solely in terms of colony efficiency, so their bystander benefits are higher than their very low punisher benefits. Non-reproductive workers at the bottom of the hierarchy (W_S), who are equally related to defaulters and dominants, should be motivated by colony efficiency gains (i.e. the average indirect component of inclusive fitness). They should punish when dominant workers (W_D) are unable or unwilling to do so, although we expect dominants (rather than non-reproductives) to punish when aggression re-asserts status. (B) Aggression against workers with developed ovaries in an advanced (organismal) insect colony with a single multiply-mated queen (Q) and committed non-reproductive workers (W_S). Some of these are specialized for aggressive policing (W_{SP}), which reduces the costs of policing to the colony. The queen (Q) will never police because of the potentially high costs of aggression to her (high) fertility ($p_t < 0$).

workers should assist defectors if the current breeder is in poor condition. Thus, we expect these ponerine colonies to also exhibit policing exclusivity, with dominant gamergates rarely punishing because of the high costs involved and sterile workers carrying the burden of policing. Experiments and observations conform to this prediction – sterile workers in the Sulawesi ponerine ant *Gnamptogenys menadensis* attack virgin egg-layers, while gamergates never do (Gobin et al. 1999). Infertile workers in the Indian jumping ant *Harpegnathos saltator* “bite and jerk” ovulating workers for several days, while mated reproductives never participate in aggression (Liebig et al. 1999). In the queenless ant *Dinoponera quadricaps*, gamergates refrain from fighting and instead smear challengers with a chemical marker that attracts low-ranking subordinates to punish and immobilize the competitor (Monnin et al. 2002). This last example not only illustrates mutual exclusivity in policing, but also offers an elegant instance of tattletaling (the gamergate chemically signals to the workers that an individual is defaulting).

2) Aggression in ‘organismal’ colonies

We can also apply the punisher/bystander approach to evolutionarily-derived eusocial species with large colonies that have “organismal” properties (Queller and Strassmann 2009) to predict which individuals should aggress defecting workers. Often headed by a single multiply-mated queen, these colonies exhibit levels of cooperation so high, and conflict so low, that they rival the integration of metazoan bodies (Seeley 1989, Queller and Strassmann 2009). Relatedness structure in these colonies – with most sisters more highly-related to the queen than to each other – selects for workers that always eat each other’s eggs in favor of the queen’s (Ratnieks 1988). The pressure of oophagy is therefore so pervasive that workers virtually never lay eggs in the presence of a well-functioning queen (Ratnieks 1993, Dijkstra et al. 2005, Kronauer et al. 2006), making them effectively sterile. As with the non-reproductives in Fig. 4A, the workers in organismal colonies are motivated purely by indirect fitness gains (in terms of Fig. 1, $b_d = p_d = 0$; therefore, $b_t = b_i$ and $p_t = p_i$).

In rare cases, workers in organismal colonies develop ovaries to begin producing male-destined eggs, prompting policing by colony-mates. The punisher/bystander approach allows us to predict which individuals should aggress these defecting workers (Fig. 4B). Queens (Q) should never punish, because the costs of aggressing a worker can jeopardize her well-being and future reproductive prospects. Non-reproductive workers (W_c) should police, because they have fitness interests aligning completely with those of the colony, and the costs that a colony suffers from a worker’s potential injury are slight compared to the overall benefits of reduced conflict. Moreover, we expect selection to utilize the already-present division of labor (Robinson 1992) to reduce the costs of policing by creating a specialized class of policing workers (W_{sp}). In honeybee colonies, workers rather than queens attack other workers with developed ovaries (Visscher and Dukas 1995), but the topic of a specialized subclass of policers remains uninvestigated in organismal colonies (although recent documentation of specialized policers in ponerine ants is encouraging; van Zweden et al. 2007). Workers in the leafcutter ant *Acromyrmex echinator* only remove eggs from

other workers, but do not aggress sister workers with developed ovaries (Dijkstra et al. 2010), seemingly violating our prediction. However, workers in this species also lay trophic eggs (Dijkstra et al. 2005), potentially making aggression towards ovary-developers too costly.

3) Selfish egg-eating

Workers and queens can police egg-layers by aggressing them, but another common tactic for stifling worker reproduction is egg-eating (oophagy). For societies with high levels of conflict, in which a singly-mated queen heads a small colony, workers are more closely related to each other than to the queen, leading researchers to predict that workers should prefer raising nephews ($r = 0.375$) to brothers ($r = 0.25$) (Sommeijer et al. 1999, Tóth et al. 2002, Brunner et al. 2005, Wenseleers and Ratnieks 2006). However, empirical studies have found that workers may favor queen sons and remove each other’s eggs in spite of their higher relatedness (Hammond and Keller 2004; but see also Wenseleers and Ratnieks 2006). Researchers studying egg-eating in vespine (yellowjacket) and polistine (paper) wasps have proposed that worker policing can be selfish, i.e. a worker consumes her sisters’ eggs to replace them with her own (Saigo and Tsuchida 2004, Wenseleers et al. 2005). The observation by Zanette et al. (2012) of selfish egg-eating in bumblebee colonies indicates that the altruistic policing of an ‘organismic’ sister lineage, the honeybees, may have developed from this form of policing.

Our punisher/bystander approach can also illustrate ‘selfish policing’ scenarios and suggest who should eat worker eggs in species characterized by small full-sibling colonies with only life-time unmated workers, such as bumblebees or vespine wasps (Fig. 5). We consider a range of possible degrees of queen control, from highly coercive (she can replace any eaten egg with her own) to low coercion (she can only replace the eggs she herself eats with her own eggs). When queens (Q) are highly coercive, drone-producing workers (W) should never eat each other’s eggs, because doing so simply replaces a sister’s egg with a queen’s, while queens should always police. When queens have little coercive power – i.e. they cannot replace all policed eggs with their own – workers (W) and queens (Q) should compete over egg-eating. We thus expect that colonies with less skewed power relationships will be more conflictual, consistent with previous theory (Ratnieks and Reeve 1992).

Possible elaborations

Multicellular animals

Cancer occurs when a single cell (a social subunit) mutates and expresses cell-division behavior opposing the interests of its clone-mates. Somatic cells in a metazoan body lack reproductive prospects, so their fitness interests should align fully with those of the organism, making transformation (the switch from regulated to cancerous growth) dangerous and costly (Crespi and Summers 2005, Frank 2007). The interests of other cells remain aligned with the organism’s survival and thus suffer from single cells becoming cancerous, a scenario of defaulting and correction with many similarities to what we outlined in Fig. 4B. As we show in Fig. 6A, there are several ways of policing that may counter outlaw cell lineages.

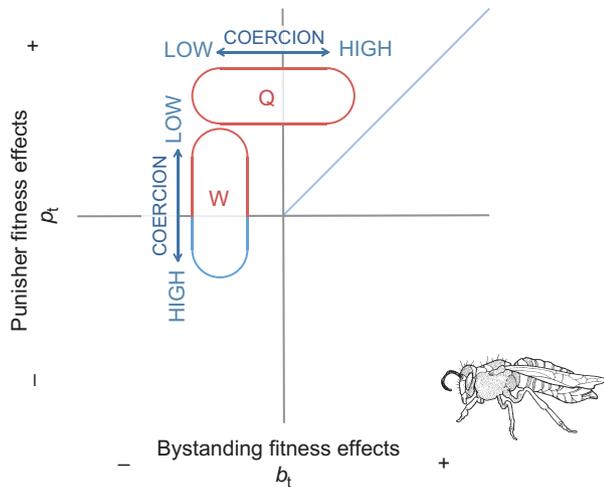


Figure 5. Egg-eating in small eusocial insect colonies with permanently unmated workers. Axes and colors are the same as in previous figures. Colonies are composed of a singly-mated queen (Q) and full-sibling workers with drone-productive potential (W). Workers lose inclusive fitness when sisters eat each other's eggs ($b_t < 0$), while the degree to which the queen benefits depend on her level of control (labeled COERCION on graph). Highly-coercive queens (those that can always replace an eaten egg with their own egg) collect high bystanding benefits, while leaving workers with low punishment effects, because workers pay the cost of egg-eating without reaping the reproductive benefit. Lowly-coercive queens cannot replace eaten eggs: they collect negative bystanding effects because egg-eating decrease colony efficiency, while worker punishment effects increase because workers can more consistently replace sisters' eggs. This may end with the queen and a subset of reproductive workers having similarly positive p_t values and similarly negative b_t values.

Because metazoan cells are clonal, self-policing by apoptosis (programmed cell death) is the most straightforward corrective mechanism (C_A). Cells express many genes that sacrifice their own fitness to curb the risk of transformation, the most notable being tumor suppressor genes, such as the RB (retinoblastoma-associated) protein. This 'gatekeeper' aids in integrating signals and deciding whether or not a cell should proceed in the growth-and-divide cycle (Burkhart and Sage 2008). Another self-policing tumor suppressor is protein 53 (p53), which can freeze a cell's progression through the division cycle if its genome suffers from excessive damage (or if the cell experiences other stresses, like a reduced nucleotide pool). If the damage to the genome goes unrepaired, p53 can take more drastic measures and trigger apoptosis (Sherr and McCormick 2002, Hanahan and Weinberg 2011).

In addition to self-policing, we might also expect cancer policing to involve specialized cell types (C_p), whose original somatic specialization pre-adapted them to adopt policing functions. Research in cancer immunology, which has recognized roles for both the innate and adaptive components of the immune system, appears to support this expectation. Tumor cells differentially express MICA/B proteins, and specific receptors expressed on natural killer and T cells seem designed to recognize these (Dunn et al. 2004). Moreover, CD8+ and CD4+ T cells respond to tumor antigens, and patients consistently develop specific antibodies for cancer cell surface antigens (Dunn et al. 2004, Swann and Smyth 2007).

Finally, we expect some somatic cell types in neighboring tissues (C_N) to have evolved mechanisms that signal (tattletale) the occurrence of transformation to a competent policing system. Whether this form of signaling exists is unclear, but several cellular phenomena suggest that cancer cells or their surroundings communicate transformation to the immune system. Researchers have found that early transformed cells and benign tumors express high levels of Fas-receptors, which, after binding with immune cell-produced ligands, trigger apoptosis (Nagata 1999, Owen-Schaub 2002). In another instance of cells advertising their transformation, dying cancer cells emit DAMPs (damage-associated molecular patterns) which help trigger the development of anti-tumor immunity (Garg et al. 2014). These danger signals allow the host immune system to identify the cells as threatening or otherwise alien. Interestingly, the emission of DAMPs shows how defectors signal their own defection to competent policers, illustrating an intersection of self-policing, specialization, and tattletale signaling.

Siphonophores, such as the well-known Portuguese man o' war, represent an unparalleled form of coloniality, compelling researchers to characterize them as an independent transition to "organismality" (Dunn 2009, Queller and Strassmann 2009). A single colony consists of many multicellular, genetically-identical subunits called zooids, which exhibit specialization analogous to the tissue division of labor of metazoan bodies and caste-differentiation in advanced eusocial insect colonies. Although siphonophore biology and development remain enigmatic, Dunn (2009) suggested that individual subunits might sometimes suffer from somatic mutations causing them to proliferate at the expense of the colony's survival. If this form of 'colony cancer' were confirmed, we expect the emergence of policing mechanisms similar to the ones outlined for metazoans and advanced eusocial insect colonies (i.e. self-policing, specialized policing and tattletale signaling).

Human societies

The deep causes of human punishment have been contested, not only because of the social relevance of such control mechanisms, but also because humans in lab environments sometimes appear to pay irrationally high costs to punish defectors, fueling arguments over the importance of cultural and multilevel selection in the evolution of human social behavior (Fehr et al. 2002, Boyd et al. 2003). While doing justice to this controversy exceeds the scope of this article, our punisher/bystander approach allows us to make two points. The first is that humans often punish more severely when power asymmetries lower the costs and increase the benefits, paralleling punishment dynamics in vertebrate and invertebrate social breeders. The second concerns the consideration of humans' ancestral social environment when interpreting the findings of laboratory economics games.

To elaborate the first point, individuals vary in resources and influence within societies, so power asymmetries may adaptively motivate some individuals to police more than others (Ruttan and Mulder Borgerhoff 1999, Hooper et al. 2010). Ethnographic studies describe policing structures that adhere to these expectations. In the Tsimane horticulturists, men with the highest (top 10%) political influence (a measure of hierarchy) adjudicated 66% of the conflicts that

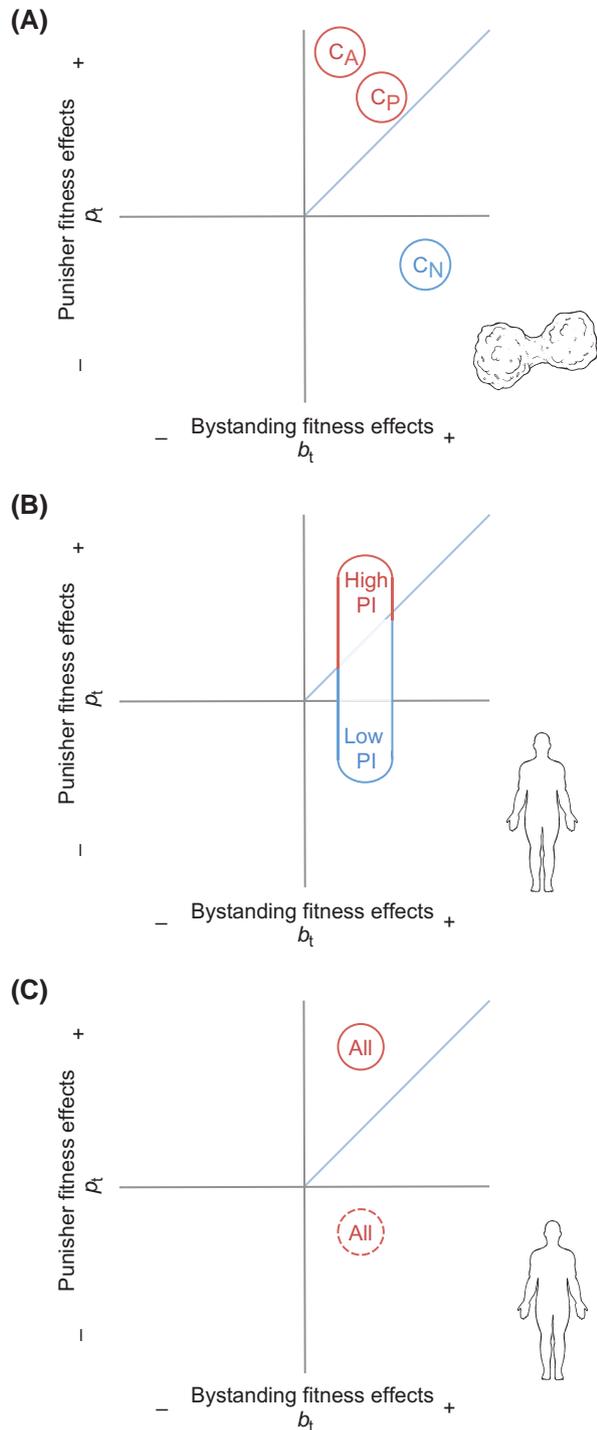


Figure 6. Possible extensions to cancer policing in metazoan bodies and conflict regulation in human societies. Red and blue colors represent punishers and non-punishers as in previous figures. (A) Cancer policing in a metazoan body. Normally, no cells should transform (become cancerous) because doing so would endanger the body and the inclusive fitness interests of every somatic cell, selecting for self-sacrificial policing by apoptosis (C_A ; the cheapest form of cancer policing). As a somatic division of labor between cell types and tissues is also generally found, specialized policer cells (C_P , usually the immune system) are expected to have evolved as a second line of defense. Cells neighboring cancerous tissue (C_N) should typically lack the power to successfully police cancerous cells, because the somatic division of labor has irreversibly committed them to other functions. However, these cells might have

emerged (von Rueden and Gurven 2012) (Fig. 6B). The researchers point out that punishment can carry the risk of retaliation in this society, but because individuals with political influence have large coalitions that deter retaliation, they may pay the lowest costs to control conflict. Researchers also observed that conflict-mediators were more likely to reap certain benefits, such as gaining support during conflicts, enjoying help during illness, having more surviving offspring, and accruing prestige, though these benefits might simply stem from having larger coalitions in the first place (von Rueden 2011, von Rueden and Gurven 2012). Ethnographers have observed similar social dynamics in societies in which authorities took on punishing roles, such as the Kapauku (Pospisil 1958) and the Cheyenne (Llewellyn and Hoebel 1941). Benefits other than those relating to power asymmetries appear to incentivize punishment as well, such as promises of reciprocity (e.g. within the sweatshop groups of the Yurok: Benson 1989) and enhanced reputation (e.g. “the brave, well-known warriors” of the Comanche: Hoebel 1954).

To detail the second point, the ethnographic observations of human punishment described above indicate that motivations for human punishment may be driven by inclusive fitness, but some authors have interpreted the findings of economic games with human participants as contradicting evolutionary logic (Fehr et al. 2002). Counter to rational agent models, humans consistently pay small costs to punish defectors in laboratory environments, prompting some authors to propose cultural and genetic multilevel selection arguments to conceptualize truly altruistic human social behavior (i.e. many individuals being in zone 4 but as a red rather than a blue circle; Fig. 6C) (Fehr et al. 2002, Henrich 2004). Others have argued that such responses do not appropriately represent the ancestral fitness payoffs of cheap punishment (Scott-Phillips et al. 2011), stressing that social dynamics like reputation, coordination and the maintenance of cooperative relationships all contribute towards making cheap punishment a beneficial behavior, especially in evolutionarily-relevant contexts (moving the red circle upwards into zone 1; Fig. 6C) (Dunbar 2004, Boyd et al. 2010, Krasnow et al. 2012, Roberts 2013). An especially striking observation is that humans appear to spontaneously police each other when the costs are very low: Wiessner (2005), for example, noted that single individuals willingly used public criticism and various forms of teasing to enforce cooperative norms among the Ju/'Hoansi bushmen. As punishing becomes more expensive, resource asymmetries and

evolved the ability to tattletale-signal to specialized policer cells (C_P) to increase their likelihood of policing. (B) Plotting human punishment as a gradient along which individuals differ in political influence (PI); those with high influence pay the lowest costs to punish (diminished chance of retaliation) and are thus motivated to take on punishing roles. (C) A summary of two alternative approaches to cheap punishment among humans. Some proponents of cultural/genetic group selection maintain that punishment is costly in inclusive fitness terms for the punisher (dashed circle), while others arguing from inclusive fitness principles maintain that our current punishment psychology must have been adaptive under the conditions in which these complex behaviors evolved (solid circle).

coordination prove effective in motivating individuals to act (Boehm 1993, Hooper et al. 2010). These points emphasize that experimenters should reflect on the conditions under which our behaviors evolved when interpreting the results of one-shot economic games.

Conclusions

Our punisher/bystander approach uses the assumption of inclusive fitness maximization to predict which individuals should punish in social groups. Although it oversimplifies many of the specific ecological details and domain-specific constraints that shape social arrangements, we believe that this approach recognizes basic principles of punishment and policing strategies across the social domains (Fig. 1, 2). Our approach makes explicit that across the domains of social evolution, 1) some individuals reap higher benefits or pay lower costs to punish, 2) this variation is important for the evolution of punishment and policing systems, and 3) the same inclusive fitness principles may produce different outcomes in each of these domains. Incorporating individual variation in costs and benefits is fundamental for understanding other public goods and coordinated group behaviors as well, such as leadership in fish schools (Krause et al. 2000) and participation in between-group aggression (Chagnon 1988, Glowacki and Wrangham 2013, 2015).

Applying a punisher/bystander approach illustrates the ways in which the dimensions of social systems – such as the distributions of power, relatedness, and reproductive options – may drive the variation in which individuals punish or police. Although both direct and indirect fitness benefits motivate individuals to punish, the two classes of fitness vary in their importance across social systems. In groups where individuals lack their own reproductive prospects, such as in advanced eusocial insect colonies and metazoan bodies, indirect fitness interests are often the exclusive motivators for policing (Fig. 4B, 6A). We consider this type of policing to be corrective, because defecting is typically at the expense of the social group as a whole and not in the interest of the individual defector. We predict these individuals both to police themselves and to freely punish others. However, in social systems where many individuals retain reproductive totipotency (e.g. vertebrate cooperative breeders and non-organismal insect societies), policing will tend to be coercive, involving dominant breeders (Fig. 3) or egg-eating selfish workers (Fig. 5) inflicting costs on defaulters to deter their selfish behaviors while obtaining direct fitness benefits (Clutton-Brock 2009, Zanette et al. 2012).

In eusocial breeding systems in which caste roles can change, both coercion (direct fitness benefits) and correction (indirect fitness benefits) may drive individuals to police, and the concurrence of both may lead to forms of mutual exclusivity between policing roles of dominants and subordinates (Fig. 4A). However, while coercion is possible in all situations where dominance hierarchies apply, correction is only possible when a large subset of workers lacks future reproductive prospects in the presence of a queen. The total spectrum of policing arrangements thus appears to range from bottom-up correction in organismal colonies of honeybees and leaf-cutting ants, via selfish policing in bumblebees and complementary policing in ponerine ants, to top-down

coercion in vertebrate social groups. In both eusociality (Boomsma 2013) and multicellularity (Fisher et al. 2013), the levels of conflict and cooperation in a social group depend on the reproductive options of social sub-units as well as the relatedness among them, and these dynamics appear critical for understanding the evolution of punishment and policing across these domains.

Acknowledgements – We thank Christopher Boehm, Andrew Bourke, Casey Dunn, Uli Ernst, Luke Glowacki, Luke Holman, Sandeep Nayak, Christopher von Rueden and Andrew Yegian for accommodating our questions and offering helpful comments. The Danish-American Fulbright Foundation provided financial support.

References

- Balshine-Earn, S. et al. 1998. Paying to stay or paying to breed? Field evidence for direct benefits of helping behavior in a cooperatively breeding fish. – *Behav. Ecol.* 9: 432–438.
- Beckerman, S. et al. 2009. Life histories, blood revenge, and reproductive success among the Waorani of Ecuador. – *Proc. Natl Acad. Sci. USA* 106: 8134–8139.
- Benson, B. 1989. Enforcement of private property rights in primitive societies: law without government. – *J. Libert. Stud.* 9: 1–26.
- Bloch, G. and Hefetz, A. 1999. Regulation of reproduction by dominant workers in bumblebee (*Bombus terrestris*) queenright colonies. – *Behav. Ecol. Sociobiol.* 45: 125–135.
- Boehm, C. 1993. Egalitarian behavior and reverse dominance hierarchy. – *Curr. Anthropol.* 34: 227–254.
- Boehm, C. 1999. Hierarchy in the forest. – Harvard Univ. Press.
- Boomsma, J. 2013. Beyond promiscuity: mate-choice commitments in social breeding. – *Phil. Trans. R. Soc. B.* 368: 20120050.
- Bourke, A. F. G. 1988. Dominance orders, worker reproduction and queen-worker conflict in the slave-making ant *Harpagoxenus sublaevis*. – *Behav. Ecol. Sociobiol.* 23: 323–333.
- Boyd, R. et al. 2003. The evolution of altruistic punishment. – *Proc. Natl Acad. Sci. US.* 100: 3531–3535.
- Boyd, R. et al. 2010. Coordinated punishment of defectors sustains cooperation and can proliferate when rare. – *Science* 328: 617–620.
- Brunner, E. et al. 2005. Reproductive conflict in social insects: male production by workers in a slave-making ant. – *Evolution* 59: 2480–2482.
- Burkhardt, D. L. and Sage, J. 2008. Cellular mechanisms of tumour suppression by the retinoblastoma gene. – *Nat. Rev. Cancer* 8: 671–682.
- Cant, M. et al. 2010. Reproductive control via eviction (but not the threat of eviction) in banded mongooses. – *Proc. Biol. Sci.* 277: 2219–2226.
- Chagnon, N. 1988. Life histories, blood revenge, and warfare in a tribal population. – *Science* 239: 985–992.
- Chaudhuri, A. 2010. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. – *Exp. Econ.* 14: 47–83.
- Clutton-Brock, T. H. 2009. Cooperation between non-kin in animal societies. – *Nature* 462: 51–7.
- Clutton-Brock, T. H. and Parker, G. A. 1995a. Punishment in animal societies. – *Nature* 373: 209–216.
- Clutton-Brock, T. H. and Parker, G. A. 1995b. Sexual coercion in animal societies. – *Anim. Behav.* 49: 1345–1365.
- Cole, B. J. 1981. Dominance hierarchies in *Leptothorax* ants. – *Science* 212: 83–84.
- Crespi, B. and Summers, K. 2005. Evolutionary biology of cancer. – *Trends Ecol. Evol.* 20: 545–52.

- Dijkstra, M. B. et al. 2005. Self-restraint and sterility in workers of *Acromyrmex* and *Atta* leafcutter ants. – *Insectes Soc.* 52: 67–76.
- Dijkstra, M. B. et al. 2010. Workers of *Acromyrmex echinator* leafcutter ants police worker-laid eggs, but not reproductive workers. – *Anim. Behav.* 80: 487–495.
- Dunbar, R. I. M. 2004. Gossip in evolutionary perspective. – *Rev. Gen. Psychol.* 8: 100–110.
- Dunn, C. W. 2009. Siphonophores. – *Curr. Biol.* 19: R233–R234.
- Dunn, G. P. et al. 2004. The three Es of cancer immunoediting. – *Annu. Rev. Immunol.* 22: 329–360.
- El Mouden, C. et al. 2010. The enforcement of cooperation by policing. – *Evolution* 64: 2139–2152.
- Fehr, E. and Gächter, S. 2002. Altruistic punishment in humans. – *Nature* 415: 137–140.
- Fehr, E. et al. 2002. Strong reciprocity, human cooperation, and the enforcement of social norms. – *Hum. Nat.* 13: 1–25.
- Fischer, S. et al. 2014. Group-size-dependent punishment of idle subordinates in a cooperative breeder where helpers pay to stay. – *Proc. R. Soc. B* 281: 20140184.
- Fisher, R. M. et al. 2013. Group formation, relatedness and the evolution of multicellularity. – *Curr. Biol.* 23: 1120–1125.
- Flack, J. C. et al. 2005a. Social structure, robustness and policing cost in a cognitively sophisticated species. – *Am. Nat.* 165: E126–139.
- Flack, J. C. et al. 2005b. Robustness mechanisms in primate societies: a perturbation study. – *Proc. Biol. Sci.* 272: 1091–1099.
- Flack, J. C. et al. 2006. Policing stabilizes construction of social niches in primates. – *Nature* 439: 426–429.
- Frank, S. A. 1995. Mutual policing and repression of competition in the evolution of cooperative groups. – *Nature* 377: 520–522.
- Frank, S. 1996. Policing and group cohesion when resources vary. – *Anim. Behav.* 52: 1163–1169.
- Frank, S. A. 2007. Dynamics of cancer: incidence, inheritance and evolution. – Princeton Univ. Press.
- Garg, A. D. et al. 2014. Danger signalling during cancer cell death: origins, plasticity and regulation. – *Cell Death Differ.* 21: 26–38.
- Gill, R. J. and Hammond, R. L. 2011. Workers determine queen inheritance of reproduction in a functionally monogynous ant population. – *Anim. Behav.* 82: 119–129.
- Glowacki, L. and Wrangham, R. W. 2013. The role of rewards in motivating participation in simple warfare. – *Hum. Nat.* 24: 444–460.
- Glowacki, L. and Wrangham, R. W. 2015. Warfare and reproductive success in a tribal population. – *Proc. Natl Acad. Sci. USA* 112: 348–353.
- Gobin, B. et al. 1999. Policing behaviour towards virgin egg layers in a polygynous ponerine ant. – *Anim. Behav.* 58: 1117–1122.
- Gobin, B. et al. 2001. Dominance interactions regulate worker mating in the polygynous ponerine ant *Gnamptogenys menadensis*. – *Ethology* 107: 495–508.
- Grafen, A. 2006. Optimization of inclusive fitness. – *J. Theor. Biol.* 238: 541–563.
- Hammond, R. L. and Keller, L. 2004. Conflict over male parentage in social insects. – *PLoS Biol.* 2: 1472–1482.
- Hanahan, D. and Weinberg, R. A. 2011. Hallmarks of cancer: the next generation. – *Cell* 144: 646–674.
- Hartmann, A. et al. 2003. Worker policing without genetic conflicts in a clonal ant. – *Proc. Natl Acad. Sci. USA* 100: 12836–12840.
- Hauser, M. D. 1992. Costs of deception: cheaters are punished in rhesus monkeys (*Macaca mulatta*). – *Proc. Natl Acad. Sci. USA* 89: 12137–12139.
- Henrich, J. 2004. Cultural group selection, coevolutionary processes and large-scale cooperation. – *J. Econ. Behav. Organ.* 53: 3–35.
- Herrmann, B. et al. 2008. Antisocial punishment across societies. – *Science* 319: 1362–1367.
- Higashi, S. et al. 1994. Worker's age regulates the linear dominance hierarchy in the queenless ponerine ant, *Pachycondyla sublaevis* (Hymenoptera: Formicidae). – *Anim. Behav.* 47: 179–184.
- Hoebel, E. A. 1954. The law of primitive man: a study in comparative legal dynamics. – Harvard Univ. Press.
- Hooper, P. L. et al. 2010. A theory of leadership in human cooperative groups. – *J. Theor. Biol.* 265: 633–646.
- Hurst, L. D. et al. 1996. Genetic conflicts. – *Q. Rev. Biol.* 71: 317–364.
- Ichinose, K. and Lenoir, A. 2009. Reproductive conflict between laying workers in the ant *Aphaenogaster senilis*. – *J. Ethol.* 27: 475–481.
- Jandér, K. C. and Herre, E. A. 2010. Host sanctions and pollinator cheating in the fig tree-fig wasp mutualism. – *Proc. R. Soc. B* 277: 1481–1488.
- Jennings, D. et al. 2011. Third-party intervention behaviour during fallow deer fights: the role of dominance, age, fighting and body size. – *Anim. Behav.* 81: 1217–1222.
- Jensen, K. 2010. Punishment and spite, the dark side of cooperation. – *Phil. Trans. R. Soc. B* 365: 2635–2650.
- Kahlenberg, S. M. et al. 2008. Immigration costs for female chimpanzees and male protection as an immigrant counter-strategy to intrasexual aggression. – *Anim. Behav.* 76: 1497–1509.
- Kiers, E. T. et al. 2003. Host sanctions and the legume–rhizobium mutualism. – *Nature* 425: 78–81.
- Krasnow, M. M. et al. 2012. What are punishment and reputation for? – *PLoS ONE* 7: e45662.
- Krause, J. et al. 2000. Leadership in fish shoals. – *Fish Fish.* 1: 82–89.
- Kronauer, D. J. C. et al. 2006. Male parentage in army ants. – *Mol. Ecol.* 15: 1147–1151.
- Liebig, J. et al. 1999. Worker policing limits the number of reproductives in a ponerine ant. – *Proc. R. Soc. B* 266: 1865–1870.
- Llewellyn, K. N. and Hoebel, E. A. 1941. The Cheyenne way: conflict and case law in primitive jurisprudence. – Univ. of Oklahoma Press.
- Lowe, S. and Lin, A. 2000. Apoptosis in cancer. – *Carcinogenesis* 21: 485–495.
- Maynard Smith, J. and Price, G. R. 1973. The logic of animal conflict. – *Nature* 246: 15–18.
- Michener, C. and Brothers, D. 1974. Were workers of eusocial Hymenoptera initially altruistic or oppressed? – *Proc. Natl Acad. Sci. USA* 71: 671–674.
- Monnin, T. et al. 2002. Pretender punishment induced by chemical signalling in a queenless ant. – *Nature* 419: 3745–3748.
- Mulder, R. A. and Langmore, N. E. 1993. Dominant males punish helpers for temporary defection in superb fairy-wrens. – *Anim. Behav.* 45: 830–833.
- Nagata, S. 1999. Fas ligand-induced apoptosis. – *Annu. Rev. Genet.* 33: 29–55.
- Owen-Schaub, L. 2002. Fas function and tumor progression: use it and lose it. – *Cancer Cell* 2: 2001–2002.
- Pospisil, L. 1958. Kapauku Papuans and their law. – Dept of Anthropology, Yale Univ.
- Queller, D. C. and Strassmann, J. E. 2009. Beyond society: the evolution of organismality. – *Phil. Trans. R. Soc. B* 364: 3143–3155.
- Raihani, N. J. et al. 2012. Punishment and cooperation in nature. – *Trends Ecol. Evol.* 27: 288–295.

- Ratnieks, F. L. W. 1988. Reproductive harmony via mutual policing by workers in eusocial Hymenoptera. – *Am. Nat.* 132: 217–236.
- Ratnieks, F. L. W. 1993. Egg-laying, egg-removal, and ovary development by workers in queenright honey bee colonies. – *Behav. Ecol. Sociobiol.* 32: 191–198.
- Ratnieks, F. L. W. and Reeve, H. K. 1992. Conflict in single-queen hymenopteran societies: the structure of conflict and processes that reduce conflict in advanced eusocial species. – *J. Theor. Biol.* 158: 33–65.
- Ratnieks, F. L. W. et al. 2006. Conflict resolution in insect societies. – *Annu. Rev. Entomol.* 51: 581–608.
- Reeve, H. K. 1992. Queen activation of lazy workers in colonies of the eusocial naked mole-rat. – *Nature* 358: 147–149.
- Ren, R. et al. 1991. The reconciliation behavior of golden monkeys (*Rhinopithecus roxellanae roxellanae*) in small breeding groups. – *Primates* 32: 321–327.
- Riedl, K. et al. 2012. No third-party punishment in chimpanzees. – *Proc. Natl Acad. Sci. USA* 109: 14824–14829.
- Roberts, G. 2013. When punishment pays. – *PLoS ONE* 8: e57378.
- Robinson, G. E. 1992. Regulation of division of labor in insect societies. – *Annu. Rev. Entomol.* 37: 637–665.
- Ruttan, L. M. and Mulder Borgerhoff, M. 1999. Are East African pastoralists truly conservationists? – *Curr. Anthropol.* 40: 621–652.
- Ruxton, G. D. and van der Meer, J. 1997. Policing: it pays the strong to protect the weak. – *Trends Ecol. Evol.* 12: 9–10.
- Saigo, T. and Tsuchida, K. 2004. Queen and worker policing in monogynous and monandrous colonies of a primitively eusocial wasp. – *Proc. R. Soc. B* 271 Suppl: S509–512.
- Scott-Phillips, T. C. et al. 2011. Evolutionary theory and the ultimate-proximate distinction in the human behavioral sciences. – *Persp. Psychol. Sci.* 6: 38–47.
- Seeley, T. D. 1989. The honey bee colony as a superorganism. – *Am. Sci.* 77: 546–553.
- Shen, S.-F. et al. 2010. The brave leader game and the timing of altruism among nonkin. – *Am. Nat.* 176: 242–248.
- Sherr, C. J. and McCormick, F. 2002. The RB and p53 pathways in cancer. – *Cancer Cell* 2: 103–112.
- Smith, J. E. et al. 2010. Evolutionary forces favoring intragroup coalitions among spotted hyenas and other animals. – *Behav. Ecol.* 21: 284–303.
- Sommeijer, M. J. et al. 1999. Behavioural data on the production of males by workers in the stingless bee *Melipona favosa* (Apidae, Meliponinae). – *Insectes Soc.* 46: 92–93.
- Swann, J. and Smyth, M. 2007. Immune surveillance of tumors. – *J. Clin. Invest.* 117: 1137–1146.
- Tóth, E. et al. 2002. Genetic and behavioral conflict over male production between workers and queens in the stingless bee *Paratrigona subnuda*. – *Behav. Ecol. Sociobiol.* 53: 1–8.
- Úbeda, F. and Duéñez-Guzmán, E. A. 2011. Power and corruption. – *Evolution* 65: 1127–1139.
- van Zweden, J. S. et al. 2007. Specialization in policing behaviour among workers in the ant *Pachycondyla inversa*. – *Proc. Biol. Sci.* 274: 1421–1428.
- Visscher, P. and Dukas, R. 1995. Honey bees recognize development of nestmates' ovaries. – *Anim. Behav.* 49: 542–544.
- von Rohr, C. R. et al. 2012. Impartial third-party interventions in captive chimpanzees: a reflection of community concern. – *PLoS ONE* 7: e32494.
- von Rueden, C. R. 2011. The acquisition of social status by males in small-scale human societies (with an emphasis on the Tsimane of Bolivia). – PhD thesis. Univ. of California, Santa Barbara.
- von Rueden, C. R. and Gurven, M. 2012. When the strong punish: why net costs of punishment are often negligible. – *Behav. Brain Sci.* 35: 43–44.
- Watts, D. 1997. Agonistic interventions in wild mountain gorilla groups. – *Behaviour* 134: 23–57.
- Wenseleers, T. and Ratnieks, F. L. W. 2006. Comparative analysis of worker reproduction and policing in eusocial Hymenoptera supports relatedness theory. – *Am. Nat.* 168: E163–E179.
- Wenseleers, T. et al. 2005. Queen and worker policing in the tree wasp *Dolichovespula sylvestris*. – *Behav. Ecol. Sociobiol.* 58: 80–86.
- West, S. A. and Gardner, A. 2013. Adaptation and inclusive fitness. – *Curr. Biol.* 23: R577–584.
- West-Eberhard, M. J. 1986. Dominance relations in *Polistes canadensis* (L.), a tropical social wasp. – *Monit. Zool. Ital. J. Zool.* 20: 263–281.
- Wiessner, P. 2005. Norm enforcement among the Ju/'hoansi Bushmen. – *Hum. Nat.* 16: 115–145.
- Wong, M. Y. L. et al. 2007. The threat of punishment enforces peaceful cooperation and stabilizes queues in a coral-reef fish. – *Proc. Biol. Sci.* 274: 1093–1099.
- Woyciechowski, M. and Kuszewska, K. 2012. Swarming generates rebel workers in honeybees. – *Curr. Biol.* 22: 707–711.
- Young, A. J. et al. 2006. Stress and the suppression of subordinate reproduction in cooperatively breeding meerkats. – *Proc. Natl Acad. Sci. USA* 103: 12005–12010.
- Zanette, L. R. S. et al. 2012. Reproductive conflict in bumblebees and the evolution of worker policing. – *Evolution* 66: 3765–3777.