LSHTM Research Online

https://researchonline.lshtm.ac.uk

Enforcement is central to the evolution of cooperation

J. Arvid Ågren[1,2,*], Nicholas G. Davies[3*], and Kevin R. Foster[4,5]

1. Department of Molecular Biology and Genetics, Cornell University

2. Department of Organismic and Evolutionary Biology, Harvard University

arvid_agren@fas.harvard.edu

3. Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine

nicholas.davies@lshtm.ac.uk

4. Department of Zoology, University of Oxford

kevin.foster@zoo.ox.ac.uk

5. Department of Biochemistry, University of Oxford

* These authors contributed equally

Summary

Cooperation occurs at all levels of life, from genomes, complex cells and multicellular organisms to societies and mutualisms between species. A major question for evolutionary biology is what these diverse systems have in common. Here we review the full breadth of cooperative systems and find that they frequently rely on mechanisms that suppress selfish behaviour. We discuss many examples including the suppression of transposable elements, uniparental inheritance of mitochondria and plastids, anti-cancer mechanisms, reciprocation and punishment in humans and other vertebrates, policing in eusocial insects, and partner choice and manipulation in mutualisms between species. To address a lack of accompanying theory we develop a series of evolutionary models that show that the enforcement of cooperation is generally predicted. We argue that enforcement is an underappreciated, and often critical, ingredient for cooperation across all scales of biological organisation.

Editorial summary

This Review discusses cooperative systems across all biological levels, including genomes, multicellularity and societies, and develops mathematical models to argue that enforcement is central to the evolution of cooperation.

Introduction

The evolution of cooperation is central to all living systems. Evolutionary history can be defined by a series of major transitions in which replicating units came together, lost their independence, and formed new levels of biological organisation[1-4]. As a consequence, life is organized in a hierarchy of cooperation: genes work together in genomes, genomes in cells, cells in multicellular organisms, and multicellular organisms in eusocial groups (Figure 1). The identification of these major transitions underlined the key conceptual challenge for all cooperative systems: why does natural selection favour investment in cooperation, rather than self-serving rebellion that would undermine a particular genome, organism, or society?

Despite this generality, discussions of social evolution commonly focus on just one or two systems. Much attention has been paid to the amazing biology of the eusocial insects – bees, ants, wasps, and termites – and the evolution of worker sterility[5-8]. Eusocial insects are an example of what Queller[9] termed a *fraternal* cooperative system, multicellularity being another[10], where the interacting individuals are genetically-related members of one species or type (Figure 1; Table S3). A key feature of fraternal systems, and fraternal major transitions, is the importance of kin selection, which explains cooperative traits via their benefits to family members that pass on the same genes[3-6, 11] (Table S3).

There is a concern, however, that the focus on particular empirical systems might be limiting our understanding of cooperative evolution. In particular, the field of social evolution has paid less attention to what Queller called *egalitarian* cooperative systems, which are those formed between unrelated individuals [10]. These include cooperation among genes in genomes, the origin of complex cells via endosymbiosis, and between-species mutualisms (Figure 1; Table S3). Between egalitarian individuals, there can be no self-sacrifice for a brother, sister or clone-mate, rendering kin selection powerless[12]. Consideration of these systems makes it clear that family life is not a general explanation for the major transitions in evolution, or for cooperative systems in general[13].

A major open question then is what, if anything, unites the evolution of cooperative systems? Here, we review cooperative evolution across all levels of biological organisation, which reveals growing evidence for the importance of enforcement. By

enforcement, we mean an action that evolves, at least in part, to reduce selfish behaviour within a cooperative alliance (see Box 1 for formal definition). These mechanisms have long been considered as way to promote cooperation[14], particularly in genomes[15-17] and humans[18], and early experiments revealed the power of enforcement in the birds[19] and the bees[20]. By now the growing list of examples covers all biological scales with recent reviews highlighting, for example, the importance of silencing of transposable elements for genomes[21, 22], conflict mediation in mitochondrial-nuclear interactions for eukaryotic cells[23, 24], cancer suppression for multicellular evolution[25, 26], and host sanctions for symbiotic mutualisms[27].

Such diverse examples, however, have resulted in an equally diverse terminology, which can hinder comparison and synthesis. In this paper, we bring together these examples and use mathematical models to formalise the common evolutionary processes at work across the full diversity of cooperative systems. It is only by considering all scales of cooperation side by side – from genomes through complex cells, multicellular organisms, and societies to mutualism between species – that the reach of enforcement becomes clear.

Genomes

The earliest alliances were molecular. One of the major theories of early life – the RNA World hypothesis – centres around cooperating RNA molecules or related polymers[28]. Similarly, the hypercycle concept provided insights on the theoretical conditions under which cooperative networks may first evolve from replicators[1, 2]. The lack of empirical details of these ancient systems make it challenging to infer how they evolved, although mathematical models allow one to explore general principles [2] (Figures 2a and 3b). What is clear is that replicating molecules eventually became stably connected in DNA-based genomes. This connection likely reduced evolutionary conflict as all loci now shared an interest in their cell's fitness, and the notion of being "in the same boat" is often considered central to the evolution and maintenance of cooperation[29, 30]. However, this effect is much weaker in genomes than is often assumed. Even small amounts of horizontal gene transfer can favour selfish genetic elements that proliferate within the genome at the cost of the host genome[31, 32]. Sexual reproduction pushes this effect to the extreme. Because element can spread to new lineages by outcrossing each generation, they can proliferate

4

despite reducing host fitness [33]. The evolution and maintenance of cooperation within genomes therefore requires additional explanation.

Some selfish genetic elements may fail to spread due to constraints on their function, or be self-limiting[34]. For example, a meiotic drive gene can spread in diploids by killing copies of the chromosome that do not carry it at meiosis, but if the driver spreads to fixation the killing will stop[34]. However, like being in the same boat, it is clear that such effects are again insufficient to protect cooperation within genomes. This is most clear from transposable elements, which are particularly interesting because of their potential similarity to early replicators[21]. Transposons are found across all domains of life, harm their hosts [35], and make up the bulk of many eukaryotic genomes, including around half of the human genome.

While the shared genomic interest that comes with being in the same boat is not expected to remove genomic conflict, it can provide the basis for the evolution of enforcement. To extend the analogy, while being the same boat does not protect you from conflict it *does* provide a boat to protect (Equation S4 versus S6 in Supplement, Model 1). Empirical observations support these predictions. The *P* transposable element in *Drosophila* was first detected in *D. melanogaster* in the 1970s[36] because it causes hybrid dysgenesis in crosses with females lacking the element. Strikingly, American populations of *D. melanogaster* evolved silencing of the *P* element in less than 40 years via a small RNA pathway, consistent with an on-going co-evolutionary arms race between the host and the *P* element [37, 38]. The *P* element has recently entered the genome of *D. simulans* from *D. melanogaster,* where it induces both morphological abnormalities and sterility, but several strains already appear to have the ability to silence the transposition[39].

Small RNA-mediated silencing is also thought to limit transposon replication in many other species[40], including plants[41]. The plant *Arabis alpina* seems to have lost the ability to effectively control long terminal retrotransposons, such that transposable element-derived sequence in *A. alpina* is now larger than the entire genome of related species[42] (Figure 1). In addition to transposable elements, there are many other forms of genomic conflict[32, 43]. As discussed, some meiotic drivers can be self-limiting but, for many others, there is evidence that the evolution of suppressors is central to maintaining organismal function[44]. These conflicts are also easy to miss because selfish genetic

elements and their suppressors commonly co-occur in populations and it often takes crosses between divergent populations to see the elements' effects[21].

Considerable effort is currently underway to engineer selfish genetic elements – so-called gene drives – to introduce desirable traits into natural populations, such as reduced carriage of malaria or dengue by mosquitos[45]. By design, these elements are intended to invade a population as well as possible. It is telling then that a key concern surrounding these approaches comes from theoretical[46] and empirical work[47] demonstrating the ease with which genomic suppressor alleles can arise and spread. The best hope comes from strategies where suppressor mutations are somehow constrained or where the presence of gene drives has little effect on host fitness, i.e. cases where the "selfish" elements actually cost less than enforcement. When the elements compromise organismal function, the prediction is clear: suppressor alleles will rapidly inhibit their spread and effects[46].

Complex cells

The evolution of the eukaryotic cell is an example of an egalitarian alliance and transition (Figure 1; Table S3). The interaction between mitochondria, plastids and nuclear genes has been crucial to eukaryotic metabolism for over a billion years[48] and is arguably the most intimate and successful of all symbiotic relationships[49]. The long co-evolutionary association between mitochondrial and nuclear genes is clearly demonstrated by the large-scale transfer of genes from cytoplasmic to nuclear DNA[49]. The mitochondrial genomes of most animal species now only carry a dozen or so genes and in *Arabidopsis thaliana* close to one in five nuclear genes are of chloroplast origin[50].

As for the evolution of genomes, the close association of organellar and nuclear genomes reduces the scope for conflict but is insufficient to remove important differences in evolutionary interest[51]. The potential for conflict was probably at its strongest in early eukaryotes where the fusion of outbred gametes would bring together genetically-different endosymbionts (Figure 2b; Supplement). The strength of this potential conflict has led several authors to argue again for the importance of enforcement, specifically that the uniparental inheritance of organelles evolved to limit the spread of selfish endosymbionts[1, 14, 51-53]. While other factors may have contributed[54], our own

modelling further suggests that isogametic fusion both strongly favoured the evolution of selfish organelles that bias their transmission, and the subsequent evolution of uniparental inheritance as a mechanism of enforcement (Figures 2b and 3c; Model 2 in Supplement). Modern isogamous species all appear to show uniparental inheritance making it challenging to test such predictions[55]. However, the unicellular green alga *Chlamydomonas reinhardtii* is informative because it is isogamous and possesses both mitochondria and chloroplasts. Interestingly, uniparental inheritance differs for mitochondria and plastids in the timing and mechanism of elimination, and which mating type passes on which endosymbiont (minus type passes on mitochondria, plus type chloroplasts)[55]. This may indicate two distinct evolutionary conflicts where selfishness and later suppression evolved for each endosymbiont independently.

The evolution of uniparental inheritance did not end conflict within complex cells. Because uniparental organellar inheritance contrasts with the bi-parental inheritance of autosomal nuclear genes, the evolutionary fate of mitochondrial and chloroplast genes in anisogamous species depends only on their performance in females[56]. As a consequence, there is a long-documented conflict driven by mitochondrial genes over sex determination and sex ratio, and several mitochondrial genes in plants are known to cause male sterility[57]. Enforcement is again important for resolving this conflict, via the evolution of nuclear alleles that suppress selfish mitochondrial phenotypes and restore organismal functioning. As for selfish genetic elements (above), however, the magnitude of co-evolutionary arms races between the organelles and the nucleus is only revealed by crossing diverged populations. For example, cytoplasmic male sterility is rarely detected in natural populations of the monkey flower *Mimulus guttatus*, but is revealed in crosses with closely related *M. nasutus*[58, 59] (Figure 1).

In contrast with plants, the mitochondria of animals do not appear to cause major conflicts, possibly due to their greatly reduced genome (but see discussions of Mother's Curse and its control[60]). However, nuclear enforcement is required to suppress other endosymbiotic bacteria, particularly from the genus *Wolbachia* that are carried by many insects. While intracellular and transmitted vertically like mitochondria and chloroplasts, *Wolbachia* are typically classified as parasites because of the reproductive problems they cause in their attempts to bias investment into female offspring[61]. For example,

Wolbachia kill male eggs in a range of insect species including the Great Eggfly butterfly (*Hypolimnas bolina*) where killing can be so effective as to drive the sex ratio down to a few percent of males[62]. However, as for the *P* transposable element, male killing can be suppressed in a few generations by the spread of a nuclear allele that restores male viability[63].

Cell groups and multicellularity

There is extensive cooperation within cellular groups. The evolution of cooperation by microorganisms is now known to be extremely common and central for both their functioning and their effects on human health[64]. For example, many bacteria secrete a suite of extracellular factors that provide benefits for neighbouring cells, including enzymes that digest nutrients and antibiotics, and molecules that scavenge rare resources like iron[65]. The evolution of cooperation within microbial groups, perhaps more than in any other system, appears to rest upon genetic relatedness and the ability of binary fission to generate a large clonal patch that can share in cooperative benefits[66]. This alignment of interests reduces the need for enforcement. But it does not remove it. The rapid proliferation of microbes creates ample opportunity for the *de novo* evolution of cells that lack cooperative traits via mutation, often known as "cheaters"[67] (Figure 2). This may select for genetic mechanisms to counter the effects of such mutations and lead cells to enforce their own cooperation ('self-enforcement'; Table 2; Model 3 in Supplement). Key mechanisms include the evolution of pleiotropy that links selfish phenotypes to personal cost, and redundancy that limits the phenotypic penetrance of mutations[68, 69] (Box S1).

The importance of enforcement in cellular groups, however, is clearest within multicellular organisms. The transition to multicellular life from independently living cells has occurred many times throughout the history of life[70]. Origins range from 3.5 billion years ago in cyanobacteria, to 200 million years ago in the volvocine algae[71]. Across these transitions, kinship has played a key role with obligate multicellularity only evolving in taxa where group formations occur clonally and relatedness is maximal between cells[72]. However, mutations pose a major threat to many multicellular systems, particularly animals where cancer, the disease associated with cells dividing in an unregulated fashion, is prevalent[26, 73].

Animal cells have evolved a wide range of mechanisms that inhibit the rise and spread of cancer[25, 26, 73, 74]. This includes strong pleiotropic links from cell proliferation and DNA damage to apoptosis, which cause many precancerous cells to undergo programmed cell death[75, 76] (Box 2; Figures 2 and 3; Model 3 in Supplement). Such self-enforcement may explain why elephants, despite their massive body size, experience a lower rate of cancer-based mortality than humans. Compared to a single copy in humans, elephants have multiple copies of the tumour suppressor gene *TP53* and their cells undergo apoptosis at a much higher rate than human cells in response to DNA damage[77, 78]. Another important response seen in pre-cancerous cells is differentiation. For example, a wide range of mutations that promote tumour formation in mammalian blood stem cells also increase the chance of becoming a non-proliferative cell type. This helps to ensure that that mutant lineages are replaced by healthy cells in the dividing stem cell population rather than progressing towards tumour formation [73]. There is even evidence that clumps of cancerous cells can be physically pushed out of proliferative tissues via a process known as apical extrusion[79]. Finally, the immune system can act as an enforcement system that targets tumour cells for destruction[80, 81]. The transmissible facial tumour in the Tasmanian devil appears to spread because low MHC (Major Histocompatibility Complex) diversity in the devils limits foreign tissue rejection [82] (Figure 1). The tumours are a threat to the species but the devils seem to be evolving resistance, thereby restoring multicellular integrity to a most charismatic of examples[83].

Interestingly, plants and fungi appear to suffer fewer harmful cancers than animals for reasons currently unknown. Features like the cell wall, lack of cell motility, and the modularity of plants may limit the spread of malign tumours[26, 84]. Recent work, however, has found that cell proliferation and cell death are again positively linked in plants, via regulators such as RBR1[85], raising the possibility that self-enforcement mechanisms have convergently evolved in plants and animals.

Societies and eusociality

The field of social evolution has long revolved around the study of animal societies. This includes cooperative breeding in vertebrates, particularly birds and mammals, where there is again a clear role for genetic relatedness. Either high average relatedness or mechanisms

of kin discrimination (that allow more related individuals to interact) are seen across cooperative breeders[86]. However, there is also evidence that enforcement is central to cooperation[76, 87], although a convincing demonstration can require challenging manipulative experiments[88]. For example, superb fairy wren helpers are punished by breeding males if they are experimentally prevented from helping[19], and if a subordinate banded mongoose is made to reproduce without the dominants, its offspring are killed [89].

The importance of enforcement in vertebrates arguably reaches its zenith in humans, where egalitarian cooperation between non-relatives is commonplace[18]. Understanding cooperation in humans is challenging owing to the complex interaction of genetics, individual learning, and culture[90]. Nevertheless, it is clear that human cooperation is enforced in a diversity of ways[91]. This includes partner choice, whereby individuals form partnerships and bonds with more cooperative individuals[92-94] (Table 1). And within particular relationships, cooperation often rests upon reciprocation whereby cooperation is directed at cooperators[95, 96], or those of good reputation[97]. There is also punishment, where transgressions are met with harm from other individuals [98], where punishment is particularly favoured in powerful individuals [99] or when punishment itself improves reputation [100].

The social insects have been central to the insight that kin selection can be a key for the evolution of cooperation[5-7]. Specifically, a morphologically-distinct worker caste appears to have only evolved in species that are monogamous, which ensures high relatedness between the workers (the queen's daughters) and the queen's other offspring[101, 102]. While relatedness is important in the origin of eusociality, its role in the derived social insects is less clear[103]. In derived species, multiply-mated queens are common and sibling relatedness is much reduced as a result. The canonical example is the honeybee, *Apis mellifera*, where queens mate tens of times. It is striking then that honeybees have one of the most conflict free societies described. To achieve this, honeybees and other derived species employ a diverse set of enforcement mechanisms that limit the potential for conflict [103, 104].

One such mechanism is "worker policing", whereby the low relatedness between workers leads them to prevent each other's reproduction and allow the queen alone to reproduce[105, 106]. Some forms of worker policing may evolve as a byproduct of worker-

worker competition, which is not an evolved enforcement mechanism in the strict sense (see discussion in Box S1). Nevertheless, the effect is to promote cooperation and colony efficiency[105]. A second route to enforcement in the social insects is the differential feeding of larvae destined to become new queens as opposed to new workers. In the honeybee, this is achieved by raising new queens in special cells under tight worker control[104]. Control over caste fate is thereby taken away from larvae. Strikingly, this control is lacking in some stingless bees, where larvae are free to determine their own caste. The result is that around a fifth of female larvae emerge as new queens, only to be immediately executed by the workers[107] (Figure 1). In other stingless bees, larger cells for queens have evolved that limit or remove this conflict[107]. This suggests that queen-worker dimorphism – the defining feature of derived eusociality and a major transition in societies[8] – can evolve as a mechanism of enforcement.

Interspecific mutualism

The evolution of cooperation between species (mutualism) occurs in a diverse range of systems, from plants and pollinators, through client and cleaner fish, to the vast diversity of microbial symbiosis of plants and animals that have been revealed by DNA sequencing[108]. Like the evolution of genomes and eukaryotic cells, interactions between species represent egalitarian alliances of individuals, where Hamilton's relatedness never occurs (except within members of each species)[109]. Coinheritance of partners is a feature of some – including endosymbionts (above), lichens, and the symbionts of leafcutter ants and termites – but is far from general. Indeed, horizontal transmission of symbionts is so common that it was recently called a "paradox", based on the assumption that vertical transmission is the best way for a host to manage conflicts[110].

This conclusion, however, overlooks a large body of theoretical and empirical work showing the importance of enforcement in mutualisms[109, 111]. As in humans (above), mutualistic species engage in partner choice to interact preferentially with more cooperative partners[12, 93] (Table 1). Bumblebees pollinate plants, receiving nectar in return, and if too little nectar is provided, a bee may leave a plant, remember and not return[112]. Many ants tend and protect aphids on plants, receiving sugary honeydew in return. If too little honeydew is provided by the aphids, the ants may leave or even eat the

aphids[113]. Figs can abort fruits containing developing fig wasp larvae if their mothers did not pollinate the fig when depositing eggs. The fig species with the most effective sanctions (that abort most readily) have more cooperative wasps that more often pollinate[114]. Finally, cleaner fish remove parasites from much-larger client fish at specific sites known as cleaner "stations". If a cleaner also bites and takes host tissue, their client will be less likely to return to that station[115].

Comparable mechanisms are used by hosts to choose and promote cooperative microbial symbionts. For example, legumes preferentially provide nutrients to root nodules that contain nitrogen fixing bacteria[116] (Box S1). Leafcutter ants farm vertically-transmitted fungi in large gardens, which digest the leaf material for the ants. Workers actively tend the gardens and remove other strains and species of fungi[117], which ensures that the beneficial fungi dominate. Mammalian hosts mount an inflammation response against bacteria that breach the gut epithelial barrier[27]. Mutualists also directly manipulate their partner's behaviour to make it more favourable (Table 1). For example, floral morphology often appears adapted to reward pollinators rather than robbers that collect nectar without pollination (Figure 2f; Supplement, Model 6), and some termites ensure only a single genotype of fungus is transmitted between generations. This, like uniparental inheritance of symbionts (Figure 2b), is hypothesised to have evolved to limit conflicts between fungal strains[118].

The bobtail squid provides a particularly striking example of enforcement[27]. It has multiple mechanisms that help to select the bacterium *Vibrio fischeri* over other species to enter its light organ and, once there, promote the luminescence that helps the squid to hunt and hide[119, 120]. The squid-*Vibrio* system is striking because enforcement appears to eliminate nearly all selfish behaviour in symbionts, despite them being horizontally acquired by the squid each generation (Figure 2e). Enforcement alone can create alliances as cooperative as a single evolutionary individual; coinheritance is not a requirement.

Evolutionary theory of enforcement

There is empirical evidence for the importance of enforcement at all biological scales. However, our survey is necessarily limited by the number of examples where enforcement has been looked for. Do these examples then reflect a general tendency for enforcement to

evolve across biological scales? This is where mathematical theory is valuable for its ability to reveal general predictions that are robust across many parameters and systems. However, there is currently no general theory of enforcement that covers all biological scales. There are several models predicting the importance of enforcement for diverse animal groups and mutualisms[14, 121-124]. This includes the seminal models of reciprocal altruism and tit-for-tat where individuals only cooperate if a partner is also cooperating[95, 96] (although this strategy can only enforce cooperation when the partner is also a reciprocator). Another important model is Frank's on mutual policing, which was aimed a diverse set of social systems[125]. However, it was subsequently criticised for relying on assumptions that, it was argued, made the evolution of policing particularly likely[11]. The general issue is that enforcement models have to be relatively complex if they are to capture both the evolution of cooperation *and* the evolution of enforcement as separate traits[14]. This can make them parameter heavy and may explain why we lack a general evolutionary model that captures enforcement across biological scales.

On this basis, we decided to build a suite of models that are each consistent with a specific biological example, where the examples span the scales of cooperation (Figure 2; Supplement). These models also allow us to mathematically define cooperation and enforcement in each case, enabling the formal comparison of very different examples. Modelling also allows one to explore ancestral scenarios that are hard to study empirically. We therefore focussed on relatively simple systems – including early replicators, isogamous cells and simple multicellular organisms – to ask in particular whether enforcement is predicted at early stages of cooperative evolution (Figure 2).

We first explored how each system functions without enforcement (Figure 3a; Supplement). Importantly, all models recapitulated the key prediction from social evolution that cooperative systems are susceptible to the evolution of selfish phenotypes that compromise higher-level function[1-4]. With this benchmark in place for each model, we then asked, across a range of parameters, whether enforcement will evolve and, if so, how this affects the evolution of cooperation. Despite being based on diverse examples across biological scales, all models identify conditions where enforcement will evolve (Figure 3). As is typical for evolutionary models[5], benefits and costs matter. In particular, high costs to enforcement can make it less likely to evolve, or less effective when it does

evolve (e.g. Figure 3g). Nevertheless, enforcement is predicted in every system and importantly this occurs over a wide range of parameters, which is consistent with the general importance of enforcement independently of both biological details and system.

Conclusions

We have considered systems ranging from genomes to interspecific mutualisms, including cases that are examples of major transitions, and cases that are not. Putting these diverse examples side-by-side reveals that enforcement plays a central role at every level of biological organisation (Figures 1 and 3). Of course, enforcement is not the only process that limits evolutionary conflicts in cooperative systems. Nor is enforcement always the most important mechanism in every example of cooperation. Some potential conflicts are constrained by pre-existing biology such that they are never expressed or are self-resolving[105], and some conflicts have a weak effect on cooperation when they are expressed[126]. It is further clear that family life plays a central role in fraternal systems, where it enables the evolution of a stable reproductive division of labour[3, 5, 6]. Egalitarian interactions lack relatedness and the reproductive division of labour, but coinheritance can nevertheless help to align evolutionary interests[29, 30, 51]. While much discussed, however, coinheritance and relatedness only apply to specific systems, such as endosymbiosis or animal groups respectively. By contrast, when costly selfish behaviour emerges, the evolution of enforcement that restores cooperation is possible (Figure 3), and observed (Figure 1), across all biological scales. It is in this sense that we find enforcement to be central to cooperative evolution (Box 2).

While generally important, the effects of enforcement vary, and it is arguably most critical in egalitarian interactions where relatedness is lacking[125]. The effect of enforcement also varies within each level of biological organisation[3], including among very similar species. Striking examples of this variation are the eusocial hymenoptera[105] and fig and fig-wasp mutualism[114]. In both cases, the effectiveness of enforcement (rates of removal of worker eggs and senescence of unpollinated figs respectively) has been experimentally measured across a set of related species, and, in both cases, more effective enforcement predicts higher cooperation (low worker ovary activation and pollen-free fig

wasps). Such large data sets are hard to obtain, but are telling because they suggest that enforcement is a dominant shaper of cooperation over evolutionary time.

At a finer scale still, the effectiveness of enforcement varies within species. The honeybee is often considered a pinnacle of conflict resolution and worker policing is typically extremely effective. However, worker lineages occur that both lay eggs and escape policing that can drive colony collapse and devastate apiaries [127, 128]. Failures of enforcement are also seen at other levels of organisation, including the *P* element in *Drosophila*[37, 38], *Wolbachia* male killing[62] and the transmissible facial tumour in the Tasmanian devil[82]. Like between-species comparisons, these examples underline the importance of enforcement; when it fails there are major consequences for cooperation that can even threaten species with extinction[82] (Figure 1). Within-species examples further emphasise how enforcement can drive complex evolutionary dynamics, where social evolution combines with coevolution to influence cooperation[34].

In spite of this complexity, such examples also show that enforcement is often highly effective. When new selfish elements are introduced to a population, the evolution of suppression is often rapid, again as seen with the *P* element [39] and *Wolbachia* male killing[63]. Such rapidity can make enforcement difficult to detect. Moreover, once in place, the most effective enforcement will typically be the least tested or seen, which may lead its importance to be underestimated[88] (Figure 1). The best evidence of enforcement often comes from challenging manipulative experiments that simulate what would happen were individuals to rebel (e.g.[19, 89, 116]). Similarly, the phenotypic consequences of selfish genetic elements are often seen in hybrids where enforcement fails, but not in within-population crosses where it functions[21].

There is now a large body of evidence that enforcement shapes cooperation across all levels of biology. This include systems that have undergone a major transition, like derived social insects, as well as those that have not, like many mutualisms. Both theory and data emphasise that biological details matter for how enforcement evolves. Looking forward, the challenge is to seek general mathematical principles for enforcement evolution[3-6, 11]. It is already clear that the costs and benefits of enforcement will be important (Figure 3; Supplement). While high costs can be prohibitive, a given cooperative system may explore diverse enforcement strategies until a low cost solution arises. Low

costs to enforcement are further facilitated by power asymmetries[3, 34, 99, 105, 129]. Enforcement can be enabled by a single powerful individual, such as a host who exercises control over symbionts (Figure 2e), or by majority rule, such as within an animal society[103-105] (Figure 2d). There is also a role for pre-adaptations; pre-existing features of the biology of a system can enable enforcement and promote its effectiveness (Box S1). Systems where enforcement is not possible will be less likely to see increases in cooperation over evolutionary time, or even see cooperation fall away. It is in those systems where enforcement does arise that we see the full elaboration, and wonders, of cooperative evolution.

Correspondence should be directed to kevin.foster@zoo.ox.ac.uk

Author contributions

All authors contributed to the writing of the article. JAÅ and KRF conceived the article, NGD and KRF developed the models, and NGD did the analysis.

We declare no competing interests.

Box 1 Formal definition of enforcement

Enforcement is an action that evolves, at least in part, to reduce selfish behaviour within a cooperative alliance. Selfish behaviours are those that benefit an actor and reduce its cooperation with one or more recipients in the alliance (Glossary). We more explicitly define selfishness (z) in a series of models in the Supplement where z always corresponds to a specific behaviour that benefits an actor while reducing the benefits they provide to others in the context of a specific cooperative venture.

Enforcement can be unilateral or a joint effort, and a cooperative alliance can be fleeting, as between a plant and pollinator that may interact only once, or long-lasting, as among genes in a genome that are partners for (at least) the lifetime of the organism. Here, we formalize our definition of enforcement by considering individuals that can invest in enforcement, a, in two key ways: either by changing the strength of interaction with different partners (partner choice; Table 1), or by manipulating the amount that their partners display selfish versus cooperative behaviours (partner manipulation; Table 1). An example of partner choice reducing selfishness is a bobtail squid host that prevents colonization by a bacterial strain that invests in its own growth rather than light production[120]. An example of partner manipulation reducing selfishness is the workers of the queenless ant *Dinoponera quadriceps* that physically restrain a member of the colony that is challenging the reproductive female, thereby restoring cooperative behaviour[106].

To be explicit about how enforcement can occur, we distinguish between the selfishness of group members z, and the selfishness experienced by a focal individual y. In the models below, these are sometimes equivalent but, where they are not, the key measure of selfishness is that experienced by an enforcer y (plotted in Figure 3), which captures any benefits of enforcement for the enforcer (e.g. Model 1 in Supplement). Consider a focal actor who experiences a level of selfishness, y, from others which is a function of that actor's investment in enforcement, a, such that $y = y(a)$. If investing in enforcement (increasing a) reduces the selfishness experienced by the actor (decreases $y(a)$), then the focal actor is enforcing cooperation. Specifically, in the absence of enforcement ($a = 0$), the focal actor's strength of interaction with individual i of n partners is defined as as $\beta_i = \beta_i(0)$ and individual i's level of selfishness as $z_i = z_i(0)$. The β_i function quantifies how much an individual's selfishness matters to the focal actor's fitness; for example, when the

17

focal individual interacts with all of its n partners equally, we would have $\beta_i = \frac{1}{n}$. When the actor invests in enforcement ($a > 0$), these change to $\beta_i(a)$ and $z_i(a)$. Formally, the focal actor's experienced level of selfishness can be written without enforcement as

$$y(0) = \sum_i^n \beta_i(0)z_i(0) \qquad (1)$$

and with enforcement as

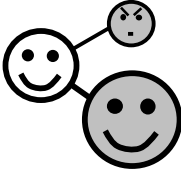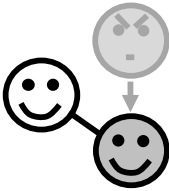$$y(a) = \sum_i^n \beta_i(a)z_i(a). \qquad (2)$$

We are interested in the evolution of selfish traits and their suppression by enforcement. A trait z is selfish if, in the absence of enforcement, a marginal increase in the trait increases its possessor's fitness w (i.e., $dw/dz > 0$) but decreases the fitness of other individuals w' (i.e., $dw'/dz < 0$). Meanwhile, a trait a causes the enforcement of cooperation if a marginal increase in the trait value decreases experienced selfishness (i.e. $dy/da < 0$). The evolution of enforcement can occur because of its effects on selfishness within a cooperative alliance, or due to other effects unrelated to these effects (Box S1). However, our discussion is focused on the former cases where the focal trait a evolves, at least in part, because of its effects on selfishness, which implies that $dw/dy < 0$ and $dy/da < 0$, and by transitive logic $dw/da > 0$. For simplicity, we phrase the above in terms of direct benefits to the enforcer, but enforcement also includes cases where an actor reduces selfish behaviour within a cooperative alliance due to benefits to its relatives (indirect benefits; Box 2; Supplement).

Box 2 Why does enforcement evolve? Kin selection, direct benefits and by-products

Why do cooperative behaviours evolve? A typical answer identifies benefits to an actor that are either direct - improves personal reproduction - or indirect, which improve the reproduction of relatives (kin selection)[5, 6]. Here, we have argued that a process is missing from many explanations: enforcement. However, like cooperation itself, enforcement is also a behaviour whose evolution needs to be explained[14]. We model diverse examples in the Supplement, which illustrate that enforcement also evolves due to either direct or indirect fitness benefits to an actor. For example, a host can benefit directly from enforcing cooperation in a symbiont (Figure 2e), while some forms of cancer repression require self-sacrifice by cells that is only possible with kin selection (Figure 2c). In this sense, therefore, kin selection and enforcement are not independent explanations for cooperation. However, the typical kin selection explanation for cooperation does not include enforcement - e.g. in the major transitions [4] - and in this sense, kin selection and enforcement evolution are alternative ways of explaining cooperation whose importance can be experimentally tested across biological systems. And, here, enforcement clearly has the greater reach as it is important in both egalitarian and fraternal interactions, which is central to our conclusion of its importance. In a similar vein, our models also identify several cases where little cooperation is directly favoured through either kin selection or direct benefits, but highly effective enforcement *is* favoured through these processes (Figure 3). For example, in Model 1, the shared interest of being in the same cell is insufficient to generate much cooperation between replicators. However, it is sufficient to drive enforcement that then allows high levels of cooperation to evolve (Figure 3b; Equation S4 versus S6 in Supplement; Model 1). Such examples emphasise the importance of including enforcement in the theories of social evolution.

We have defined enforcement as processes that evolve because they limit selfishness and promote cooperation (Box 1). However, similar mechanisms that limit selfishness can also evolve independently of their effects on cooperation[14, 88]. We discuss such by-product enforcement in more detail in the Supplement, along with the potential for species-level selection to enrich for such mechanisms in Box S1.

Table 1 Mechanisms of enforcement

Mechanism	Illustrative examples
Partner choice[1] and related mechanisms *Increase interactions with cooperators (changes β_i, Eq.1)*	Humans choose to preferentially interact with cooperators[98]. Choice in mutualisms[12, 93, 94] e.g. cleaner and client fish, plants and pollinators. Increased resources to beneficial mutualists, including mammalian microbiota, legume sanctions of rhizobia, fruit abscission by figs[27, 114, 116]. Immune suppression of harmful microbiota[27], Bobtail-squid enzymes that harm non-light producing bacteria[120]. Self-enforcement that removes non-cooperators e.g. apoptosis of proliferative cells in mammalian development[2] [75].
Partner manipulation *Change behaviour of individuals (changes z_i, Eq.1)*	Genomic suppressors of conflict e.g. silencing of transposable elements[37, 38]. Limiting competition between partners[14] e.g. uniparental inheritance of endosymbionts and organelles[51, 52] and termites enforcing high relatedness in their fungus [118]. Reciprocal altruism and indirect reciprocity promotes cooperation in other reciprocators [95, 97]. Self-constraint via genetic redundancy[69], e.g. regulation of mammalian cell proliferation is robust to mutation[25]. Punishment in vertebrate societies; some worker policing[2] in insects[19, 76, 87, 89, 103, 106]. Compartmentation to limit resource access e.g. specialised queen cells in honeybees[2][104, 107], epithelial barrier in microbiome[27], floral morphology that directs nectar only to pollinators (Model 6, Supplement).

1) Partner choice is a form of enforcement that increases the likelihood that a partner is a cooperator, thereby reducing selfishness in an alliance. However, it does not guarantee that cooperators preferentially associate as cooperation and choice can be unlinked traits (e.g. a selfish host will benefit from selecting cooperative symbionts) [12]. 2) Individuals receiving direct benefits of enforcement (white face in cartoon) are relatives of the enforcer, not the enforcer itself (Box 2).
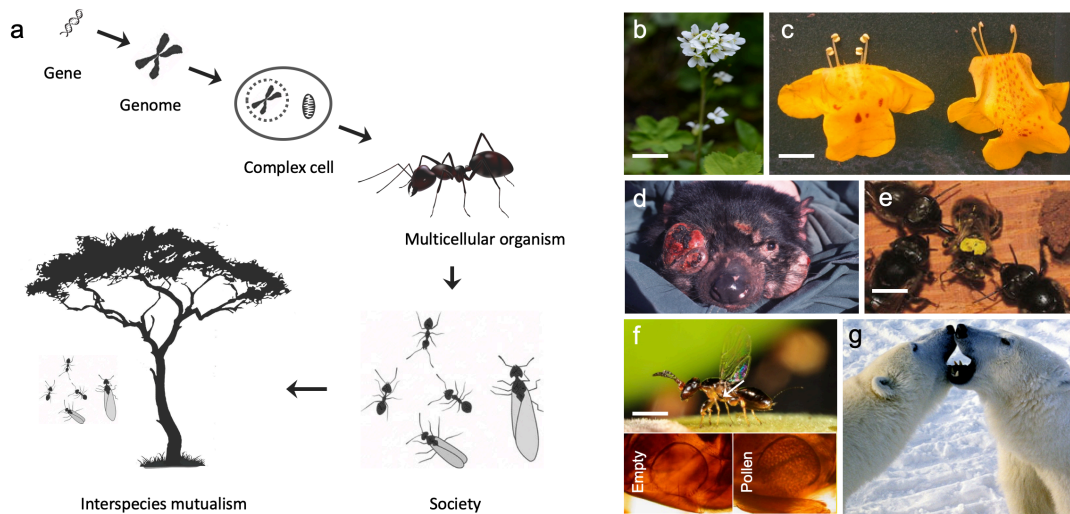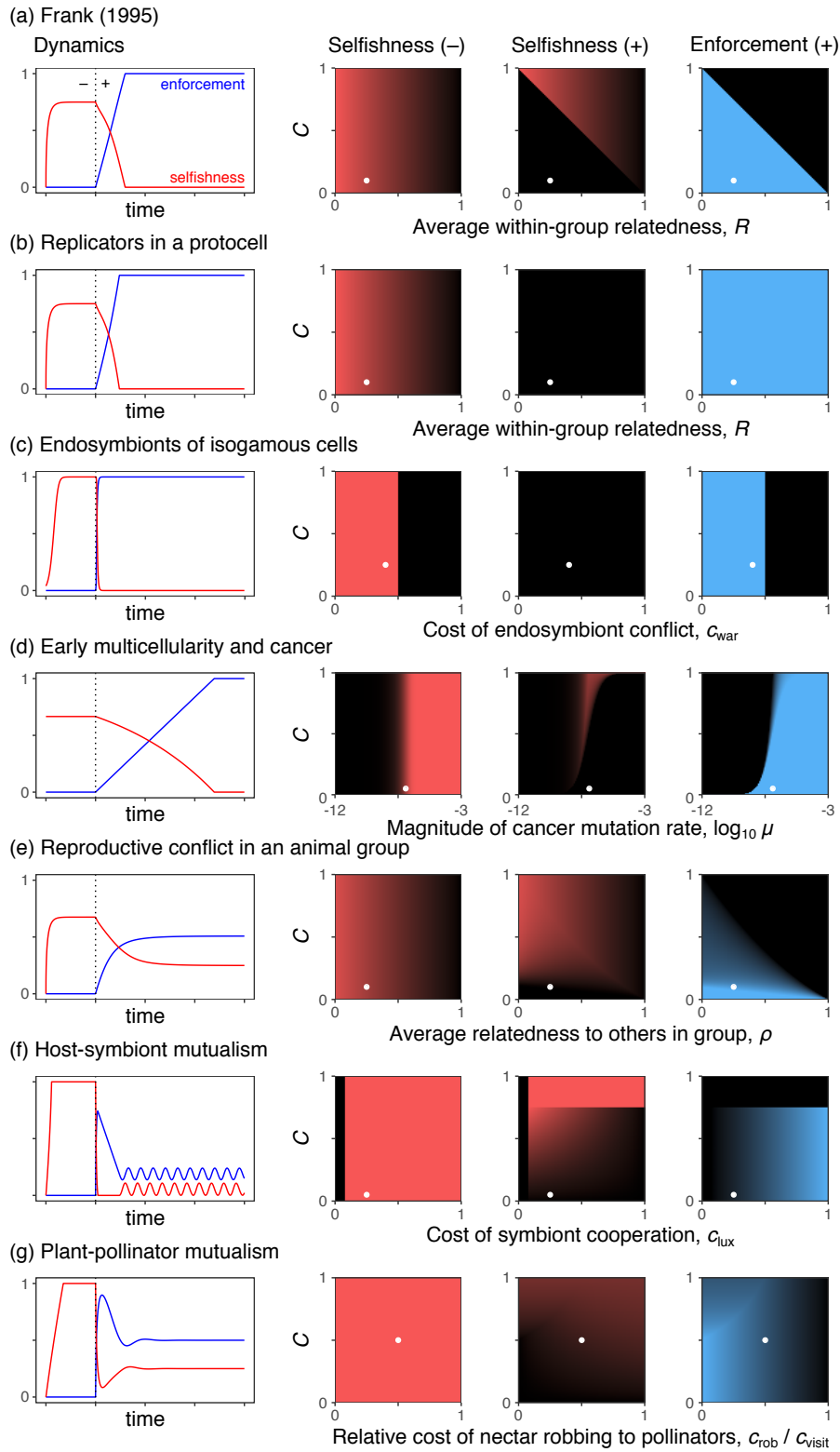
Figure 1. The importance of enforcement is revealed by its absence at all levels of biological organisation. (a) Life has a nested organisation that rests upon cooperation at successive levels. (b) Genomes: *Arabis alpina* appears to have a reduced ability to silence and remove retrotransposons. Transposons contribute more DNA to its genome than occurs in the entire genome of some relatives[42]. Scale bar equals 5 cm. (Picture credit: Jon Ågren). (c) Complex cells: Cytoplasmic male sterility driven by mitochondria revealed in *Mimulus guttatus × nasutus* F_2 hybrids. Male fertile flowers with pollen-producing anthers (left) and male sterile with deformed sterile anthers (right)[58]. Scale bar equals 1.5 cm. (Picture credit: Andrea Case). (d) Multicellularity: Uncontrolled growth of the transmissible facial cancer tumour in Tasmanian devil, which is associated with a loss of MHC diversity and tissue rejection[82, 83] (Picture credit: Menna Jones under creative common licence). (e) Sociality: Unlike many other species, caste fate is not forced upon larvae in several species of stingless bees. As a result, far more larvae develop into queens than the colony needs, which are then killed by workers upon emergence[107]. Scale bar equals 4.5 mm. (Picture credit: Tom Wenseleers). (f) Mutualism: Some fig wasps do not carry pollen in their pollen pocket (arrow, insets) but figs can abort fruits containing developing fig wasps if the wasps do not pollinate. In mutualisms where the fig readily abort, the wasps more often carry pollen[114]. Scale bar equals 1 mm (Picture credit: Charlotte Jander). (g) No cooperative system: Many reproductive conflicts occur without little prospect of enforcement, simply because there is no collective interest to protect, such as competition between males over a mate. Some images still require permission for use (Picture credit Brocken Inaglory under creative commons licence).
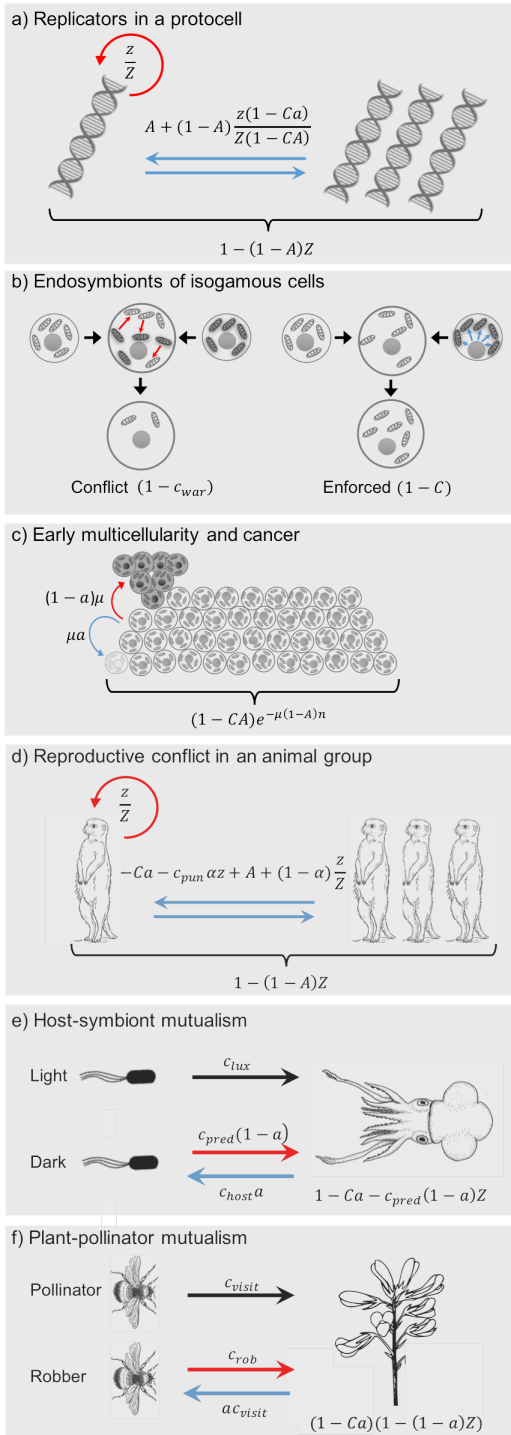
Figure 2. Models of enforcement across the levels of biological organisation. Selfishness is indicated in red, enforcement in blue. (a) Replicators and early genome evolution. A set of replicators in a protocell that can either just replicate (red arrow) or replicate more slowly to invest in the mutual inhibition of others' replication (blue arrows), which improves collective fitness (bottom). (b) Endosymbionts and the evolution of

complex cells. Costly endosymbiont conflict (red, left) can be reduced by nuclear modifiers that enforce uniparental inheritance of symbionts (blue, right). (c) Multicellular evolution and cancer. Cells mutate to become cancerous (red arrow) at a cost to the cellular group (bottom), but a cell can invest in self enforcement (blue arrow) that suppress the impact of cancerous mutations e.g. by undergoing apoptosis. (d) Animal groups. An individual can either just invest in reproduction (red arrow) or reproduce less and invest in punishment of other individuals who reproduce a lot (blue arrows), which improves group fitness (bottom). (e) Host-symbiont mutualism. A symbiont can either invest in cooperation (black arrow), or in reproduction (red arrow) and face an enforcement cost from the host (blue arrow). (f) Plant-pollinator mutualism. A pollinator can either pollinate (black arrow) or attempt to rob the flower (red arrow), which can fail if the plant protects the flower and forces the insect to visit it normally (bottom, blue arrow). See Supplement for full model descriptions and Table S2 for variables and parameters.

(a) Frank (1995)

Dynamics | Selfishness (−) | Selfishness (+) | Enforcement (+)

Average within-group relatedness, R

(b) Replicators in a protocell

Average within-group relatedness, R

(c) Endosymbionts of isogamous cells

Cost of endosymbiont conflict, c_{war}

(d) Early multicellularity and cancer

Magnitude of cancer mutation rate, $\log_{10} \mu$

(e) Reproductive conflict in an animal group

Average relatedness to others in group, ρ

(f) Host-symbiont mutualism

Cost of symbiont cooperation, c_{lux}

(g) Plant-pollinator mutualism

Relative cost of nectar robbing to pollinators, c_{rob} / c_{visit}

Figure 3. The evolution of enforcement is predicted to reduce selfishness and promote cooperation across all levels of biological organisation. The first column shows example

25

dynamics of selfishness (red) and enforcement (blue); first without the possibility of enforcement and then when it is able to evolve (right of dotted line). Columns two to four show parameter sweeps of i) evolved selfishness (red intensity) without enforcement ii) evolved selfishness (red intensity) when enforcement can evolve and iii) evolved enforcement (blue intensity). In the parameter sweeps, the y axis is always the cost of enforcement C, while the x axes sweep a second key parameter from each model. White dots show the parameters used in the example dynamics plots. See Supplement for full model descriptions and Table S2 for variables and parameters.

Glossary

Altruism: Following Hamilton[5, 6], defined as an action that evolves to increase the fitness of a recipient while decreasing the lifetime reproductive fitness of the actor, sometimes known as strong altruism[7].

Cooperative system: Alliance, however transient, between entities (genes, cells, individuals), which arises via the evolution of cooperative phenotypes.

Cooperation: Social phenotype that has a positive fitness effect on another individual and evolved, at least in part, because of this effect.

Direct benefit: A fitness benefit to a focal individual via increasing its lifetime personal reproduction.

Egalitarian alliance: Cooperative system made up of unrelated individuals or entities e.g. different loci in a genome.

Enforcement: an action that evolves, at least in part, to reduce selfish behaviour within a cooperative alliance.

Fraternal alliance: Cooperative system made up of related entities of the same species e.g. cells in a multicellular organism.

Genetic relatedness: Genetic similarity at a locus for a social phenotype as compared to the population average allele frequency. Often driven by kinship, which is the probability of this similarity based on recent common ancestry.

Indirect benefit: A fitness benefit to a focal individual's relatives. Actions that yield indirect benefits to individuals can be promoted by kin selection.

Individual: A stable, physically integrated collective.

Kin selection (as a process): Natural selection involving non-zero genetic relatedness between individuals. Kin selection within family groups is a major explanation for cooperation.

Major transition: A shift in the way information is transmitted across generations. Often involves the coming together of formerly independently reproducing entities to form a new level of individuality, where entities become irreversibly dependent upon one another.

Preadaptation: An adaptation that evolved for one function in an ancestor that later enables a different function in a descendent.

Selfishness: a phenotype that benefits an actor and harms a recipient. When selfish traits harm other members of a cooperative system, they also reduce cooperation.

References

1.   Maynard Smith, J., and Szathmáry, E. (1995). The Major Transitions in Evolution, (Oxford; New York: W.H. Freeman Spektrum).

2.   Michod, R.E. (1999). Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality, (Princeton, N.J.: Princeton University Press).

3.   Bourke, A.F.G. (2011). Principles of Social Evolution, (Oxford; New York: Oxford University Press).

4.   West, S.A., Fisher, R.M., Gardner, A., and Kiers, E.T. (2015). Major evolutionary transitions in individuality. Proc. Natl. Acad. Sci. U.S.A *112*, 10112-10119.

5.   Hamilton, W.D. (1964). The genetical evolution of social behaviour. I. J. Theor. Biol. *7*, 1-16.

6.   Hamilton, W.D. (1964). The genetical evolution of social behaviour. II. J. Theor. Biol. *7*, 17-52.

7.   Foster, K.R. (2009). A defense of sociobiology. Cold Spring Harb. Symp. Quant. Biol. *74*, 403-418.

8.   Boomsma, J.J., and Gawne, R. (2018). Superorganismality and caste differentiation as points of no return: how the major evolutionary transitions were lost in translation. Biol. Rev. Camb. Philos. Soc. *93*, 28-54.

9.   Queller, D.C. (1997). Cooperators since life began. Q. Rev. Biol *72*, 184-188.

10.   Wheeler, W.M. (1911). The ant-colony as an organism. J Morphol *22*, 307-325.

11.   El Mouden, C., West, S.A., and Gardner, A. (2010). The enforcement of cooperation by policing. Evolution *64*, 2139-2152.

12.   Foster, K.R., and Wenseleers, T. (2006). A general model for the evolution of mutualisms. Journal of evolutionary biology *19*, 1283-1293.

13.   Szathmáry, E. (2012). Transitions and Social Evolution. Philosophy & Theory in Biology *4*.

14.   Frank, S.A. (2003). Perspective: Repression of competition and the evolution of cooperation. Evolution *57*, 693-705.

15.   Hamilton, W.D. (1967). Extraordinary sex ratios. Science *156*, 477-488.

16.   Leigh, E.G. (1977). How does selection reconcile individual advantage with the good of the group? Proc. Natl. Acad. Sci. U.S.A. *74*, 4542-4546.

17. Leigh, E.G. (1971). Adaptation and Diversity: Natural History and the Mathematics of Evolution, (San Francisco,: Freeman).

18. Alexander, R.D. (1987). The biology of moral systems, (Transaction Publishers).

19. Mulder, R.A., and Langmore, N.E. (1993). Dominant males punish helpers for temporary defection in superb fairy-wrens. Anim. Behav. *45*, 830-833.

20. Ratnieks, F.L.W., and Visscher, P.K. (1989). Worker policing in the honeybee. Nature *342*, 796-797.

21. Ågren, J.A. (2014). Evolutionary transitions in individuality: insights from transposable elements. Trends Ecol Evol *29*, 90-96.

22. Friedli, M., and Trono, D. (2015). The developmental control of transposable elements and the evolution of higher species. Annu Rev Cell Dev Bi *31*, 429-451.

23. Blackstone, N.W. (2013). Why did eukaryotes evolve only once? Genetic and energetic aspects of conflict and conflict mediation. Philos T R Soc B *368*, 20120266.

24. Radzvilavicius, A.L., and Blackstone, N.W. (2015). Conflict and cooperation in eukaryogenesis: implications for the timing of endosymbiosis and the evolution of sex. J R Soc Interface *12*, 20150584.

25. Nunney, L. (2013). The real war on cancer: the evolutionary dynamics of cancer suppression. Evol. Appl. *6*, 11-19.

26. Aktipis, C.A., Boddy, A.M., Jansen, G., Hibner, U., Hochberg, M.E., Maley, C.C., and Wilkinson, G.S. (2015). Cancer across the tree of life: cooperation and cheating in multicellularity. Philos T R Soc B *370*, 20140219.

27. Foster, K.R., Schluter, J., Coyte, K.Z., and Rakoff-Nahoum, S. The evolution of the host microbiome as an ecosystem on a leash. Nature *548*, 43-51.

28. Higgs, P.G., and Lehman, N. (2015). The RNA World: molecular cooperation at the origins of life. Nat Rev Genet *16*, 7-17.

29. Dawkins, R. (1976). The Selfish Gene, (New York: Oxford University Press).

30. Herre, E.A., Knowlton, N., Mueller, U.G., and Rehner, S.A. (1999). The evolution of mutualisms: exploring the paths between conflict and cooperation. Trends Eco. Evol. *14*, 49-53.

31.   Iranzo, J., Puigbo, P., Lobkovsky, A.E., Wolf, Y.I., and Koonin, E.V. (2016). Inevitability of genetic parasites. Genome Biol Evol, doi: 10.1093/gbe/evw1193.

32.   Burt, A., and Trivers, R. (2006). Genes in Conflict: The Biology of Selfish Genetic Elements, (Cambridge, Mass.: Belknap Press of Harvard University Press).

33.   Hickey, D.A. (1982). Selfish DNA - a sexually transmitted nuclear parasite. Genetics *101*, 519-531.

34.   Hurst, L.D., Atlan, A., and Bengtsson, B.O. (1996). Genetic conflicts. The Quarterly Review of Biology *71*, 317-364.

35.   Wagner, A. (2006). Periodic Extinctions of Transposable Elements in Bacterial Lineages: Evidence from Intragenomic Variation in Multiple Genomes. Molecular Biology and Evolution *23*, 723-733.

36.   Kelleher, E.S. (2016). Reexamining the P-Element invasion of Drosophila melanogaster through the lens of piRNA silencing. Genetics *203*, 1513-1531.

37.   Lu, J., and Clark, A.G. (2010). Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in Drosophila. Genome Res. *20*, 212-227.

38.   Kelleher, E.S., and Barbash, D.A. (2013). Analysis of piRNA-mediated silencing of active TEs in Drosophila melanogaster suggests limits on the evolution of host genome defense. Mol. Biol. Evol. *30*, 1816-1829.

39.   Hill, T., Schlotterer, C., and Betancourt, A.J. (2016). Hybrid dysgenesis in Drosophila simulans associated with a rapid invasion of the P-element. PLoS Genet. *12*, e1005920.

40.   Blumenstiel, J.P. (2011). Evolutionary dynamics of transposable elements in a small RNA world. Trends Genet. *27*, 23-31.

41.   Ågren, J.A., and Wright, S.I. (2015). Selfish genetic elements and plant genome size evolution. Trends Plant Sci. *20*, 195-196.

42.   Willing, E.M., Rawat, V., Mandakova, T., Maumus, F., James, G.V., Nordstrom, K.J., Becker, C., Warthmann, N., Chica, C., Szarzynska, B., et al. (2015). Genome expansion of Arabis alpina linked with retrotransposition and reduced symmetric DNA methylation. Nat. Plants *1*, 14023.

43. Ågren, J.A., and Clark, A.G. (2018). Selfish genetic elements. PLoS genetics *14*, e1007700.

44. Lindholm, A.K., Dyer, K.A., Firman, R.C., Fishman, L., Forstmeier, W., Holman, L., Johannesson, H., Knief, U., Kokko, H., Larracuente, A.M., et al. (2016). The ecology and evolutionary dynamics of meiotic drive. Trends Ecol Evol *31*, 315-326.

45. Burt, A. (2003). Site-specific selfish genes as tools for the control and genetic engineering of natural populations. Proceedings of the Royal Society of London B: Biological Sciences *270*, 921-928.

46. Unckless, R.L., Clark, A.G., and Messer, P.W. (2017). Evolution of resistance against CRISPR/Cas9 gene drive. Genetics *205*, 827-841.

47. Champer, J., Reeves, R., Oh, S.Y., Liu, C., Liu, J., Clark, A.G., and Messer, P.W. (2017). Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. PLoS genetics *13*, e1006796.

48. Dyall, S.D., Brown, M.T., and Johnson, P.J. (2004). Ancient invasions: from endosymbionts to organelles. Science *304*, 253-257.

49. Archibald, J.M. (2015). Endosymbiosis and eukaryotic cell evolution. Curr Biol *25*, R911-R921.

50. Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., and Penny, D. (2002). Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. U.S.A. *99*, 12246-12251.

51. Cosmides, L.M., and Tooby, J. (1981). Cytoplasmic inheritance and intragenomic conflict. J. Theor. Biol. *89*, 83-129.

52. Eberhard, W.G. (1980). Evolutionary consequences of intracellular organelle competition. Q Rev Biol *55*, 231-249.

53. Hastings, I.M. (1992). Population genetic aspects of deleterious cytoplasmic genomes and their effect on the evolution of sexual reproduction. Genetics Research *59*, 215-225.

54. Christie, J.R., Schaerf, T.M., and Beekman, M. (2015). Selection against heteroplasmy explains the evolution of uniparental inheritance of mitochondria. Plos Genet. *11*, e1005112.

55. Sato, M., and Sato, K. (2013). Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1833*, 1979-1984.

56. Frank, S.A., and Hurst, L.D. (1996). Mitochondria and male disease. Nature *383*, 224-224.

57. Touzet, P., and Meyer, E.H. (2014). Cytoplasmic male sterility and mitochondrial metabolism in plants. Mitochondrion *19*, 166-171.

58. Fishman, L., and Willis, J.H. (2006). A cytonuclear incompatibility causes anther sterility in Mimulus hybrids. Evolution *60*, 1372-1381.

59. Case, A.L., Finseth, F.R., Barr, C.M., and Fishman, L. (2016). Selfish evolution of cytonuclear hybrid incompatibility in Mimulus. Proc. R. Soc. B. *283*, 20161493.

60. Ågren, J.A., Munasinghe, M., and Clark, A.G. (2018). Sexual Conflict through Mother's Curse and Father's Curse. Theoretical Population Biology, in press.

61. Fenn, K., and Blaxter, M. (2006). Wolbachia genomes: revealing the biology of parasitism and mutualism. Trends Parasitol. *22*, 60-65.

62. Dyson, E.A., and Hurst, G.D.D. (2004). Persistence of an extreme sex-ratio bias in a natural population. Proc. Natl. Acad. Sci. U.S.A. *101*, 6520-6523.

63. Charlat, S., Hornett, E.A., Fullard, J.H., Davies, N., Roderick, G.K., Wedell, N., and Hurst, G.D.D. (2007). Extraordinary flux in sex ratio. Science *317*, 214.

64. Nadell, C.D., Drescher, K., and Foster, K.R. (2016). Spatial structure, cooperation and competition in biofilms. Nat. Rev. Microbiol. *14*, 589-600.

65. Griffin, A.S., West, S.A., and Buckling, A. (2004). Cooperation and competition in pathogenic bacteria. Nature *430*, 1024-1027.

66. Nadell, C.D., Foster, K.R., and Xavier, J.B. (2010). Emergence of spatial structure in cell groups and the evolution of cooperation. PLoS Comput. Biol. *6*, e1000716.

67.   Harrison, F., and Buckling, A. (2005). Hypermutability impedes cooperation in pathogenic bacteria. Curr Biol *15*, 1968-1971.

68.   Foster, K.R., Shaulsky, G., Strassmann, J.E., Queller, D.C., and Thompson, C.R.L. (2004). Pleiotropy as a mechanism to stabilize cooperation. Nature *431*, 693-696.

69.   Foster, K.R. (2011). The sociobiology of molecular systems. Nat. Rev. Genet. *12*, 193-203.

70.   Niklas, K.J. (2014). The evolutionary-developmental origins of multicellularity. Am. J. Bot. *101*, 6-25.

71.   Rainey, P.B., and De Monte, S. (2014). Resolving conflicts during the evolutionary transition to multicellular life. Annu Rev Ecol Evol S *45*, 599-620.

72.   Fisher, R.M., Cornwallis, C.K., and West, S.A. (2013). Group formation, relatedness, and the evolution of multicellularity. Curr Biol *23*, 1120-1125.

73.   DeGregori, J. (2013). Challenging the axiom: does the occurrence of oncogenic mutations truly limit cancer development with age? Oncogene *32*, 1869.

74.   Michod, R.E., and Roze, D. (2001). Cooperation and conflict in the evolution of multicellularity. Heredity (Edinb) *86*, 1-7.

75.   Evan, G.I., and Vousden, K.H. (2001). Proliferation, cell cycle and apoptosis in cancer. Nature *411*, 342-348.

76.   Singh, M., and Boomsma, J.J. (2015). Policing and punishment across the domains of social evolution. Oikos *124*, 971-982.

77.   Abegglen, L.M., Caulin, A.F., Chan, A., Lee, K., Robinson, R., Campbell, M.S., Kiso, W.K., Schmitt, D.L., Waddell, P.J., Bhaskara, S., et al. (2015). Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. Journal of the American Medical Association *314*, 1850-1860.

78.   Sulak, M., Fong, L., Mika, K., Chigurupati, S., Yon, L., Mongan, N.P., Emes, R.D., and Lynch, V.J. (2016). TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. elife *5*, :e11994.

79. Hogan, C., Dupré-Crochet, S., Norman, M., Kajita, M., Zimmermann, C., Pelling, A.E., Piddini, E., Baena-López, L.A., Vincent, J.-P., and Itoh, Y. (2009). Characterization of the interface between normal and transformed epithelial cells. Nature cell biology *11*, 460.

80. Schreiber, R.D., Old, L.J., and Smyth, M.J. (2011). Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. Science *331*, 1565-1570.

81. Feng, M., Marjon, K.D., Zhu, F., Weissman-Tsukamoto, R., Levett, A., Sullivan, K., Kao, K.S., Markovic, M., Bump, P.A., and Jackson, H.M. (2018). Programmed cell removal by calreticulin in tissue homeostasis and cancer. Nature communications *9*, 3194.

82. Morris, K., Austin, J.J., and Belov, K. (2013). Low major histocompatibility complex diversity in the Tasmanian devil predates European settlement and may explain susceptibility to disease epidemics. Biol. Lett. *9*, 20120900.

83. Epstein, B., Jones, M., Hamede, R., Hendricks, S., McCallum, H., Murchison, E.P., Schonfeld, B., Wiench, C., Hohenlohe, P., and Storfer, A. (2016). Rapid evolutionary response to a transmissible cancer in Tasmanian devils. Nat. Commun. *7*, 12684.

84. Doonan, J.H., and Sablowski, R. (2010). Walls around tumours - why plants do not develop cancer. Nat. Rev. Cancer *10*, 794-802.

85. Biedermann, S., Harashima, H., Chen, P., Heese, M., Bouyer, D., Sofroni, K., and Schnittger, A. (2017). The retinoblastoma homolog RBR1 mediates localization of the repair protein RAD51 to DNA lesions in Arabidopsis. The EMBO Journal *36*, 1279-1297.

86. Cornwallis, C.K., West, S.A., and Griffin, A.S. (2009). Routes to indirect fitness in cooperatively breeding vertebrates: kin discrimination and limited dispersal. J. Evol. Biol. *22*, 2445-2457.

87. Clutton-Brock, T.H., and Parker, G.A. (1995). Punishment in animal societies. Nature *373*, 209-216.

88. Riehl, C., and Frederickson, M.E. (2016). Cheating and punishment in cooperative animal societies. Philos Trans R Soc Lond B Biol Sci *371*, 20150090.

89.   Cant, M.A., Nichols, H.J., Johnstone, R.A., and Hodge, S.J. (2014). Policing of reproduction by hidden threats in a cooperative mammal. Proc. Natl. Acad. Sci. U.S.A. *111*, 326-330.

90.   Lehmann, L., Foster, K.R., Borenstein, E., and Feldman, M.W. (2008). Social and individual learning of helping in humans and other species. Trends Ecol. Evol. *23*, 664-671.

91.   West, S.A., El Mouden, C., and Gardner, A. (2011). Sixteen common misconceptions about the evolution of cooperation in humans. Evol. Hum. Behav. *32*, 231-262.

92.   Whitehouse, H., McQuinn, B., Buhrmester, M., and Swann, W.B., Jr. (2014). Brothers in arms: Libyan revolutionaries bond like family. Proc. Natl. Acad. Sci. U.S.A. *111*, 17783-17785.

93.   Hammerstein, P., and Noe, R. (2016). Biological trade and markets. Philos T R Soc B *371*, 20150101.

94.   Noë, R., and Hammerstein, P. (1994). Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. Behav. Ecol. Sociobiol. *35*, 1-11.

95.   Trivers, R.L. (1971). Evolution of reciprocal altruism. Q. Rev. Biol *46*, 35-57.

96.   Axelrod, R., and Hamilton, W.D. (1981). The evolution of cooperation. science *211*, 1390-1396.

97.   Nowak, M.A., and Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. Nature *393*, 573-577.

98.   Barclay, P., and Raihani, N. (2016). Partner choice versus punishment in human Prisoner's Dilemmas. Evol. Hum. Behav. *37*, 263-271.

99.   Úbeda, F., and Duéñez-Guzmán, E.A. (2011). Power and corruption. Evolution: International Journal of Organic Evolution *65*, 1127-1139.

100.   Raihani, N.J., and Bshary, R. (2015). The reputation of punishers. Trends in ecology & evolution *30*, 98-103.

101.   Hughes, W.O., Oldroyd, B.P., Beekman, M., and Ratnieks, F.L. (2008). Ancestral monogamy shows kin selection is key to the evolution of eusociality. Science *320*, 1213-1216.

102.	Boomsma, J.J. (2009). Lifetime monogamy and the evolution of eusociality. Philos T R Soc B *364*, 3191-3207.

103.	Ratnieks, F.L., and Wenseleers, T. (2008). Altruism in insect societies and beyond: voluntary or enforced? Trends Ecol Evol *23*, 45-52.

104.	Ratnieks, F.L.W., Foster, K.R., and Wenseleers, T. (2006). Conflict resolution in insect societies. Annu Rev Entomol *51*, 581-608.

105.	Wenseleers, T., and Ratnieks, F.L.W. (2006). Enforced altruism in insect societies. Nature *444*, 50.

106.	Monnin, T., Ratnieks, F.L.W., Jones, G.R., and Beard, R. (2002). Pretender punishment induced by chemical signalling in a queenless ant. Nature *419*, 61.

107.	Wenseleers, T., and Ratnieks, F.L.W. (2004). Tragedy of the commons in Melipona bees. Proc. R. Soc. B *271*, 310-312.

108.	Bronstein, J.L. (2015). Mutualism, First edition. Edition, (Oxford, United Kingdom: Oxford University Press).

109.	Foster, K.R., and Wenseleers, T. (2006). A general model for the evolution of mutualisms. J. Evol. Biol. *19*, 1283-1293.

110.	Hartmann, A.C., Baird, A.H., Knowlton, N., and Huang, D. (2017). The Paradox of Environmental Symbiont Acquisition in Obligate Mutualisms. Current Biology *27*, 3711-3716. e3713.

111.	Jones, E.I., Afkhami, M.E., Akcay, E., Bronstein, J.L., Bshary, R., Frederickson, M.E., Heath, K.D., Hoeksema, J.D., Ness, J.H., Pankey, M.S., et al. (2015). Cheaters must prosper: reconciling theoretical and empirical perspectives on cheating in mutualism. Ecol Lett *18*, 1270-1284.

112.	Goulson, D. (1999). Foraging strategies of insects for gathering nectar and pollen, and implications for plant ecology and evolution. Perspectives in plant ecology, evolution and systematics *2*, 185-209.

113.	Fischer, M.K., Hoffmann, K.H., and Völkl, W. (2001). Competition for mutualists in an ant–homopteran interaction mediated by hierarchies of ant attendance. Oikos *92*, 531-541.

114. Jandér, K.C., and Herre, E.A. (2010). Host sanctions and pollinator cheating in the fig tree–fig wasp mutualism. Proceedings of the Royal Society of London B: Biological Sciences *277*, 1481-1488.

115. Bshary, R., and Schäffer, D. (2002). Choosy reef fish select cleaner fish that provide high-quality service. Animal Behaviour *63*, 557-564.

116. Kiers, E.T., Rousseau, R.A., West, S.A., and Denison, R.F. (2003). Host sanctions and the legume-rhizobium mutualism. Nature *425*, 78-81.

117. Zhang, M.M., Poulsen, M., and Currie, C.R. (2007). Symbiont recognition of mutualistic bacteria by Acromyrmex leaf-cutting ants. The ISME journal *1*, 313-320.

118. Aanen, D.K., de Fine Licht, H.H., Debets, A.J.M., Kerstes, N.A.G., Hoekstra, R.F., and Boomsma, J.J. (2009). High Symbiont Relatedness Stabilizes Mutualistic Cooperation in Fungus-Growing Termites. Science *326*, 1103-1106.

119. Visick, K.L., Foster, J., Doino, J., McFall-Ngai, M., and Ruby, E.G. (2000). Vibrio fischeri lux genes play an important role in colonization and development of the host light organ. J Bacteriol *182*, 4578-4586.

120. Schwartzman, J.A., and Ruby, E.G. (2016). A conserved chemical dialog of mutualism: lessons from squid and vibrio. Microbes Infect. *18*, 1-10.

121. Ratnieks, F.L.W. (1988). Reproductive harmony via mutual policing by workers in eusocial hymenoptera. Am Nat *132*, 217-236.

122. Cant, M.A. (2010). The role of threats in animal cooperation. Proceedings of the Royal Society B: Biological Sciences.

123. West, S.A., Kiers, E.T., Simms, E.L., and Denison, R.F. (2002). Sanctions and mutualism stability: why do rhizobia fix nitrogen? Proceedings of the Royal Society of London B: Biological Sciences *269*, 685-694.

124. Foster, K.R., and Kokko, H. (2006). Cheating can stabilize cooperation in mutualisms. Proceedings of the Royal Society of London B: Biological Sciences *273*, 2233-2239.

125. Frank, S.A. (1995). Mutual policing and repression of competition in the evolution of cooperative groups. Nature *377*, 520-522.

126. Foster, K.R. (2004). Diminishing returns in social evolution: the not-so-tragic commons. Journal of evolutionary biology *17*, 1058-1072.

127. Martin, S.J., Beekman, M., Wossler, T.C., and Ratnieks, F.L.W. (2002). Parasitic Cape honeybee workers, Apis mellifera capensis, evade policing. Nature *415*, 163.

128. Montague, C.E., and Oldroyd, B.P. (1998). The Evolution of Worker Sterility in Honey Bees: An Investigation into a Behavioral Mutant Causing Failure of Worker Policing. Evolution *52*, 1408-1415.

129. Beekman, M., Komdeur, J., and Ratnieks, F.L.W. Reproductive conflicts in social animals: who has power? Trends in Ecology & Evolution *18*, 277-282.

Supplementary Materials for:

Enforcement is central to the evolution of cooperation

J. Arvid Ågren[1,2,*], Nicholas G. Davies[3*], and Kevin R. Foster[4,5]

1. Department of Molecular Biology and Genetics, Cornell University

2. Department of Organismic and Evolutionary Biology, Harvard University

arvid_agren@fas.harvard.edu

3. Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine

nicholas.davies@lshtm.ac.uk

4. Department of Zoology, University of Oxford

kevin.foster@zoo.ox.ac.uk

5. Department of Biochemistry, University of Oxford

* These authors contributed equally

Contents:

Definition of enforcement (summary of Box 1)

We define enforcement as an action that evolves, at least in part, to reduce selfish behaviour within a cooperative alliance. Individuals can use enforcement in two key ways: either to change the strength of interaction with different partners (partner choice; Table 1; Box 2), or to manipulate the amount that their partners display selfish behaviours rather than cooperative ones (partner manipulation; Table 1; Box 2). We distinguish between the selfishness of group members, z, and the selfishness experienced by a focal individual y. In the models below, these are sometimes equivalent but, where they are not, the key measure of selfishness is that experienced by an enforcer y (plotted in Figure 3), which captures any effects of enforcement from the enforcer's perspective (the key to the evolution of enforcement, see next section).

Consider a focal actor who experiences a level of selfishness, y, from others that is a function of that actor's investment in enforcement, a, such that $y = y(a)$. Here, $a = 0$ represents no enforcement and $a > 0$ represents enforcement. If investing in enforcement (increasing a) reduces the selfishness experienced by the focal actor (decreases $y(a)$) then the focal actor is enforcing cooperation. Specifically, in the absence of enforcement ($a = 0$), define the focal actor's strength of interaction with individual i of n partners as $\beta_i = \beta_i(0)$ and individual i's level of selfishness as $z_i = z_i(0)$. The β_i function quantifies how much an individual's selfishness matters to the focal actor's fitness; for example, when the focal individual interacts with all of its n partners equally, we would have $\beta_i = \frac{1}{n}$. When the actor invests in enforcement ($a > 0$), these change to $\beta_i(a)$ and $z_i(a)$. Formally, the focal actor's experienced selfishness can be written without enforcement as

$$y(0) = \sum_i^n \beta_i(0)z_i(0) \tag{S1}$$

and with enforcement as

$$y(a) = \sum_i^n \beta_i(a)z_i(a). \tag{S2}$$

We are interested in the evolution of selfish traits and their suppression by enforcement. A trait z is selfish if, in the absence of enforcement, a marginal increase in the trait increases its fitness w (i.e., $dw/dz > 0$) but decreases the fitness of other individuals w' (i.e., $dw'/dz < 0$). Meanwhile, a trait a causes the enforcement of cooperation if a marginal increase the trait value decreases experienced selfishness (i.e. $dy/da < 0$). The evolution of enforcement can occur because of its effects on selfishness, or due to other effects unrelated to these effects (byproduct enforcement; Box S1). We are most interested in the former cases where the focal trait a evolves, at least in part, *because* of its effects on selfishness within a cooperative alliance, which implies that $dw/dy < 0$ and $dy/da < 0$, and by transitive logic $dw/da > 0$. Finally, we assume in the models below that enforcement carries a cost, C. For simplicity, we phrase the above in terms of the direct benefits to the enforcer, but enforcement can also evolve due to indirect fitness benefits, as we discuss next.

Direct and indirect benefits of enforcement

The evolution of enforcement, like cooperation, rests upon there being either direct benefits to an enforcer (via improved lifetime reproduction), or indirect fitness benefits via improved reproduction of the enforcer's relatives (Box 2). Examples of direct benefits include a plant that limits the access of nectar robbing insects, but not pollinators, using complex floral morphology[1] (Model 6, below). Another example is reciprocal altruism in humans, which uses cooperation to make other reciprocators cooperate back[2]. Direct benefits can also come via effects on the whole social group[3]. For example, the evolution of a silencing allele that reduces the transposition rate of transposable elements will benefit the suppressor because it increases genome-level fitness[4]. Here, the benefit from enforcement is direct because it falls back on the allele that silences the elements. An example of indirect benefits driving enforcement is the evolution of distinct queen cells in corbiculate bees, which allows workers to control which larvae become queens and thereby prevents caste conflict[5-7]. This requires relatedness within the colony, such that there is an indirect fitness benefit to the non-reproductive workers from improving colony productivity.

With relatedness, enforcement can also occur via an allele that reduces the expression of selfishness in carriers (self-enforcement). For example, there is a tight pleiotropic link between proliferation and apoptosis in the regulatory networks of mammalian cells, which means many potentially cancerous cells simply enter programmed cell death[8, 9]. For this to be favoured by natural selection, clonemates have to benefit from the suicide of a pre-cancerous cell (Model 3). This then satisfies the condition that an individual carrying an enforcement allele (a) will experience reduced selfishness, i.e. $dy/da < 0$ because group members that did not undergo apoptosis will also tend to carry the allele, which allows $da/dw > 0$ to be satisfied (above).

In the models of enforcement that follow, we analyse examples of enforcement involving both direct and indirect benefits to the enforcer. To account for indirect benefits of enforcement, experienced selfishness $y(a)$ could be defined separately for each class of relatives and then prorated by genetic relatedness as appropriate. The models analysing indirect benefits of enforcement below (Models 1 and 4) simply assume that each relative in a cooperative group experiences the same level of selfishness, but this can be extended to models where different relatives may each have a separate experienced level of selfishness.

Models of Enforcement

We develop a series of models to study the evolution of enforcement. Our goal is to identify the broad conditions that favour, or disfavour, the evolution of enforcement and thereby assess its generality as an explanation for cooperation. In order to preserve key biological details, we do not seek to reduce the problem to a single general model. Instead, we develop one or more dedicated models at each level of biological cooperation (Figures 1 and 2; Table S1). The examples also cover the different mechanisms of enforcement (partner choice versus manipulation; Table 1) as well as enforcement that benefits the enforcer (direct fitness) and cases that benefit relatives

Table S1 Summary of the models

Model	Level[a]	Mechanism[b]	Benefit	Relatedness[c]
Replicators in a protocell	Genome	manipulation	Direct/Indirect	yes
Endosymbionts of isogamous cells	Complex cell	manipulation	Direct	no
Early multicellularity and cancer	Organism	choice	Indirect	yes
Reproductive conflict in animals	Society	manipulation	Direct/Indirect	yes
Host-symbiont mutualism	Mutualism	choice	Direct	no
Plant-pollinator mutualism	Mutualism	manipulation	Direct	no

a) Level of biological organisation (Figure 1). b) Mechanism of enforcement (Box 1, Table 1). c) Relatedness between enforcer and enforcees is possible in the model.

(indirect fitness). The evolution of enforcement has the potential to drive the evolution of novel selfish strategies that escape a particular form of enforcement[10, 11], followed by the evolution of new more-effective enforcement, and so on[12]. While we do not consider this process explicitly, our models describe the conditions for selfishness and enforcement for any one round of this process. Finally, though the study of enforcement has often focused on derived systems, particularly in the social insects[13-16], our examples often reflect an early stage in the evolution of a given social system. We do this to challenge the notion that enforcement is only important in the maintenance, and not the origin, of cooperative systems.

In the models that follow, we notate individual investment in enforcement and selfishness as a and z, respectively; group mean investment in these traits as A and Z; population mean investment in these traits as $\bar{a}$ and $\bar{z}$; and C as the cost of enforcement.

Model 1: Replicators in a protocell (and review of Frank's policing model)

We first consider a model of a simple group of replicating molecules in an early cell (Figure 2a)[17]. In the absence of investment into selfishness (z), we assume that replicators are transmitted fairly each cell division. However, replicators can attract a catalytic molecule (equivalent to a retrotransposase in modern systems) with affinity z to afford them an additional duplication per cell division at the cost of other replicators. Additionally, we assume that a replicator can spend a fraction a of its time bound to the catalytic molecule without undergoing replication, which limits the potential for other replicators to replicate selfishly, at cost Ca to itself; such that when $A = 1$, no selfish replication can occur. This simple scenario can be captured with a modification of Frank's classic model of mutual policing[18]. We therefore first review Frank's model before we modify and apply it to the case of replicators.

Mutual policing is a type of enforcement by manipulation in which the enforcers are themselves also the targets of enforcement. Therefore, a focal individual expressing selfishness z and

4

enforcement *a* belongs to a larger group expressing average selfishness *Z* and average enforcement *A*. From Frank[18], an individual's fitness is then

$$w = (-Ca + A + (1 - A)\frac{z}{Z})(1 - (1 - A)Z) \ , \qquad (S3)$$

where *C* is the individual marginal cost of enforcement, accounting for the *-Ca* term. A proportion *A* of all social interactions are policed. The fruits of a policed interaction are divided fairly among all members of the group, while the fruits of an unpoliced interaction accrue to each individual proportionally to their relative selfishness, z/Z. This accounts for the *A + (1-A)z/Z* term. Finally, the group as a whole experiences a loss of cooperation-derived fitness due to any unpoliced selfishness, which accounts for the 1 - (1 - *A*)*Z* term.

Frank found that the equilibrium level of selfishness in the absence of enforcement was *z** = 1 - *R*, where *R* is the average group relatedness. Since there is no enforcement, we also have the equilibrium experienced selfishness *y** = *z**. If we allow enforcement to evolve, we have *a** = 1 and *y** = 0 if *C* < 1 - *R*, and *a** = 0 and *y** = *z** = 1 - *R* if *C* > 1 - *R*. That is, full policing evolves if the cost of enforcement is less than the equilibrium selfishness prior to enforcement *z** = 1 - *R*, while if the cost of enforcement is less than this cost of selfishness, enforcement does not evolve.

A number of modifications have been made to Frank's basic model[19]. For example, one key assumption in the original model is that the cost of enforcement is independent of an individual's investment selfishness per se. This assumption is likely to be violated in our replicator system, where an individual's investment in enforcement by blocking the catalytic molecule will trade off with its replication proper. We can capture this tradeoff with

$$w = (A + (1 - A)\frac{z(1-Ca)}{Z(1-CA)})(1 - (1 - A)Z), \qquad (S4)$$

where $A + (1 - A)\frac{z(1-Ca)}{Z(1-CA)}$ captures the individual benefits of competition to replicators and $(1 - (1 - A)Z)$ captures the shared fitness interest ("being in the same boat", see main text) of the replicators, which is harmed by competition *Z* but improved by investment into enforcement *A*. In the absence of enforcement (*a* = *A* = 0), this is equivalent to Frank's model, meaning that equilibrium selfishness in the absence of enforcement is *z** = 1 - *R*. As for Frank's model then, without enforcement the model predicts the evolution of selfishness that will reduce cooperation and compromise the function of the protocell. This is also consistent with the prediction that early replicator systems would be susceptible to the evolution of selfish genetic elements that invest in biasing personal replication over replication of other sequences and the cell[17, 20].

We can then ask whether enforcement, *a*, will increase from zero. This will occur when

$$\frac{\partial w}{\partial a}|_{a=A=0,z=Z=1-R} + R\frac{\partial w}{\partial A}|_{a=A=0,z=Z=1-R} > 0, \qquad (S5)$$

which eventually simplifies to

5

$C < 1$.

Moreover, we find that when $C = 1$, any nonzero enforcement ($a > 0$) makes enforcement easier to evolve, meaning that any unilateral expression of enforcement can select for group enforcement. As the system evolves, we find that $a^* = 1$ and $y^* = 0$ for all $C \leq 1$, meaning that policing always becomes absolute and reduces experienced selfishness to zero under these conditions. That is, for our scenario – where the cost of policing is in terms of an individual's ability to selfishly replicate – policing is easier to evolve than under Frank's original model (Figure 3a vs 3b). Evolution then is cost dependent but so long as policing is equally or less costly per unit investment than replication, it is predicted that enforcement removes all conflict (Figure 3b).

Note that, in this case, the evolution of enforcement requires a group benefit to succeed, which is exemplified by the shared fitness term $(1 - (1 - A)Z)$ of equation (S4). Unless enforcement improves group function, it does not evolve. Specifically, if we remove this effect from the fitness function above, yielding the alternative form

$$w = (A + (1 - A)\frac{z(1-Ca)}{Z(1-CA)})(1 - Z), \tag{S6}$$

the condition for natural selection to favour an increase in enforcement becomes $R > 1$, which can never be fulfilled. This formalizes the intuition that while "being in the same boat" is insufficient to prevent conflict, it can create a shared collective interest that allows highly-effective enforcement to evolve (see main text).

Relation to definitions of selfishness and enforcement — In the model above, affinity for the catalytic molecule, z, meets the definition of selfishness. That is because, in the absence of enforcement, $w = \frac{z}{Z}(1 - Z)$, and hence an increase in z improves the fitness of a focal replicator ($\frac{dw}{dz} = \frac{1-Z}{Z}$, and so $\frac{dw}{dz} > 0$ when $0 < Z < 1$), while an increase in Z decreases the expected fitness of any replicator ($\frac{dw}{dZ} = -\frac{z}{Z^2}$, and so $\frac{dw}{dZ} < 0$), which fulfils the conditions for selfishness.

Interference with the catalytic molecule, a, meets the definition of enforcement because it reduces experienced selfishness and hence restores fitness. Specifically, experienced selfishness is $y = (1 - A)Z$, so $\frac{dy}{dA} < 0$ when $Z > 0$. Moreover, $w = (A + (1 - A)\frac{z(1-Ca)}{Z(1-CA)})(1 - y)$, and so $\frac{dw}{dy} < 0$. Together, these fulfil the conditions for enforcement.

Model 2: Endosymbionts of isogamous cells

We next consider a well-mixed population of single-celled eukaryotic diploid organisms that carry endosymbiotic bacteria, such as early mitochondria or chloroplasts. Sexual reproduction was present in the last universal common ancestor of eukaryotes[21]. We therefore assume the host organisms occasionally undergo meiosis to form haploid gametes of two mating types (m_1 and m_2), where gametes are assigned a mating type at random (Figure 2b). As for the protocell, we

first consider the potential for endosymbiont selfishness to evolve, before secondly asking whether enforcement will evolve to suppress any selfishness that arises.

The scope for selfish symbionts is as follows. When two gametes fuse, the endosymbionts carried by one of the two gametes stand to gain a twofold transmission advantage if they can eliminate the endosymbionts of the other gamete, e.g. via one of the toxin-immunity protein pairs that are widespread in bacteria[22]. To analyse this case, we use the logic of classical game theory before performing simulations to support our conclusions. Suppose that the population-level frequency of endosymbionts engaging in selfish behaviour is $\bar{z}$; endosymbionts engaging in killing (killer type) can spread if the fitness cost of endosymbiont warfare, c_{war}, is less than 1/2, since costs greater than 1/2 cancel the twofold transmission benefit of endosymbiont warfare. This cost may capture both the effects of coping with a half-complement of endosymbionts as well as the cost of synthesizing toxins for endosymbiont warfare.

When gametes carrying killer and passive endosymbionts fuse, the killer symbionts are retained and the resulting zygote has fitness $1 - c_{war}$. When killer symbionts mix, both sets are killed and the zygote has negligible fitness. Despite the high cost of conflict when killers meet, natural selection will promote the killer phenotype so long as $\bar{z} < 1$, since passive endosymbionts always stand to gain from adopting the killer phenotype. This prediction occurs because both phenotypes are killed upon encountering a killer; but as long as there are passives left in the population, a killer stands to gain from its strategy. A simple game theory model then predicts that there is the potential for strong conflict between endosymbionts that will be very costly for cellular function (Figure 3b), which is consistent with the predictions of more extensive population genetic analysis[23].

We next consider the potential for enforcement to affect the outcome of this conflict. Specifically, we consider the evolution of a nuclear modifier (a_+) that degrades the DNA of its endosymbionts shortly before or shortly after gametic fusion, conditional on the gamete's mating type. This is consistent with the known biology of isogamous unicellular species, such as *Chlamydomonas reinhardtii* where with an hour of mating, the DNA in chloroplasts of one mating type is selectively degraded before the chloroplasts lacking DNA fuse with those of the other mating type[24]. Without loss of generality, we assume the uniparental phenotype is only expressed in m_1 gametes. This phenotype prevents endosymbiont conflict at cost C to the resulting zygote. Degrading endosymbiont DNA in this way may carry a relatively low cost as it occurs without destroying the organelles themselves. This suggests that $C \leq c_{war}$, i.e. a cell that sacrifices one set of endosymbionts in an energy-efficient manner will do so with lower cost to the cell than a killer endosymbiont strain using toxins to kill a competing passive strain. In the worst-case scenario, an m_1 gamete can sacrifice its endosymbionts completely, which ought to be no costlier for the cell than it is for killer endosymbionts to destroy opposing endosymbionts, i.e. at most $C \approx c_{war}$. If true, enforcement results in the zygote having an expected fitness of $1 - C$ rather than 0, and so enforcement should always invade in a population approaching complete endosymbiont selfishness ($\bar{z} \to 1$).

To verify this prediction, we simulate mating in a very large population of gametes with non-overlapping generations, which come in eight types, corresponding to mating type / endosymbiont type / enforcement type. We track the frequencies of the eight types: m_1/passive/a_-, m_1/passive/a_+, m_1/killer/a_-, m_1/killer/a_+, m_2/passive/a_-, m_2/passive/a_+, m_2/killer/a_-, and m_2/killer/a_+. When two gametes fuse, the fitness of the resulting zygote can be read from the following table:

<table>
<tr><td colspan="2"></td><td colspan="2">No enforcement by + gamete</td><td></td><td colspan="2"></td><td colspan="2">Enforcement by + gamete</td></tr>
<tr><td colspan="2"></td><td colspan="2">m_1 gamete</td><td></td><td colspan="2"></td><td colspan="2">m_1 gamete</td></tr>
<tr><td colspan="2"></td><td>passive</td><td>killer</td><td></td><td colspan="2"></td><td>passive</td><td>killer</td></tr>
<tr><td rowspan="2">m_2 gamete</td><td>passive</td><td>1</td><td>$1-c_{\text{war}}$</td><td rowspan="2">➔</td><td rowspan="2">m_2 gamete</td><td>passive</td><td>$1-C$</td><td>$1-C$</td></tr>
<tr><td>killer</td><td>$1-c_{\text{war}}$</td><td>0</td><td>killer</td><td>$1-C$</td><td>$1-C$</td></tr>
</table>

After all gametes fuse into zygotes, they undergo viability selection (zygotes survive with probability proportional to their fitness w) and then undergo meiosis again, creating new gametes. Zygotes with both killer and passive endosymbionts pass on only the killers, while passive-passive zygotes pass on only passive endosymbionts. Gametes are randomly assigned a mating type, and mutation causes gametes to change their phenotype from killer to passive during meiosis with probability $\mu = 0.001$.

As predicted from game theory logic, we find that selfishness (the killer phenotype) sweeps to near-fixation in this scenario for any $c_{\text{war}} < 1/2$, and that when we allow enforcement (a_+ phenotype) to evolve, it also sweeps to fixation and causes realised selfishness (the frequency of gametic pairings resulting in endosymbiont warfare) to drop to zero for any cost of enforcement $C < 1$ (Figure 3c).

Relation to definitions of selfishness and enforcement — The killer gamete phenotype meets the definition of selfishness when $c_{war} < 1/2$ because the killer phenotype benefits gametes: in the absence of enforcement, a gamete's expected fitness is $w = (1-z)(1-\bar{z}) + 2z(1-\bar{z})(1-c_{war})$, so $\frac{dw}{dz} = -(1-\bar{z}) + 2(1-\bar{z})(1-c_{war})$, which is positive when $c_{war} < 1/2$; and $\frac{dw}{d\bar{z}} = -(1-z) - 2z(1-c_{war})$, which is always negative. Accordingly, when $c_{war} < 1/2$, expressing the killer phenotype benefits a focal gamete ($\frac{dw}{dz} > 0$) and harms other gametes ($\frac{dw}{d\bar{z}} < 0$), fulfilling the condition for selfishness.

The a_+ nuclear modifier meets the definition of enforcement. With enforcement, an m_1 gamete's expected fitness is $w = (1-a)[(1-z)(1-\bar{z}) + 2z(1-\bar{z})(1-c_{war})] + a(1-C)$, which can be rewritten $w = (1-a-y)[(1-z) + 2z(1-c_{war})] + a(1-C)$, where $y = (1-a)\bar{z}$ is the gamete's experienced selfishness, and the term in square brackets is the gamete's expected fitness in the absence of any enforcement or selfishness by others. Note that experienced selfishness y reduces this fitness component when the focal gamete is not an enforcer ($a = 0$), but has no impact upon fitness when the focal gamete is an enforcer ($a = 1$, which also makes $y = 0$). Since $\frac{dy}{da} < 0$ when $\bar{z} > 0$, enforcement decreases experienced selfishness.

Model 3: Early multicellularity and cancer

Here we consider the evolution of cancer suppression in a simple multicellular animal (Figure 2c). In some cases, like the facial tumour in the Tasmanian devil, a cancer that arises within one individual can be transmitted onto others[25]. However, the majority of cancers are not transmitted and, as such, are short term evolutionary processes in the sense that they begin and end within the life of a multicellular organism. Nevertheless, cancer formally corresponds to the *de novo* evolution of a selfish phenotype within an organism that – within the timescale of one host generation – spreads for its own benefit but at a cost at the level of the organism[26-30]. Moreover, phenotypes that suppress cancer and restore multicellular function correspond to the definition of enforcement against selfish phenotypes[26-30] (above).

Our model assumes that individuals start development as a single, non-cancerous cell, but that there is a per-cell-division risk of mutating to a cancer cell. For simplicity, we focus on those cancers in an organism that are ultimately fatal as these are the ones that are expected to have the greatest impact on multicellular function. To derive an organism of n cells from a single cell requires $n - 1 \approx n$ divisions. If the per-cell-division risk of cancer is μ, then the lifetime risk of cancer for an n-celled animal is

$$z = 1 - (1 - \mu)^{n-1} \approx 1 - exp(-\mu n). \qquad \text{(S8)}$$

For simplicity, the amount of selfishness z in the system is the per-individual risk of getting cancer, which differs from our other models in that z is not expressed in terms of selfishness enacted at the lower level of biological organization, which would be per-cell selfishness. However, the two are equivalent in the sense that reducing z will also reduce the selfishness experienced per cell y. For fatal cancers, an individual's expected fitness is

$$w = 1 - z = exp(-\mu n). \qquad \text{(S9)}$$

We then consider the potential evolution of self-enforcement whereby a fraction $A = a$ of cancerous cells undergo apoptosis, which is consistent with the well-known link between cell proliferation and apoptosis in animals[8, 9, 31]. Then the effective rate of cancer mutation is $(1 - A)\mu$, experienced selfishness is $1 - exp(-\mu(1 - A)n)$, and fitness is $w = (1 - CA)exp(-\mu(1 - A)n)$, where C is the cost of enforcement.

Natural selection will promote a small increase in enforcement A if

$$dw/dA > 0 \qquad \text{(S10)}$$

$$(-C)(exp(-\mu(1 - A)n) + (1 - CA)(exp(-\mu(1 - A)n)n\mu > 0 \qquad \text{(S11)}$$

$$n\mu > \frac{C}{1 - CA} . \qquad \text{(S12)}$$

9

This last line can be interpreted as follows: Natural selection will increase enforcement if the per-individual rate of cancer mutations ($n\mu$) exceeds the marginal relative cost of enforcement ($\frac{C}{1-CA}$).

As expected then, the evolution of self-enforcement in cancerous cells will depend on the size of organism, because large organisms will, all else being equal, experience more cancerous mutations[30, 32]. However, an organism does not need to be very large to benefit from enforcement. From inequality S9, we can see that self-enforcement becomes important once n gets of the order $1/\mu$. Eukaryotes exhibit about 3×10^{-3} coding-sequence mutations per genome per cell division[33]. However, the proportion of mutations that lead to a cancer phenotype in early organisms is unknown. Nevertheless, by comparison with modern single-celled organisms, many routes to increasing cell division rate within an organism are expected, such as a loss-of-function mutations in pathways that generate cooperative secretions[34]. To give an example, if we assume that one in a thousand nonsense mutations led to cancer phenotypes in early multicellular organisms, and a given point mutation causes a nonsense mutation with approximate probability 0.038 (ref. [35]), then the per-cell-division rate of cancer mutations is μ = $3\times10^{-3} \cdot 1\times10^{-3} \cdot 0.038$ = 1.1×10^{-7}, and our model predicts that enforcement will be important for organisms of on the order of millions of cells or above (c.f. 10^{13} cells in the human body[36]). But whatever the exact ancestral mutation rate, whenever cancer is prevalent enough to damage multicellular function, the prediction is that self-enforcement will rapidly evolve that restores cooperation (Figure 3d).

Relation to definitions of selfishness and enforcement — Cancer meets the definition of selfishness because, at the scale of selection between cells in a developing individual, the fitness of an individual cell is positively correlated with the cancer phenotype (the unchecked proliferation and growth of cancer cells are what make them cancerous by definition). Although we do not model this process explicitly here, the increase in frequency of mutant cells during cancer evolution is a form of natural selection within individuals[26-30].

Apoptosis A meets the definition of enforcement because fitness can be written $w = (1 - CA)(1 - y)$, where experienced selfishness y is $1 - exp(-\mu(1 - A)n)$. Hence, we have both $\frac{dw}{dy} < 0$ (cancer is harmful to the multicellular organism) and $\frac{dy}{dA} = -\mu n \exp(-\mu(1 - A)n) < 0$ (enforcement decreases the effect of selfishness on the group).

Model 4: Reproductive conflict in an animal group

We next consider the evolution of enforcement in a non-eusocial animal group, akin to a cooperatively breeding vertebrate like meerkats (Figure 2d). Individuals can invest in aggression that gains them a greater share of group resources, and thereby reproduction, but at a cost to group productivity. They also have the possibility of investing in enforcement that punishes selfish individuals in the group. As for the case of replicators, we capture this scenario with a modification of Frank's mutual policing model because enforcers are again also subject themselves to enforcement[19]. Specifically, we make two changes to Frank's model to better fit punishment in an animal group: 1) punishment incurs a direct cost for the selfish individual being policed; and 2)

an individual's investment in punishment only affects others in the group, rather than also affecting themselves (individuals do not self-police). Under these conditions, an individual's fitness is

$$w = (-Ca - c_{\text{pun}}\alpha z + A + (1-\alpha)\frac{z}{Z})(1-(1-A)Z), \tag{S13}$$

where c_{pun} is the cost of being punished and α is the average level of enforcement among individuals in the group *other* than the focal individual.

We calculate group relatedness R from the parameters ρ (relatedness to other members other than oneself) and n (total group size). This average relatedness to the entire group then is

$$R = \frac{n-1}{n}\rho + \frac{1}{n} \tag{S14}$$

As in the replicator model, equilibrium selfishness in the absence of enforcement is $z^* = 1 - R$ in this model. Enforcement of others will increase from zero when

$$\frac{\partial w}{\partial a}|_{a=\alpha=A=0,z=Z=1-R} + \rho\frac{\partial w}{\partial \alpha}|_{a=\alpha=A=0,z=Z=1-R} + R\frac{\partial w}{\partial A}|_{a=\alpha=A=0,z=Z=1-R} > 0, \tag{S15}$$

which simplifies to

$$-C - (c_{\text{pun}}z + 1)\rho + (1+\frac{z}{1-z})R > 0. \tag{S16}$$

The term $-C$ captures the individual's cost of enforcement; the term $-(c_{\text{pun}}z+1)\rho$ captures the indirect-fitness cost of punishing relatives, $c_{\text{pun}}z\rho$, and the indirect cost of reducing relatives' selfishness, ρ. The term $(1+\frac{z}{1-z})R$ captures the benefit R of the group sharing group resources due to policing and the benefit $\frac{z}{1-z}R$ of reduced degradation of the environment due to less expressed selfishness. Note that when c_{pun} = 0 and $\rho = R$, this recovers Frank's condition for the evolution of altruism, $C < z$. In our model, natural selection for enforcement is higher in smaller groups because policing impacts upon fewer relatives in the group. However, increasing relatedness to the other group members ρ decreases the benefits of enforcement because of the negative affect of punishing relatives (Figure 3e). And, as for the other models, increases to the cost of punishment c_{pun} also reduce natural selection for enforcement. The result of the additional punishment cost c_{pun} is that there is enforcement is less favoured than in Frank's original model (Figure 3a vs 3e), and intermediate levels of enforcement ($0 < a < 1$) are stable. However, for a low cost C of enforcement, the result is again highly effective enforcement and little impact of selfishness on a group.

Relation to definitions of selfishness and enforcement — The trait z meets the definition of selfishness. That is because, in the absence of enforcement, $w = \frac{z}{Z}(1-Z)$, and hence an increase in z improves the fitness of a focal animal ($\frac{dw}{dz} = \frac{1-Z}{Z}$, and so $\frac{dw}{dz} > 0$ when $0 < Z < 1$),

while an increase in Z decreases the expected fitness of all group members ($\frac{dw}{dZ} = -\frac{z}{Z^2}$, and so $\frac{dw}{dZ} < 0$), which fulfils the conditions for selfishness.

The trait a meets the definition of enforcement because it reduces experienced selfishness and hence restores fitness. Specifically, experienced selfishness is $y = (1 - A)Z$, so $\frac{dy}{dA} < 0$ when $Z > 0$. Moreover, $w = (-Ca - c_{\text{pun}}\alpha z + A + (1 - \alpha)\frac{z}{z})(1 - y)$, and so $\frac{dw}{dy} < 0$. Together, these fulfil the conditions for enforcement.

Model 5: Mutualism, partner choice in host-symbiont systems

Many host-symbiont systems have large amounts of horizontal transmission whereby symbionts can move between hosts, and often are picked up *de novo* each generation[37]. Such transmission introduces a large potential for conflict, both between different symbionts and between the host and symbionts. The mammalian microbiome is one such system[38], but we motivate our model here on one of the best-studied examples of horizontally acquired symbionts, the mutualism between the bobtail squid *Euprymna scolopes* and luminescent *Vibrio fischeri* bacteria[39, 40] (Figure 2e). The bacteria are housed in a special light organ by the squid that provides them with nutrients. Light produced by the bacteria is thought to help the squid to limit its moonlight shadow at night, aiding in both predation and avoiding predation. However, dark mutants of *V. fischeri* are easily observable, raising the question of what maintains the energetically costly reaction.

We assume that a bacterial strain with selfishness z has a probability z of not luminescing. The expected level of selfishness (dark phenotype) is $\bar{z}$. If a cell is luminescent it pays cost c_{lux}. There is a potential cost for the dark phenotype if it leads to the squid getting eaten by a predator, which is fatal for both host and symbiont. With high rates of horizontal transfer and mixing of genotypes (low average relatedness among *V. fischeri* cells, R), there is the potential for luminescent strains to be outcompeted within the squid light organ. However, we also assume that the squid has the option of investing in an enforcement mechanism – such as enzymes that produce toxin compounds in the absence of the light reaction[39]. This selectively harms dark strains of bacteria by subjecting them to a cost c_{host}.

The level of enforcement in the a focal squid is captured by a; the average level of enforcement among all squids is $\bar{a}$. A squid pays the cost Ca to enforce. Moreover, a squid that experiences selfishness y pays the cost $c_{\text{pred}}y$, where c_{pred} is the probability that an unilluminated squid gets eaten. The squid's experienced selfishness is $y = (1 - a)Z$, where Z is the average selfishness of *V. fischeri* cells in the squid. Accordingly, the fitness of a focal squid is

$$w_{squid} = 1 - Ca - c_{\text{pred}}(1 - a)\bar{z}, \tag{S17}$$

while the fitness of a focal *V. fischeri* cell is

12

$$w_{Vibrio} = 1 - c_{\text{lux}}(1 - z) - c_{\text{host}}z\bar{a} - c_{\text{pred}}(1 - \bar{a})Z .\qquad\text{(S18)}$$

Natural selection will promote an increase in *V. fischeri* selfishness if

$$c_{\text{lux}} - c_{\text{host}}\bar{a} - Rc_{\text{pred}}(1 - \bar{a}) > 0.\qquad\text{(S19)}$$

Above, the first term c_{lux} gives the benefit of selfishness owing to saving on investment in luminescence; the second term $-c_{\text{host}}\bar{a}$ gives the marginal risk of experiencing enforcement from the squid; and in the third term, $-Rc_{\text{pred}}(1 - \bar{a})$ is the marginal cost of the risk of being eaten.

In the absence of enforcement, *V. fischeri* selfishness will increase from zero if

$$c_{\text{lux}} > Rc_{\text{pred}} .\qquad\text{(S20)}$$

This formally captures the intuition that with high rates of horizontal transfer that leads to many strains associating with a host (low *R*), the cost of light production will lead to the loss of light producing strains. However, natural selection will promote an increase in squid enforcement if

$$c_{\text{pred}}\,\bar{z} - C > 0 .\qquad\text{(S21)}$$

That is, increased enforcement by the squid is selected for if the benefit of avoiding predation exceeds the cost of increasing enforcement. If light production is important to squid survival, therefore, enforcement is expected to evolve, even if it is costly. The prediction then is that enforcement is strongly favoured in this system so long as the benefits of mutualism are high (Figure 3f).

The model also predicts that there may be oscillations in the investment into cooperation and punishment. This mirrors an earlier model studying the coevolutionary dynamics of enforcement and cooperation in mutualisms[41]. Instabilities can occur if, once cooperation has evolved, the benefits of enforcement are reduced leaving only its cost ($\bar{z} = 0$, equation S21). The lack of natural selection for enforcement can then cause it to decline, but then cooperation also declines, which restores the benefit of, and natural selection for, enforcement, and so on. The magnitude and persistence of these dynamics in real systems remains unclear; frequent mutation to, or immigration of, less cooperative symbionts is predicted to maintain the benefits of enforcement and remove instabilities[41]. Both mutation and immigration of symbionts are expected to be important for bobtail squid which i) sample environmental bacteria each generation, many of which are not light producing and ii) carry bacteria that can become dark via a single mutation (e.g. loss of function mutation in *luxA*)[39, 40]. Nevertheless, the prediction of oscillations underlines again the potential for complex coevolutionary dynamics in cooperative systems under enforcement[12].

Relation to definitions of selfishness and enforcement — Being dark is beneficial to a focal *V. fischeri* cell in the absence of enforcement when $c_{lux} > Rc_{pred}$ (Equation S20)—under these

conditions, darkness increases the inclusive fitness of the cell. A lack of luminescence is also harmful to the squid in the absence of enforcement because $\frac{dw_{squid}}{d\bar{z}} = -c_{\text{pred}} < 0$. Hence, the dark phenotype meets the definition of selfishness.

Selective harming of dark bacteria meets the definition of enforcement because a squid's experienced selfishness is $y = (1 - a)Z$, and hence $\frac{dy}{da} < 0$ when $Z > 0$.

Model 6: Mutualism, partner manipulation in plant-pollinator interactions

The biology of enforcement is particularly clear and diverse in mutualisms. We therefore finish with a second major example from mutualisms, which corresponds to partner manipulation, rather than partner choice (Box 1; Table 1). A well-known threat to the plant-pollinator mutualism is nectar robbers that cut through the side of a flower and remove nectar without collecting pollen[1] (Figure 2f).

We assume for simplicity that pollinators visit flowers at random, attempting to selfishly rob with probability z. By robbing, the nectar-robber has a fitness bonus, as it can visit more flowers in a shorter amount of time (getting more nectar per unit time), suggesting that robbing will invade if left unchecked.

However, cutting through a plant can be time consuming; it can take twenty minutes for small insect to reach nectar this way[1]. A plant therefore can significantly inhibit robbing by thickening or toughening the flower corolla or sepals where a robber attempts its cut (or equivalently investing in the production of secondary compounds in these regions that are toxic to a robber). We assume that this investment comes at the cost of flower production. At $a = 0$ it invests fully in flowers and pollinators can always rob. At $a = 1$ it has flowers that are impossible to rob but so few flowers that the plant's reproductive fitness is negligible.

Upon attempting to rob nectar, an insect will try chewing through the plant, with probability of success $1 - a$. Each action taken by the insect incurs a cost in terms of time and energy expended. Assume that the cost of chewing through the plant is c_{rob}; if successful, the pollinator moves on. If unsuccessful, the pollinator visits the flower legitimately at cost c_{visit}. Both costs can be expressed relative to the cost of travelling between plants, which we fix at 1. A pollinator's expected fitness can then be written as

$$w_{pollinator} = z/(1 + c_{\text{rob}} + \bar{a}c_{\text{visit}}) + (1 - z)/(1 + c_{\text{visit}}) , \qquad (S22)$$

while a plant's expected fitness can be written as

$$w_{plant} = (1 - Ca)(1 - (1 - a)\bar{z}) , \qquad (S23)$$

where the plant produces a number proportional to $1 - Ca$ flowers and $1 - (1 - a)\bar{z}$ is the probability that, given a visit, a flower is legitimately visited.

14

A pollinator should increase selfishness z if

$$1/(1 + c_{\text{rob}} + \bar{a}c_{\text{visit}}) - 1/(1 + c_{\text{visit}}) > 0, \tag{S24}$$

which means

$$c_{\text{rob}} + \bar{a}c_{\text{visit}} < c_{\text{visit}},$$

that is, nectar robbing will invade when it costs less than a legitimate visit.

A plant will evolve to increase its enforcement a if

$$-C(1 - (1 - a)\bar{z}) + (1 - Ca)\bar{z} > 0 . \tag{S25}$$

That is, if the loss of fecundity from investing in enforcement (first term) is smaller than the gain in legitimate visits through enforcement (second term). More generally, we see that enforcement is again broadly predicted so long as it carries a relatively low cost ($C < 1$ in Figure 3g). However, as in the squid system, a benefit to enforcement – thickening the flower corolla or sepal – requires that there is some robbing ($\bar{z} > 0$, equation S25). As a result, oscillations also occur for some parameters (not shown) and the equilibrium level of enforcement often occurs at a point where there remains a significant level of robbing (Figure 3g). This is consistent with the observation that both robbers and legitimate pollinators often co-occur in pollination mutualisms[1].

Relation to definitions of selfishness and enforcement — Nectar robbing is beneficial to a focal pollinator in the absence of enforcement when $c_{\text{rob}} < c_{\text{visit}}$, by equation (S24), and hence under these conditions nectar robbing increases the fitness of the pollinator. Robbing is harmful to a plant in the absence of enforcement because then $w_{plant} = 1 - \bar{z}$, and so $\frac{dw_{plant}}{d\bar{z}} < 0$. Nectar robbing therefore meets the definition of selfishness.

A plant investing in reducing the efficacy of nectar robbing is enforcement because $w_{plant} = (1 - Ca)(1 - y)$, where experienced selfishness is $y = (1 - a)\bar{z}$, and hence $\frac{dy}{da} < 0$ when $\bar{z} > 0$.

Table S2 Parameters

Parameter	Actor	Parameter meaning	Range (point estimate)[a]	Rationale of range (rationale of point estimate)
Replicators in a protocell				
C	gene	cost of enforcement	0-1 (0.1)	full range (relatively low cost)
R	gene	group relatedness (including self)	0-1 (0.25)	full range (for comparison with animal group model)
Endosymbionts of isogamous cells				
C	zygote	cost of enforcement	0-1 (0.1)	full range (assumed lower than c_w, see model description)
c_{war}	zygote	cost of losing one set of mitochondria	0-1 (0.4)	full range (intermediate value)
Early multicellularity and cancer				
C	organism	cost of enforcement	0-1 (0.05)	full range (estimate)
n	organism	number of cells in primitive adult organism	(10^7)	estimate
μ	cell	per-division mutation-to-cancer rate	10^{-12} to 10^{-3} (1.1×10^{-7})	wide range (conservative estimate, see model description)
Reproductive conflict in an animal group				
C	animal	cost of enforcement	0-1 (0.1)	full range (relatively low cost)
c_{pun}	animal	cost of being punished	(1)	(exiled individuals cannot reproduce)
ρ	animal	group relatedness (not including self)	0-1 (0.25)	full range (half sibs, consistent with data[42])
n	animal	group size	(10)	(small group)

Mutualism: partner choice in host-symbiont systems				
C	squid	cost of enforcement	0-1 (0.05)	full range (low cost)
c_{pred}	squid	risk of death for a wholly-dark squid	(0.75)	(assume light production is important)
c_{lux}	bacteria	cost of luminescence	0-1 (0.25)	full range (luminescence assumed costly)
c_{host}	bacteria	cost of host squid detecting cheating	(1)	(assume enforcement is effective)
R	bacteria	relatedness between bacterial cells in squid light organ	(0.1)	(6 crypts in squid light organ, 1-2 cells per crypt)
Mutualism: partner manipulation in plant-pollinator interactions				
C	plant	cost of enforcement	0-1 (0.5)	full range (assume that thickening corolla is relatively costly to future flower production)
c_{visit}	pollinator	cost of legitimately visiting a flower	(1)	(arbitrary)
c_{rob}	pollinator	cost of nectar robbing	$0–c_{visit}$ (0.5)	full range for which $c_{rob} \leq c_{visit}$ (robbing is twice as efficient as legitimate visit)

[a] Point estimates are used for the illustrative "Dynamics" plots in Fig. 3.

Table S3: Fraternal versus egalitarian alliances[*]

	Egalitarian	Fraternal
Example of cooperative alliance forged	Different molecules in compartments; genes in chromosomes; nuclear and organelle in cells, mutualisms between species	Same molecules in compartments; cells in multicellular organisms; individuals in family group
Units	Unlike, non-fungible	Like, fungible
Reproductive division of labour	No	Possible
Initial advantage	Division of labour; combination of functions	Economies of scale; later division of labour

*Modified from[43, 44]. To encompass the full spectrum of cooperative systems, we do not only consider major transitions in evolution and instead discuss fraternal and egalitarian *alliances* that may or may not have undergone a major transition.

Box S1 Enforcement as an evolutionary byproduct

We have defined enforcement relatively narrowly as processes that evolve because they limit selfishness and promote cooperation (Definition of enforcement, above). However, similar mechanisms that limit selfishness can also evolve independently of their effects on cooperation[26, 45]. One potential example is worker policing via egg eating in the eusocial hymenoptera[16]. Here, workers are thought to remove other workers' eggs primarily because they are more related to the queen's eggs. By removing the incentive for workers to lay their own eggs, this behaviour leads to the evolution of increased worker cooperation[16]. The increase in cooperation only occurs after a delay, however, and is unlikely to contribute to natural selection for policing, which would be required for enforcement proper. That said, policing may immediately promote cooperation when it interacts with another conflict (over sex allocation), which would render it enforcement *sensu stricto*[46] according to our definition. There are many other potential examples of byproduct enforcement. For example, genetic recombination breaks down linkage groups that may act against the interest of the genome[47] but may have evolved for a different reason[48]. Similarly, pleiotropy (Model 3, above) is a major feature of all genomes so there may be many cases where this limits the mutational target size that generates rebellious phenotypes as a byproduct[34].

There are two important caveats to this discussion of byproduct enforcement. First, byproduct enforcement may be improved upon by subsequent natural selection for cooperation. Take for example, the finding that legumes use root senescence to cut off non-cooperative (non-nitrogen fixing) bacteria[49]. It has been argued that this may be a preadaptation because the senescence of low productivity roots is a general feature of plant biology that existed before root nodules[50]. This would be byproduct enforcement. However, the mechanism of root senescence was likely fine-tuned once root nodules evolved, thus rendering it enforcement proper. The second caveat is that species-level selection is predicted to enrich for enforcement mechanisms, byproduct or otherwise, if within-species conflict is linked to extinction risk[51]. Consistent with this, there is evidence that transposon abundance is associated with extinction in bacteria, and that termite species with lower societal conflict will outcompete species with higher levels[52]. Species-level processes are often overlooked in the literature, probably due to long-standing nervousness associated with the misapplication of group-selection thinking[53]. However, these processes may greatly enrich for cooperative systems with enforcement or byproduct enforcement[51] (in the same way they are thought to enrich for sexual species[54]). In this sense, then, byproduct enforcement is also likely to be selected due to its effects on cooperation.

References

1. Irwin, R.E., Bronstein, J.L., Manson, J.S., and Richardson, L. (2010). Nectar robbing: ecological and evolutionary perspectives. Annual Review of Ecology, Evolution, and Systematics *41*, 271-292.
2. Trivers, R.L. (1971). Evolution of reciprocal altruism. Q. Rev. Biol *46*, 35-57.
3. Gardner, A., and Foster, K.R. (2008). The evolution and ecology of cooperation: history and concepts. In Ecology of Social Evolution, J. Korb and J. Heinz, eds. (Springer), pp. 1-36.
4. Ågren, J.A. (2014). Evolutionary transitions in individuality: insights from transposable elements. Trends Ecol Evol *29*, 90-96.
5. Wenseleers, T., Ratnieks, F.L.W., and Billen, J. (2003). Caste fate conflict in swarm-founding social Hymenoptera: an inclusive fitness analysis. J. Evol. Biol. *16*, 647-658.
6. Ratnieks, F.L.W. (2001). Heirs and spares: caste conflict and excess queen production in Melipona bees. Behav. Ecol. Sociobiol. *50*, 467-473.
7. Wenseleers, T., and Ratnieks, F.L.W. (2004). Tragedy of the commons in Melipona bees. Proc. R. Soc. B *271*, 310-312.
8. Evan, G.I., and Vousden, K.H. (2001). Proliferation, cell cycle and apoptosis in cancer. Nature *411*, 342-348.
9. Grosberg, R.K., and Strathmann, R.R. (2007). The evolution of multicellularity: A minor major transition? Annual Review of Ecology Evolution and Systematics *38*, 621-654.
10. Martin, S.J., Beekman, M., Wossler, T.C., and Ratnieks, F.L.W. (2002). Parasitic Cape honeybee workers, Apis mellifera capensis, evade policing. Nature *415*, 163.
11. Montague, C.E., and Oldroyd, B.P. (1998). The Evolution of Worker Sterility in Honey Bees: An Investigation into a Behavioral Mutant Causing Failure of Worker Policing. Evolution *52*, 1408-1415.
12. Hurst, L.D., Atlan, A., and Bengtsson, B.O. (1996). Genetic conflicts. The Quarterly Review of Biology *71*, 317-364.
13. Ratnieks, F.L., Foster, K.R., and Wenseleers, T. (2006). Conflict resolution in insect societies. Annu Rev Entomol *51*, 581-608.
14. Ratnieks, F.L., and Wenseleers, T. (2008). Altruism in insect societies and beyond: voluntary or enforced? Trends Ecol Evol *23*, 45-52.
15. Ratnieks, F.L.W., Foster, K.R., and Wenseleers, T. (2006). Conflict resolution in insect societies. Annu Rev Entomol *51*, 581-608.
16. Wenseleers, T., and Ratnieks, F.L.W. (2006). Enforced altruism in insect societies. Nature *444*, 50.
17. Szathmáry, E. (2006). The origin of replicators and reproducers. Philosophical Transactions of the Royal Society of London B: Biological Sciences *361*, 1761-1776.
18. Frank, S.A. (1995). Mutual policing and repression of competition in the evolution of cooperative groups. Nature *377*, 520-522.
19. El Mouden, C., West, S.A., and Gardner, A. (2010). The enforcement of cooperation by policing. Evolution *64*, 2139-2152.
20. Maynard Smith, J. (1979). Hypercycles and the origin of life. Nature *280*, 445-446.
21. Goodenough, U., and Heitman, J. (2014). Origins of Eukaryotic Sexual Reproduction. Cold Spring Harbor Perspectives in Biology *6*.
22. Riley, M.A., and Wertz, J.E. (2002). Bacteriocins: evolution, ecology, and application. Annu Rev Microbiol *56*.
23. Hastings, I.M. (1992). Population genetic aspects of deleterious cytoplasmic genomes and their effect on the evolution of sexual reproduction. Genetics Research *59*, 215-225.

24. Sato, M., and Sato, K. (2013). Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1833*, 1979-1984.
25. Morris, K., Austin, J.J., and Belov, K. (2013). Low major histocompatibility complex diversity in the Tasmanian devil predates European settlement and may explain susceptibility to disease epidemics. Biol. Lett. *9*, 20120900.
26. Frank, S.A. (2003). Perspective: Repression of competition and the evolution of cooperation. Evolution *57*, 693-705.
27. Nunney, L. (2013). The real war on cancer: the evolutionary dynamics of cancer suppression. Evol. Appl. *6*, 11-19.
28. Nunney, L. (1999). Lineage selection and the evolution of multistage carcinogenesis. Proc. R. Soc. B *266*, 493-498.
29. Aktipis, C.A., Boddy, A.M., Jansen, G., Hibner, U., Hochberg, M.E., Maley, C.C., and Wilkinson, G.S. (2015). Cancer across the tree of life: cooperation and cheating in multicellularity. Philos T R Soc B *370*, 20140219.
30. Michod, R.E., and Roze, D. (1999). Cooperation and conflict in the evolution of individuality. III. Transitions in the unit of fitness. Lectures on mathematics in the life sciences, 47-92.
31. Singh, M., and Boomsma, J.J. (2015). Policing and punishment across the domains of social evolution. Oikos *124*, 971-982.
32. Nunney, L. (1999). Lineage selection and the evolution of multistage carcinogenesis. Proceedings of the Royal Society of London. Series B: Biological Sciences *266*, 493-498.
33. Drake, J.W., Charlesworth, B., Charlesworth, D., and Crow, J.F. (1998). Rates of spontaneous mutation. Genetics *148*, 1667-1686.
34. Foster, K.R., Parkinson, K., and Thompson, C.R.L. (2007). What can microbial genetics teach sociobiology? Trends Genet. *23*, 74-80.
35. Johnson, L.J., Cotton, J.A., Lichtenstein, C.P., Elgar, G.S., Nichols, R.A., Polly, p.D., and Le Comber, S.C. (2011). Stops making sense: translational trade-offs and stop codon reassignment. BMC Evolutionary Biology *11*, 227-227.
36. Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., Vitale, L., Pelleri, M.C., Tassani, S., Piva, F., et al. (2013). An estimation of the number of cells in the human body. Annals of Human Biology *40*, 463-471.
37. Hartmann, A.C., Baird, A.H., Knowlton, N., and Huang, D. (2017). The Paradox of Environmental Symbiont Acquisition in Obligate Mutualisms. Current Biology *27*, 3711-3716. e3713.
38. Foster, K.R., Schluter, J., Coyte, K.Z., and Rakoff-Nahoum, S. (2017). The evolution of the host microbiome as an ecosystem on a leash. Nature *548*, 43-51.
39. Schwartzman, J.A., and Ruby, E.G. (2016). A conserved chemical dialog of mutualism: lessons from squid and vibrio. Microbes and Infection *18*, 1-10.
40. Nyholm, S.V., and McFall-Ngai, M. (2004). The winnowing: establishing the squid–Vibrio symbiosis. Nature Reviews Microbiology *2*, 632-642.
41. Foster, K.R., and Kokko, H. (2006). Cheating can stabilize cooperation in mutualisms. Proceedings of the Royal Society of London B: Biological Sciences *273*, 2233-2239.
42. Griffin, A.S., Pemberton, J.M., Brotherton, P.N.M., McIlrath, G., Gaynor, D., Kansky, R., O'Riain, J., and Clutton-Brock, T.H. (2003). A genetic analysis of breeding success in the cooperative meerkat (Suricata suricatta). Behavioral Ecology *14*, 472-480.
43. Queller, D.C. (2000). Relatedness and the fraternal major transitions. Philos T R Soc B *355*, 1647-1655.
44. Queller, D.C. (1997). Cooperators since life began. Q. Rev. Biol *72*, 184-188.

45.   Riehl, C., and Frederickson, M.E. (2016). Cheating and punishment in cooperative animal societies. Philos Trans R Soc Lond B Biol Sci *371*, 20150090.
46.   Foster, K.R., and Ratnieks, F.L. (2001). The effect of sex-allocation biasing on the evolution of worker policing in hymenopteran societies. Am Nat *158*, 615-623.
47.   Haig, D., and Grafen, A. (1991). Genetic scrambling as a defense against meiotic drive. J. Theor. Biol. *153*, 531-558.
48.   Otto, S.P., and Lenormand, T. (2002). Resolving the paradox of sex and recombination. Nat. Rev. Genet. *3*, 252-261.
49.   Kiers, E.T., Rousseau, R.A., West, S.A., and Denison, R.F. (2003). Host sanctions and the legume-rhizobium mutualism. Nature *425*, 78-81.
50.   Frederickson, M.E. (2013). Rethinking mutualism stability: cheaters and the evolution of sanctions. Q Rev Biol *88*, 269-295.
51.   Rankin, D.J., Lopez-Sepulcre, A., Foster, K.R., and Kokko, H. (2007). Species-level selection reduces selfishness through competitive exclusion. J. Evol. Biol. *20*, 1459-1468.
52.   Korb, J., and Foster, K.R. (2010). Ecological competition favours cooperation in termite societies. Ecol. Lett. *13*, 754-760.
53.   Foster, K. (2009). A defense of sociobiology. Cold Spring Harbor symposia on quantitative biology *74*, 403-418.
54.   de Vienne, D.M., Giraud, T., and Gouyon, P.H. (2013). Lineage selection and the maintenance of sex. Plos One *8*.